



# Open Library 2023 Planning

*Open Library's [mission](#) is to catalog the world's books and make them accessible and useful.*


Official  2023: Open Library - Priorities

Last year:  2022: Open Library - Planning

## Background

In 2005, the Internet Archive co-founded the [Open Content Alliance \(OCA\)](#) to digitize and archive the world's books. Open Library was born shortly after, circa 2006, as a card catalog for books. Since its inception, OpenLibrary.org has featured [links](#) (where available) to read scans of public domain and unrestricted books. In 2007 the Internet Archive received official [Special Library](#) (Rush Brandis, California State Library) [accreditation](#) by the State of California. In 2011, the Internet Archive piloted its online [lending program](#) on OpenLibrary.org, in collaboration with several [OCA partners](#). In 2014, the Internet Archive received a [funding grant](#) under LSTA from the California State Library system to support the digitization of books and the development of OpenLibrary.org as an online lending library service. After the completion of the grant, Open Library became unstaffed, entered a period of hiatus, and was slated to be decommissioned. The project was revived in 2015 and between 2016-2018, Open Library and its [community](#) entered a period of resurgence. The project delivered key improvements, like supporting mobile and the “[Want to Read](#)” button. In 2019, Open Library [launched](#) the [Book Sponsorship program](#), enabling the community to help fund the digitization of books missing from the lending program. In 2020, Open Library invested in sustainability (upgrading to [Python3](#), establishing an [import pipeline](#), and production Docker), simplifying the [patron experience](#), and improving [discovery](#) & adoption by booklovers. In 2021, we focused on imports, partnerships, search, and becoming the Open Book Catalog for the Internet. In 2022, we focused on increasing direct value to patrons by improving the core usability and experience of the service: book page design, edition-level search, navigation, mobile, and performance.

## Recap (added 2022-10-07)

In  2022: Open Library - Priorities , we focused our efforts on improving the core experiences of the Open Library service: the Books Page & Search.

## Analysis & Theory of Change

In its early stages of development, the Open Library benefited from being an underdog and learning from the examples set by other industry-leaders. With limited resources, we felt it made more sense to focus on core, broadly appealing features with outsized potential for improving the average experience than to segment and understand the specific needs of our 6M patrons. But as we approach parity with these services and notice certain things working or not working, it's becoming increasingly important, to the patrons we serve, that we acknowledge the ways in which the Open Library is different and to focus on our unique value proposition. When we design for everyone, we risk building something that [isn't ideal for anyone](#) and we miss opportunities to steward high impact changes for key audiences who may disproportionately derive essential benefit from using the Open Library. This year we'd like to invest in better understanding our core patrons and their needs so we can offer the best for those who uniquely rely on our services.

## *Open Library's theme for 2023 is **focus**: to identify, understand, & serve key audience segments more effectively*

**Context:** This roadmap is an initial proposal from staff and is not a commitment to what will end up in the

☰ 2023: Open Library - Priorities document. For planning, consider also viewing the

☰ Open Library - Long-Term Vision (5-10yr) document and ☰ 2022: Open Library - Planning .

**Story:** At the end of 2023, I'm hoping that the Open Library's catalog has lots of well readable books from many different/new sources. I'm hoping that we can generally continue to make the experience simpler, less confusing, and easier to use. I hope we're able to isolate personas: to identify why core groups of patrons are deciding to come to the site and focus on collecting high-quality, organized collections of useful content and present them in ways which meaningfully support their goals. Whether this is underserved communities looking for equitable content, self-learners, researchers. I think this year the changes we make to our experience, to the catalog, and the tools we provide to the community should try to address the specific core audiences (e.g. focus on importing edu books? organize edu books into curriculum? banned books? organize diverse collections? building a wikipedia companion library?). Lets ask patrons and learn what their needs are and provide unique, meaningful, essential value.

### *Priorities → Epics*

★ Yearly Commitments (~1Q = 3mo = 12wk):

- a. Project kickoffs (2wk), October event + community celebration (2wk), special projects (2 wk), board report (1wk), performance reviews (1wk), EOY Community planning + Priority doc (2wk), 3 weekly calls: ABC, OLCC, Design (2 wk)

#### 1. Design Objectives - *Do fewer things, better, for the audiences that matter.*

- b. ☒ **Conduct Lending Survey.** Understand readers needs → increase good borrows
- c. ☒ **Personas & Design Research.** Gaining a unified understanding of our patron segment
- d. **Segment specific portals.** e.g. k-12? textbooks portal? Add reading levels, OSP, subjects to Search
- e. **Consistent UI Components.** All book cards (search, trending, my books) should have the same style/options: **Reading Log dropper button** (search vs books page), consolidated read button.
- f. **Design System.** Develop pattern libraries for major components in Figma (Fellowship)
- g. **Onboarding.** Simplify + streamline registration process, send patrons to Open Library, POST activate to archive.org and then login? send patrons to Archive.org which redirects to Open Library with a key? Add a welcoming landing page.

#### 2. Catalog, Imports & Holdings Fix IA imports, crons for promise items jchamp@archive.org

- a. ☒ **Covers.** Import + archive Book Covers (BWB, Amz)
- b. **Enable Development & Administration.** Fixing IA imports, error handling & sentry reporting, sureing architecture, if time enable community contribution via ☒ **ability to run importer pipeline locally**
- c. **Central Data Quality Filter.** Central filter to prevent bad imports. Prevent insubstantial amz imports (e.g. asin only) and low quality promise items, notebooks. **Better Author & Work resolution** (than exact str match; avoid creating new authors/works). **Source-specific filters?**
- d. **QoS: Throughout & Performance.** haproxy, ☒ multi-threading, separate db table for pending v. completed imports, run audit
- e. **Establish & codify new data pipelines.** MARC, Promise Cron augmented w/ amazon, **Google Books, Web books, Trusted Book Providers**
- f. **Increase good unsolicited book donations** (e.g. wishlist books) DWWI → Internet Archive

- g. **Identify top books to accession**, per major audience segment (use reading log stats? borrow stats?)  
→ Release roll-up? of anonymized WTR **from** data dump for top 10k books + add reading Log data in solr. `drini@archive.org`
  - h. **Book Series**. Support for creating, viewing, and searching for book series
3. Core & Devops
- a. **Site Performance Epic**. Faster book pages & search, js
  - b. **100% off trusty** → on docker (`ol-db[1,2]`, `ol-backup0`, `ol-home`, `ol-www1`) w/ [monit](#) (via ansible)

- **Notes**

- **Personas Research**: Gaining a unified understanding of our patron segment: who uses Open Library (*Librarians? Teachers? Parents? Students? Researchers? Self-learners? Fiction readers? Developers?*), how, and why (what are their reasons/needs & how may we be essential to them?).
  - Phase I: Research Questions → Sort + Categorize into sources: survey, interview, data, etc  
→ ~5 Screener Q's w/ cutoffs (usage, age, location) to control for audience → conducting user research via channels → Synthesis → Themes (group synthesis) → [Personas](#) (1p max per persona; quotes, needs, desires) → Phase II: direct original Q's to personas (~30h @ \$150)

## Aspirational, Ideas, Projects for Fellows, & Overflow

See also: [Open Library - Long-Term Vision \(5-10yr\)](#)

1. [Open Library - Canonical Tags](#) : Unifying bookshelves to community tags (lists, reading log, community reviews, subjects → easy tags)
2. Experiment with Open Library as a publishing platform?
3. Book Trails & Book Tournaments + OSP @mekarpeles
4. [Book Page Score](#) ([demo bookmarket](#))
5. [Book Citations](#) Fatcat + Wikipedia? Book Genome? Lexile Reading Levels? OBGp
6. Librarian Merge Queue v2 Redesign (Fellow)
7. Direct Manipulation Editing by Confirmed Editors
8. Ability to go into read-only during power outage?
9. Enable [Follow Patrons](#) & activity feeds? (needs doc)
10. Federation & Decentralized Open Library clients via activity-pub?
11. Edition merging. `mark.heiman@gmail.com` ?

## 2023 Community Notes

**Homework:** Below, please add a list of **3 to 5** things you'd like to see happen in 2023:

This could include...

- A specific feature (e.g. *Series, Follow Users*)
- An area (e.g. *APIs*)
- A metric (e.g. *Performance, fewer errors*)
- Systems / tooling / research (e.g. *a Design System, User Research to answer some Q*)
- A specific problem & solution (e.g. *Dead Ends? Confusion? Inconsistency? Unfocus?*)

## Summary

- **Design system, research & resources**
- **Site performance**
- **Improving imports quality**
- **Better onboarding for patrons**
- **Book Series**


- **I have this book** → **Sponsorship / Donation**
  - **Covers**
  - Children's Books, Summer reading (related to **personas**)
  - **Consistent book cards**; e.g. reading log button across the site
  - Canonical Tags
- 
- **Scott Barnes**
    1. **Performance, Metrics:** **Improve load times.** consistently keep Largest Contentful Paint below... some threshold for some percentage of users. (<= 2s? For what percentage of users I have no idea.)
    2. **Stakeholder Exercises** (Information gathering): **Codifying processes for getting feedback from stakeholders.** develop objectives, metrics, and a process for an ongoing, articulated-in-writing, effort to collect information from a diverse set of partners/users/patrons (and possibly developers?), and a set time to revisit the information collected (annually?), along with a set schedule for analyzing the information and making decisions upon it. I am suggesting this because I think it is nice to have current, concrete information collected directly from partners/users/patrons/contributors so there isn't a need to guess about the problems users/patrons face, or their proportion or magnitude, when we try to come up with solutions. It may be because of my ignorance about what's collected, but I am not sure if there is objective data on the basis of which to assert (1) what partners/users/patrons/etc. want or the problems they face, (2) what categories of patrons/users there are, (3) what the proportions are, (4) or how much group category membership overlap.
- 
- **Lisa**
    1. Correctly matched imports. (I do realize this is a problem on the IA side to be fixed as well.)
      - A large chunk of patron services requests are about fixing metadata/links across IA/OL
    2. Selecting specific editions for the reading log.
    3. Ability to add multiple editions to the reading log, ie print AND audio.
- 
- **Dana**
    1. Design System: Convert the discussion we've been having to be more concrete/structured
      - **Open Library Design System** (it's a start!)
    2. Reading Challenges might increase engagement
- 
- **Samuel**
    1. Design System: working towards a public-facing, cohesive, clear, and to the point design system document or page would be really cool → something similar to the [WikiMedia Design Style Guide](#), including principles, visual style, and ideally components and resources that contributors would be able to use via Figma or GitHub
      - A sub-point to this would be defining a more consistent procedure for UX research and usability testing, internally or to be included in Design System
    2. Education/Social instructional design: I think it would be fun to work on something akin to the project I initially proposed when first reaching out to Open Library, which was adding interactive front-end modules for the education/help pages (e.g. [Reading Log & Lists](#), [Borrowing Books](#), [Editing](#), [Reading on OL](#), etc.)
      - A related idea is **adding a Share button to Lists** and **creating a new social card for Lists and Reading Logs**, this goes along with interactive education modules because this functionality

could be included/promoted in the help guides (e.g. create a new list, put more books in it, share a card of your list on Twitter, etc.)

- **Allegra**

1. Make the website easier to navigate (+1 Sam's educational instructional design)

- **Mark**

1. **Series!**
2. **Tags.** See:  Open Library - Canonical Tags . Subjects as first class objects (possibly publishers and contributors as well)
3. **Book Data Structures:** Tags, pen names, work titles, series (metadata weaknesses)
4. **Import improvements.** The import process creates a huge amount of extra work for us, creating many duplicate authors and works that have to be manually fixed. I think an engineering focus on imports could improve this process significantly.
5. **Pen Names:** We have a fair bit of noise and confusion around pen names in the catalog. The recommended practice has been to merge all author records for the various names into one; this significantly reduces findability for the pen names, and creates confusion when the author on the works doesn't match the author on the actual books. Leaving them separate, however, provides no consistent mechanism for finding all the books by a particular author. We also have syndicated authors (e.g. Carolyn Keene) where a number of individuals write under a single name. Developing data structures and interfaces for managing these relationships would be a big win.
6. **Work title handling:** I've been doing a lot of cleanup on authors that have titles translated into many different languages (and sometimes multiple different titles in the same language). Currently title data is purely descriptive, based on what editions are present. Having a more structured way to handle title data at the work level would be very powerful, and also contribute to #3 above.

- **Onno**

1. **Focus.** OpenLibrary needs to decide whether it wants to be primarily an online library, an online archive, a social book site (like TheStoryGraph, Bookwyrms, Goodreads etc) or a library support service. While some of the functions are complementary, others are not:
  - A library mainly needs trending books, and will discard many or all copies once the trend runs out. Duplicates are therefore necessary for a library. It only needs broad categories for readers to find books and some kind of suggestion system.
  - An archive mainly needs books that are in danger of being lost - those will usually be uncommon books only few people want to read. It also needs a very detailed and accurate metadata system to support researchers in finding the books they need as sources.
  - A social book site mainly needs ways for users to interact, as well as a good collection of trending books. (Short addendum: We do not need to do everything on our own. If we create a good library support service, tools like Bookwyrms can use that data to create a social site.)
  - A library support service mainly needs background data that makes librarians work easier (which might be gathered via social book site features). For example, ways to import title data in their catalogue, find exchange partners (open library as "lender of last resort", where libraries can "order" a copy they do not locally have), or access to user data to decide which books to buy with their limited budget.
2. **Librarian Metadata Guides.** In any case, we need better metadata for books - mainly tags and dates, but also translations and publishers. How detailed the metadata needs to be depends on



point 1. I think we should focus on improving storage, retrieval, management, and accessibility of metadata (I already opened several issues with feature requests in this regard, but there are probably plenty of areas I am not aware of).

3. **Finding new books.** If we go the library or social route, we also need to improve personalization and recommendations. (I wrote a detailed proposal for that as issue #7102)
4. **Donating books**

- **Bernat**

1. **Problem: There's no easy (and scaled) way to know: is a book for a young audience?**
  - Context: most libraries have a "children's corner". And when you go there you know that you and your kids will find appropriate books (I'm talking about reading skill, length, topics, etc..). Through my work at dreambooks.club I've been trying to create such "children's corner".
  - As far as I know, from an OL data perspective the only way to know if a book is for the appropriate audience is by filtering through "Juvenile/Children" subjects. (This is a standardized approach)
  - Issue: Import bot doesn't always import the subject. Is it because it's not available? Or maybe we're skipping it?
  - Solutions:
    - Improve ImportBot to add subjects. Especially those that will be helpful in defining the audience.
2. **Problem: Many covers are non existent or low resolution**
  - Solution 1: Add covers to works without one: Go through recently added books and/or books that were published in the last 5-15 years AND don't have covers. Then go one by one and find the best possible cover for it. It's easy to scrape from amazon, there's a url that takes a isbn so you could confidently know that you're getting the right cover. The % of wrong images with this approach is probably low.
  - Solution# 2: Improve the quality of the main cover (that's the one that shows up on search results). Similar to the above but you check if the cover in OL is low resolution in which case, scrape and update.
  - Solution #3: Better Cover Queue tool. Essentially a queue of suggested updates. It would be a tool for librarians or admins. You see a list of books (name, current cover, suggested cover). Then a human decides to commit the cover update after looking at meta info. This is probably more expensive to build and maintain but would yield the best results overtime. If it works well, it could be made public, and the community could contribute. Think of "10 books you could improve today".
3. **Problem: Duplicated works with the same name**
  - Context: I know very little of the existing efforts and maybe the proposed solutions below have already been considered. That being said I'd think of the problem as two parts: 1. Prevent new books/works with the same name from being created. 2. Reduce the existing books wit the same name.
  - Solution #1: bot dupes: Check how many books a bot adds that are duplicates. If that happens, evaluate if it would be ok preventing bots from adding dupes. If there's value in the bot creating it, add those to the merge queue.
  - Solution #1.2: bot dupes for other languages: Let's assume when a bot imports we know that a book is being imported in a language different from english. In that case, try different things to see if it's a dupe, if so, add it to the merge queue. Some ideas on how to know the language:

- [This is an example of a language dupe](#):
  - You could check for the notes value: Source title: Neu (Catalan Edition). if you see {language} edition. You know it's a language dupe and then you kick it to the merge queue.
  - It's probably pretty hard to auto-merge. But you could try translating the title to english, and see if there's a book for any of the authors with the translated title.
  - Alternatively, you could try to find the right work. This case is easy because you could translate the book title from catalan to english and get "Snow". Then you'd query {authorKey}
  - ([the following edition](#) is an import bot created book of a "catalan" edition of Sam Usher Snow). I'm not sure if there's a better way, but you could scrape amazon via the isbn link. In this case the isbn corresponds to amazon UK so it might be tricky. But maybe the import bot knew that?
  - Solution #2: human dupes: Improve the UX in the "add book" form. Treat the title as a "search input", right above the "add" button show a list of potential duplicates. Clicking will send the user to that page and take them out of the creation flow.
  - Solution #3: reduce duplicated books. x% of search queries kick a parallel process (non-blocking for the user) that evaluates the search results. If it finds that two results have the same title it adds them to the merge queue. You can play with the % to not burn the server and/or a max per day. This process will optimize for the most searched books. Overtime you can try to auto-merge given the right context.
4. **Problem: L size Book covers are still low res when fully showing a cover**
- Context: for children's books the cover is critical, It's nice to be able to see it
  - Example: <https://covers.openlibrary.org/b/olid/OL34782729M-L.jpg>
  - Children's book have beautiful illustrations. It would be great to be able to have access to the highest resolution. Ideally x2 from -L.
  - Solution: add a -XL cover ending. Maybe rate limit it?

## • Nazar

1. Inconsistency in Reading Log "Want to Read" button. It is only shown near books in search results, but not on author page, homepage or subject page
2. Edition merging

## • Allan LEONARD

1. Re **Onboarding** (above), having read Open Library Long-Term Vision heading [Decentralization](#), which of the two Onboarding options would be more consistent/keep more options open: (a) "send patrons to Open Library, POST activate to archive.org and then login" or (b) "send patrons to Archive.org which redirects to Open Library with a key"? I'm a registered user of a [Bookwyrms instance](#) and I wonder how patrons would be linked to Open Library and/or Archive.org.
  - There are two efforts which seems to be focusing on federation and decentralization: <https://annas-archive.org/> and <https://bookwyrms.social/>
2. Re Onno (above), FWIW my preference would be for Open Library to NOT pursue or develop as a social book site; such is better served in a federated/Bookwyrms universe. This discussion is about core competencies, and developing and maintaining a singular universal social book site (a la Good Reads) could consume evermore technical resources. Rather, **my preferred priority would be: the best universal catalogue**, for use by libraries and individual readers. This lends the

discussion to Onno's points about better metadata, etc. I also think researching and cataloguing rare/out-of-print books is important and consistent with the Archive.org mission.

3. Re Bernat (above), I'm all for using automation for book cover discovery. Perhaps my own bias, but I also think setting up a manual book cover discovery project for human volunteers would be successful. (Volunteers respond to specific task requests.) Perhaps even a **combination of automation and human discovery/verification for book covers**, akin to [Galaxy Zoo](#)?

- **Chris Clauss**

1. **Metrics.** We have done a good job on finding and tracking errors and exceptions in our code but we could do a better job on metrics. Capturing **statsd metrics and improving our Grafana user experience** could help us to better understand our performance and UX pain points.
2. **Benchmarks.** We need to **write more benchmarks** for measuring performance of key processes. Using periodic runs of these benchmarks could give us a sense of improvements or regressions.
3. **API Docs & Experience.** We have started a migration towards **Swagger documentation of our Application Programming Interfaces (APIs)** so we should drive that forward as a way to bring new developers to our community.

- **Arseny (architectural theory)**

1. **Research:** Finding translations & other Editions
2. **Donate books** (to nearby libraries)

- **Ave**

1. **Open Library Summer Activities.** To correspond with the existing K-12 Student Library carousel adding curated worksheets/resources for students and parents to use over the Summer.
  - Concerns: Re Onno above mentioned having a conversation about the audience of the Open Library and the focus. This project would definitely view the Open Library as a more traditional library for patrons.
  - Check out the brainstorm for this project. [Summer Reading Activities](#)
2. **Overall design of the website.** The site would benefit from updating the use of fonts. The colors and the graphics used to make the website more appealing to patrons.

## 2023 Breakdown

# 2023 Open Library Breakdown

Process	1
Project Kickoffs	2
2023 Priorities	3
Commitments	4
Mek Karpeles	4
jchamp@archive.org	5
drini@archive.org	6
cclauss@archive.org	7
scottreidbarnes@gmail.com	8












Roadmap	9
Menu for Contributors	10
Wins Log	11

## Process

1. Schedule all [Project Kickoff](#) meetings using [template](#) to identify + assign commitments
2. After each kickoff, import each project's [Commitments](#) into this planning doc
3. Staff decides what will be done in H1 v. H2, creates issues for all H1 tasks and moves them on to the [2023 milestone](#)
4. All issues for H2 get added as checklist items on a tracking issue which gets added to 2023 project board
5. 1:1s to double check H1 & H2 priorities

## Project Kickoffs



- ✓ ~~Tues -  Open Library - Unified v2 APIs  Extending Solr Data~~
- ✓ ~~Wed -  Open Library: Imports (Improving Catalog Quality Every Month)~~
- ✓ ~~Thursday -  Open Library - Canonical Tags~~
- ✓ ~~Jan 25, 2023 -  Open Library - Site Performance~~
- ✓ ~~Jan 26, 2023 ( cclauss@archive.org + Mek Karpeles + drini@archive.org 1:1)~~
  - ~~ Open Library: Dev Ops (Provisioning & Deployment)~~
- ✓ ~~Jan 31, 2023 -  Open Library - Design Strategy + Communications & Development Strategy?~~
  - ** Open Library - Communications Team - Community Call**
    - Blog strategy? Rachel (  Open Library Marketing presentation ) + Nick
  - **Development**
    - Work with Development Team (Christina) to explore funding opportunities (IASoC & Design Research budget)
  - **Design Research**
    - **Personas Design Research.** Gain unified understanding of audience segments
    - **More targeted collections** What do patrons need? e.g. K-12? textbooks?
    - **Conduct Lending Survey.** Understand readers needs → promote good borrows + **Seamless Lending Flow**. 1-app, offline, simple, clear, unified lending flow

## 2023 Priorities





1. **Enable librarians** ASAP to create [Series](#) & [Tags](#) + select & manage in [Bulk Editor](#)
2. **Edu collections** for learners using Subjects, [Tags](#) – requires [design research](#)
3. **Fundamentals:** [Performance](#), [Provisioning](#), [Imports](#) + [Covers](#), Core Design
4. **Futureproof:** Safeguard [5-year strategy](#) of OL; **BookServer**, **TBP**

2023: 121 issues, 58 unassigned

Each item on this list should have an epic **or** managed `tag` + a set of issues on 2023 milestone we can query for

- ☒ [Books Page: mobile Redesign & cleanup - Issue #6484](#) jchamp@archive.org
- ☐ [List Improvements - #7390](#):  GSoC 2023 - Open Library - Debbie jchamp@archive.org
- ☒ [Yearly Reading Goals](#) jchamp@archive.org
- ☐ [Book Covers](#) Mek Karpeles
- ☐ [Provisioning](#)  Open Library: Dev-Ops (Provisioning & Deployment) cclauss@archive.org
- ☐ Community
  - o [My Books Desktop Redesign #7651](#) d.feinsch@gmail.com
  - o [My Books Mobile Redesign #7376](#) samzgrunebaum@gmail.com
  - o [Librarian Merge Queue Redesign - Issue #6816](#) Mek Karpeles → Ed Tee [#7512](#)
  - o [Redesign Account Registration #7647](#) & [#7694](#) ruijingya.tang@gmail.com
  - o [Open Library Team Page #7650](#) debbie.san16@gmail.com
  - o Information Architecture – Connor

### Needs Planning!

- ☐ [Improve imports #7684](#)  **Open Library - Imports** scottreidbarnes@gmail.com jchamp@archive.org Charles Horn drini@archive.org
  - o Community batch imports, [Don't create duplicate author records for M. <surname> with exact date match - Issue #7349](#)
  - o [Book Scores #7661](#) jchamp@archive.org
- ☐ [Site Performance \(Epic #4485\)](#)  Open Library - Site Performance Jayden
- ☐ [Canonical Tags \(Epic #2819\)](#)  Open Library - Canonical Tags Mek Karpeles
  - o [ILE Toolbar Bulk Edit #7486](#) mark.heiman@gmail.com
  - o Subjects / collections page design w/ Tags? d.feinsch@gmail.com
- ☐ [Book Series - Book #6718](#) drini@archive.org
-  Open Library - Design Team - Persona Research abbey.ripstra@gmail.com + ruijingya.tang@gmail.com
- **BookServer** (needs doc + slides) Mek Karpeles
  - o [OPDS endpoint PR #7078](#)
  - o [Experiment with direct book providers #6585](#) + availability + dwwi in solr drini@archive.org
  - o Providers in Solr (augmented per page by infogami, but not in index) → OPDS
  - o Query princes in bulk Prices in bulk → augment solr
  - o Pm press, Bookshop

# Commitments

## Mek Karpeles

1. **Open Library - Adding Book Data (Infogami & Solr)**
  - ☐ Add field to Edit UI for reading\_level [grade](#) → infogami
  - ☐ Adding reading\_level field → solr
  - ☐ Adding ia\_wishlist → solr
  - ☒ ~~OSP co-assignments → solr (if time permits)~~
2. **APIs**
  - ☐ Open Library OPDS endpoint (+ petabox)
  - ☐ Tokens for scoping [Book Availability](#)
  - ☐ Update the Books API Docs for v2 (announce a date for deprecation, EOL for books.json) + give docs a refresh
3. **Design Research**
  - ☒ ~~Develop Personas → Identify 5 key collections~~
4. **[Site Performance \(Epic #4485\)](#)**
  - ☐ [Books Page: Get Best Edition](#) → [Solr #7452](#)
5. **[Continuous import channels & data sources \[epic\]](#)**
  - ☐ [ISBNdb/OL Overlap #7658](#) Charles Horn
  - ☐ Add @username to batch\_import db?
  - ☒ ~~Daily covers archival cron → archive all existing items? (tar v. zip?)~~
    - ☐ Migrate covers DB + archive items to use .zip instead of .tar
  - ☐ Import BWB Monthly [Covers](#) Mek Karpeles [scottreidbarnes@gmail.com](#)
    - ☐ Monthly BWB covers need to be archived to archive.org (along with BWB monthly cron)
    - ☐ Cover import process that gets kicked off via BWB monthly cron
    - ☐ Some cron for archiving covers (from ol-covers0 disk) so we don't break everything
6. **Covers**
  - ☒ [Investigate migrating cover tars → zips - Issue #7478](#) Making sure we have daily archival → zips on archive.org (pair w/ mek)
7. **[Canonical Tags \(Epic #2819\)](#)**
8. **Librarians, Development & Communications**
  - ☒ ~~Develop process for communications contributions~~
  - ☐ Develop process for librarian collections contributions
  - ☐ Collaborate with the Development team to secure resources for IASoC or design research.

1. **Open Library - Adding Book Data (Infogami & Solr)**

- ☐ Add field to Edit UI for Summary / Synopsis (hide spoilers) → infogami
- ☐ [Add ia\\_wishlist boolean tag to solr search engine · Issue #6187](#)
- ☐ [Create a UI field \(list\) for Fatcat + Wikipedia Citations](#) mark@archive.org + Bryan Newbold ?

2. **Site Performance (Epic #4485)**

- ☐ [ReCaptcha: Only load where needed #7453](#) + invisible
- ☐ Partials -> .json (gzip)? Async components? [Issue #7447](#)
- ☒ [Cache Patron's Active Loans · Issue #7591](#)

3. **APIs**

- ☐ Light survey on the existing API docs page (form) to collect more intel on needs ([issue #7576](#)) → banner on api docs page, similar to the my books YRG banner

4. **Imports**

- ☒ [\(H1 2023-02\) Updated Promise Item Cron](#)
  - ☒ [Ensure books have bwbsku/barcode #7425](#)
  - ☒ [180k books have multiple bwb local ids #7430](#)
- ☐ Improve Author & Work resolution:
  - [Don't create duplicate author records for M. <surname> with exact date match · Issue #7349](#)
- ☐ Improve Centralize Import Validation (epic) → at least create issue + document
  - Have more specific errors for service-level failure (e.g. timeout, 503)
  - Fail on Future dates (unless overridden)
  - Block imports for works w/ 100+ editions (if e.g. independently published)
  - Ensure a modification only occurs in catalog if some additive data (unless overridden, e.g. bwb\_sku). Do not add source\_record if no other data changed (do not only add source\_record) → exception is promise items always need source\_record
- ☐ Extending imports API to allow patrons to submit batch imports
  - Specify a batch\_name + give chunks of data, queue in import table (for accounts with API access); extend table to have 'submitter'; adding revision number if we have it back to table
- ☐ A MARC Import pipeline using batches for Charles Horn (store MARCs in IA → OL)
- ☐ Reboot [/admin/imports](#) → public, performant, show more (epic)
  - [import throughout & performance](#). Audit performance and explore haproxy, multi-threading, separate db table for pending v. completed imports
  - Show aggregate 'status' per batch
  - Ability to paste a json obj and either validate or import
  - *Ability to fetch data by ID (e.g. librivox, archive.org) to synthesize import record*
  - **Local Imports / Developer mode.** Design import pipeline to run on dev instances to enable better testing + community contributions. ([issue #7578](#))
  - Fix slow-loading [/admin/imports](#) page ([issue #7577](#))

5. **Yearly Reading Goals (epic)**

- ☒ [\(H2\) Yearly Read Lists: A way patrons can see books they've read in year /account/books/already-read/2023 \(issue #7379\)](#)
  - Bonus: Could use existing List preview from yearly check-ins e.g.

6. **Canonical Tags (Epic #2819) & Lists**

- ☒ [Content Warning tags: Add "Safe For Work" Mode to blur NSWF covers · Issue #7115](#)

7. **Onboarding Q: [Adds "Safe Mode" to blur NSWF covers #7416](#) Robert Wu? jchamp@archive.org**

1. **Bookreader**
2. **Book Series** – [Issue #6718 · internetarchive/openlibrary - Book Series](#)
3. **Open Library - Adding Book Data (Infogami & Solr)**
  - ☒ [Solr should including Ratings stats – Issue #3990](#)
  - ☒ [Add Book Reading Log counts into Search Engine – Issue #7321](#)
  - ☐ **Reinstate donate button**
  - ☐ Listener for changes to ratings, reading log counts?
4. **BookServer**
  - ☐ Having opds “providers” infogami field indexed in solr – needs discussion (may not be worth it?)
5. **Site Performance (Epic #4485)**
  - ☐ [Book Page: Other Editions Table → Solr #7451](#)
  - ☐ [Bulk Book Availability in Solr #7450](#)
6. **APIs**
  - ☒ [Adding leen, oclc, bwb\\_sku \(+ all local\\_ids\) to solr for scanning centers](#)
  - ☐ A canonical v2 Book API endpoint for ~editions, likely using the search.json endpoint (what data is needed, etc) + [Mek Karpeles](#) on assist
7. **Imports**
  - ☐ Investigate auto-work-merge opportunities
  - ☐ Archive.org → OpenLibrary imports [scottreidbarnes@gmail.com](#) + [drini@archive.org](#)
    - ☐ Fixing pipeline [drini@archive.org](#)
    - ☐ Legacy imports (improving the query) [scottreidbarnes@gmail.com](#)
  - ☐ *Standardized Import Template/class for data sources* [drini@archive.org](#)
8. **Canonical Tags (Epic #2819)**
  - ☐ An effort to survey + dedupe/merge subjects [t3ohjingxiang@gmail.com](#) (case insensitivity)



cclauss@archive.org

1. Fix Affiliate-server cclauss@archive.org
  - ☒ [Affiliate Server: Data Fetch Thread Dies - Issue #7569](#) → [Make amazon lookup thread a daemon #7601](#)
2. **Open Library: Dev-Ops (Provisioning & Deployment)**
  - ☒ Fix docker compose → docker compose
  - ☒ **Complete ol-www1 → ol-www0**
    - ☒ Fix monthly sitemap rsync
    - ☐ Reclaiming trusty ol-home (which has been replaced by ol-home0)
      - We need to get `solr\_stats` off ol-home → ol-home0 & update cron so ol-www uses ol-home0
  - ☒ **Patch holes in reporting** (affiliate\_server, covers, cron jobs)
  - ☐ **Begin migrating db from trusty**
    - ☐ **Create provisioning guide (based on 2020 provisioning scripts)** → <https://github.com/internetarchive/openlibrary/wiki/Production-Service-Architecture>
    - ☐ Spin up a new server for ol-db\_new w/ docker + self-contained Open Library
    - ☐ Attempt a dump of ol-db1 → ol-db\_new
  - ☐ **Improve deploy safety** (get running services from docker container) → docker exec -it ps ax | scripts.sh (ask ChatGPT?)
3. **Open Library - Book Covers[epic]**
  - ☐ Documentation needed for **book covers**
  - ☐ Backfill isbn10/13 in coverdb + ensure db gets correct values moving forward
  - ☐ Publish a monthly dump of covers db → item/offset? + width height, etc
4. **APIs (Lead on swagger)**
  - ☐ **Issue: MVP release plan for swagger (not an epic, specifically what the 1st release will look like and a checklist – even if all the work is already done and just has to be enacted; it would be good to have issue to close, put on milestone, + record & debate decision.**
  - ☐ <https://github.com/internetarchive/openlibrary/issues/5914> Formal documentation of our existing & v2 APIs (e.g. swagger). Per service, but *also*, spanning services (json index)
5. **Imports**
  - ☒ Fix affiliate-server dead threads
  - ☒ Fix amazon imports with pre-check + statsd
  - ☒ Why is amazon importing insubstantial records? (i.e. source\_record only)
  - ☒ Catalog Sentry instrumentation (check if import system using sentry); may be done
  - ☐ Update Amazon API code to pull author ID (lower priority)
  - ☐ **Extend affiliate server to hit + import more sources · Issue #7513**

scottreidbarnes@gmail.com

- Petabox Py2 → Py3 audit (petabox)
- Covers
- Imports
  - **Catalog Import DRY Refactor: A basic commonsense DRY / refactor of the Catalog add\_book import pipeline w/ a few tests – emphasis is on making code pieces testable, removing duplicate logic, adding docs, understanding what's going on: document here** <https://github.com/internetarchive/openlibrary/wiki/Developer's-Guide-to-Data-Importing>

# Roadmap

2023-01

- Yearly Planning
- Team Reviews

2023-02

- Team Reviews
- In solr

2023-10

- Library Leaders Forum (issue needed)
- Community celebration (issue needed)

2023-11

- Board Report & presentation (issue needed)

2023-12

- Yearly Planning (issue needed)

## NOW: What's in flight?

This table only includes contributors/fellows who have been with us >1mo and contribute 5h+ a week: ~15 contributors

Contributor	Efforts In Progress (limit to 3)	Notes & Considerations
Mek	<ul style="list-style-type: none"><li>Strategic Planning</li><li>Tags</li></ul>	
Chris Clauss	<ul style="list-style-type: none"><li>Fixing <a href="#">sitemaps</a></li><li>Retiring ol-home &amp; ol-www</li></ul>	
Drini Cami	<ul style="list-style-type: none"><li>DWWI: Work-level Synonyms (e.g. xisbn)</li><li>Implementing <a href="#">Book Series</a></li></ul>	
Jim Champ	<ul style="list-style-type: none"><li>Releasing <a href="#">mobile Books Page</a></li><li>API: Letting Archive.org update + pull from Open Library</li></ul>	
Scott Barnes	<ul style="list-style-type: none"><li><a href="#">Petabox py2→py3 audit</a></li></ul>	
Jayden	<ul style="list-style-type: none"><li><a href="#">GSoC Applicant</a></li><li><a href="#">Canonical Tags</a></li><li><a href="#">Site Performance</a></li></ul>	2023 GSoC Hopeful
Mark Heiman	<ul style="list-style-type: none"><li>Google sign-in for unofficial mobile Open Library Reader app</li><li><a href="#">Bulk Editing</a> and <a href="#">Tagging</a> selected books</li></ul>	Book Fair obligations this month
Design Team		
Ruijingya (Fellow)	<ul style="list-style-type: none"><li>Registration redesign</li><li>Formalizing <a href="#">design system</a></li></ul>	
Debbie (Fellow)	<ul style="list-style-type: none"><li><a href="#">Team Page design</a></li><li>Standardizing <a href="#">My Books Component</a></li></ul>	w/ Dana on Design
Abbey Ripstra	<ul style="list-style-type: none"><li><a href="#">Persona research</a></li></ul>	Coordinating w/ Ruijingya
Jasmine Connor	<ul style="list-style-type: none"><li><a href="#">Standardizing Button Designs</a></li><li>Information Architecture</li></ul>	
Samuel Grunebaum	<ul style="list-style-type: none"><li><a href="#">My Books Mobile page</a></li></ul>	
Dana	<ul style="list-style-type: none"><li><a href="#">My Books Desktop</a> page</li><li><i>Subject Page redesign (planning)</i></li></ul>	Begins Masters program @ Tufts soon 😊
Keita	<ul style="list-style-type: none"><li><a href="#">Accessibility Plan</a></li></ul>	Expertise in Accessibility
Nick	<ul style="list-style-type: none"><li>Establishing an <a href="#">Open Library Communications strategy</a></li></ul>	
Sivan + Riya	<ul style="list-style-type: none"><li>Moderation Tags</li></ul>	
Sidarth	<ul style="list-style-type: none"><li>???</li></ul>	
Georgia		

## Menu for Contributors

### Good First Tasks

1. **APIs**
  - a. Implement a Bulk List add/create API [drini@archive.org](mailto:drini@archive.org)
2. **Front End**
  - a. Hamburger open/close on mobile [drini@archive.org](mailto:drini@archive.org)
3. **Settings**
  - a. Enable safe-for-work mode [jchamp@archive.org](mailto:jchamp@archive.org)
  - b. specify primarily book language [drini@archive.org](mailto:drini@archive.org)

### Design Tasks

- Series Page
- Series Editing Flow
- Series Integration into Books Page
- Subjects Page
- Want to Read Button (w/ Lists)
- Consolidated Borrow/read button (download options/durations)
- **Consistent UI Components.** All book cards (search, trending, my books) should have the same style/options: **Reading Log dropper button** (search vs books page)
  - consolidated read button. Standardize buttons.
- **Design System.** Develop pattern libraries for major components in Figma (Fellowship)
- **Search Design Mobile:** Fix search cards (mobile)
- **0-Results Search.** [Show fulltext results async beneath results #1011](#)
- Direct Manipulation Editing by Confirmed Editors

### Projects

1. Barcode scanner → Vue [drini@archive.org](mailto:drini@archive.org)
2. Dismissible banner [jchamp@archive.org](mailto:jchamp@archive.org)
3. Move Book Extractor™ into Open Library
4. Bulk “Add to List” widget (takes a selection of books → adds/creates list)
5. Adding facets
6. **Twitter.** Reinstate [@borrowbot](https://twitter.com/borrowbot)
7. **Covers.** investigate OL...A +
8. **OSP:** Add Learning Objectives
- [Edition merging](#) [mark.heiman@gmail.com](mailto:mark.heiman@gmail.com)
9. [Book Trails](#)

### Wins Log

- Star ratings + Reading Log counts in solr
- List API documentations improvements

- Switched to ol-www0, upgraded ol-www1
- Bulk availability upgrades

## Notes

- rayz321z@gmail.com
  - Managing Proposals for Project Management
- @tmanarl
  - 1:1 to discuss "Librarian Tags" Project
- @Scott Barnes
  - Trusted Book Providers? Librivox?
- Scott Hewlitt
  - K-12 Library
- Connor Forsyth
  - Information Architecture
- Jasmine
  - Design System: Buttons & Information Architecture
- Keita
  - Accessibility
- Jaye
  - Team Page
- Kira
  - ...
- **Riya + Sivan:** Hamburger open/close on mobile drini@archive.org
  -
- **Joanna:** Lead on Barcode Scanner? Barcode Scanner → vue
- Jayden T → Singapore, GSoC drini@archive.org ?
  - Widget for Create List / Bulk Add to List?
- **BookServer**
  - Enable publishers to import their catalogs & list their books for sale (esp. stripe/paypal)
- **Design Research**
  - We should better understand what our patrons need
- **Collections**
  - We need to better connect Librarians with curation tasks our patrons need
  - Open Library should prioritize academic use cases and also serve pleasure readers
  - The website should have more useful collections & subject search
- **Quality & Usefulness**
  - The website should have useful book pages
  - The website should be faster
  - The website should have cleaner metadata
  - The website should have some targeted way of getting good missing books
- The website should have more web books
- The website should have series
- Visiting a book page should enhance it
- The book page should be editable in place
- The website should be more tailorable (onboarding)
- We should enable others to submit batch-import



1. Prioritize some tasks that will help the Librarian, Communications, & Developers teams work more autonomously @ better utilization
  - a. Getting batch imports working locally → enable developers to help with bulk imports
  - b. Trusted Book Providers; if there's a way to add support for providers that doesn't require Drini to be lead → document & lead?
  - c. Bulk List API ← Jayden
    - i. Turn barcode scanner into list
    - ii. Take extractor turn into list
2. Can we find a way to leverage leads to do issue triaging, prioritization, organization

## (old) Open Library: 1-off Design Efforts 2022 H2

### Topics

- CTA Button Default Action (check libraries) Mek Karpeles [#1184](#)
- ~~Standardize the Waitlist btn (blue, hide text in carousels) – Mek Karpeles [#6652](#)~~
- Main Site Navigation Redesign [#6681](#) Dana, **Lead:** Mek Karpeles
- **My Books** Experience <https://openlibrary.org/account/books> **needs issue** Lead for Epic: Mek Karpeles
  - Nav sidebar should appear on settings pages and profile page (there's already an issue somewhere) – **Rodrigo?** Lead: [jchamp@archive.org](mailto:jchamp@archive.org)
  - Proposing a Summary Page Lead: Mek Karpeles
  - Cleanup of Loans + Holds **Rodrigo**
  - Redesign Navigation on My Books (mobile) **Hayoon**, Lead: [drini@archive.org](mailto:drini@archive.org) **create issue**
    - Treat desktop nav as the contents of a dropdown menu
      - May hide too much
    - Have a horizontal navigation (e.g. Books Page) w/ Headers
      - May require 2 levels of navigation / complexity (for patron)
      - More require more design / thought
    - 2 components: 1 is horizontal nav, 2nd is accordion for relevant section
- ~~Search Page Navigation – **Rodrigo?** Lead: [drini@archive.org](mailto:drini@archive.org)~~
- Search Cards (mobile) [#2152](#) Lead: [drini@archive.org](mailto:drini@archive.org)
  - Reading Log, star ratings
- Unified Type Search UI Exploration **needs issue** Lead: [drini@archive.org](mailto:drini@archive.org)

# 2023 H2 Review

Jul 7, 2023

## 2023 Priorities

5. **Enable librarians** ASAP to create [Series](#) & [Tags](#) + select & manage in [Bulk Editor](#)
6. **Edu collections** for learners using Subjects, [Tags](#) – requires [design research](#)
7. **Fundamentals:** [Performance](#), [Provisioning](#), [Imports](#) + [Covers](#), Core Design
8. **Futureproof:** Safeguard [5-year strategy](#) of OL; **BookServer**, **TBP**

## 2023 H1 Wins

- **Stats**
  - Community involvement way up :)
  - +1.5M registered patrons ([equal to the entirety of last year](#))
- **Design**
  - [Lending Survey](#)
- Books Page redesign
- My Books mobile Page
- Team's Page
- Ability to mark account as SFW
- Librarian Merge Queue
  - Bulk-Undo Work Merges
  - Proposing Author Merges
  - Author ID now appears with correct order
  - Doesn't break when lists not loading
  - Column order
- ILE
  - Select books across pages
  - "Edit in bulk" (in progress)
- Search
  - Ratings & Reading Log available in Search
  - Diacritic insensitive
- Added Tag Types
- New Librarian documentation (Onno)
- Covers Archival
- xISBN LCCN overlaps
- Ops: Ol-www and ol-home both migrated + ol-home0 solr removed 🎉
- [Yearly Reading Goals Summary](#)
- Jayden's Search Inside 5x performance improve
- Swagger Open API docs live
- Designs proposed for [Account Registration #7647](#)

## 2023 H2 Planning

**To Discuss:** BookServer? [Performance](#)? [Imports](#)?

abbey.ripstra@gmail.com + ruijingya.tang@gmail.com

- Design Research

Jayden [t3ohjingxiang@gmail.com](mailto:t3ohjingxiang@gmail.com)

- Tags & Subjects – Roya, Dana, Lisa
  - Better intersecting relevant subject faceting? Mentor: [drini@archive.org](mailto:drini@archive.org)
- Site Performance
- [Bulk Editor](#) (Metadata & Tags) – [mark.heiman@gmail.com](mailto:mark.heiman@gmail.com)

[cclauss@archive.org](mailto:cclauss@archive.org)

- Resilience & Failover for VMs + Live-migration #8043 [cclauss@archive.org](mailto:cclauss@archive.org)
- Move off Trusty – Db (requires us to investigate migrating from one postgres to another)
  - Trusty hosts: ol-db1, ol-db2, ol-home, ol-backup0

[scott.barnes@archive.org](mailto:scott.barnes@archive.org)

- Imports – Refactoring + Testing

[jchamp@archive.org](mailto:jchamp@archive.org)

- Design Components (Droppers)
- **Site Performance** (Jayden + Jim)

Mek Karpeles

- Covers
- OBGp Copyright + Extracting TOC (from Drini) into pipeline w/ Jake & Jude
- Design Research
- Tags
- Re-enable Amazon in a way that doesn't suck™

[drini@archive.org](mailto:drini@archive.org)

- Series
- Trusted Book Provider
- Extracting TOC (LLM) → OL Book Pages
- V2 API – Work + Edition Search to include **Author** (image, metadata)

## Ideas / Wishlist

- OBGp + AI + Books + Book Pages?
  - Executive Summaries? Index? Entities / topics? OCR correction?
  - Suggestions engine? Tell us criteria (include vampires and comedy) or other similar books.
- Search & APIs
  - Reading Levels?
  - Open Syllabus?
  - /type/pages?
- Book Donations, BookServer, ISBNdb