Sustaining an Open Scholarly Ecosystem with Community-Based Open Source

ELPUB 2019
Robert Cartolano
June 3, 2019
Welcome and Thanks - ELPUB 2019
Sustain... Over Time...

Marseille: Braun, Georg, 1540 or 1541-1622
https://digital.tcl.sc.edu/digital/collection/braunhogen/id/303
Sustaining an Open Scholarly Ecosystem with Community-Based Open Source
Sustain

- Definition: “use and stewardship of resources today that preserves them for tomorrow”

- Components:
  - Economy - management, or stewardship, of the resources;
  - Ecology - relationship of the community with its environment
  - Equity - fairness to all

- ALA Sustainability and Libraries,
  https://libguides.ala.org/SustainableLibraries
“...free to access, use, modify, and share”

- Ensure content, technology and service can be sustained
- Avoid “traps” that impede economy, ecology, equity
- Support bibliodiversity
- Minimize conservation, remediation, migration costs
“Scholarly communication is the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use.”

- Principles and Strategies for the Reform of Scholarly Communication
  http://www.ala.org/acrl/publications/whitepapers/principlesstrategies
Community-Based

“...larger community of users, programmers, administrators, governing agencies, and sponsors are involved in setting development priorities, providing user support, fixing bugs, defining policies, encouraging adoption, and otherwise maintaining a viable product.”

- It Takes a Village: Open Source Software Sustainability
  https://doi.org/10.7916/D89G70BS
“Open source software is made by many people and distributed under an OSD-compliant license which grants all the rights to use, study, change, and share the software in modified and unmodified form. Software freedom is essential to enabling community development of open source software.”

- Open Source Initiative, https://opensource.org/
Sustaining an Open Scholarly Ecosystem with Community-Based Open Source

WHY?

...to better serve our faculty, students, researchers and society - now and in future
Foster Sustainability

- Sustain access to information for the long-term
- Foster discovery for reuse across disciplines
- Support scientific reproducibility
- Conserve, reformat and migrate over time
Foster Sustainability

- Evolving roles
- New approaches for new forms of scholarship
- Need for increased collaboration at scale
UC and Cambridge University Press agree to open access publishing deal

The University of California and Cambridge University Press have entered into an agreement to advance the global shift toward an open access future for research.

Open Access books on JSTOR

PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM).
Content = Service?
Open Access to content is not enough!
It Takes a Village - Open Source Sustainability

Three Phases:
- Getting started,
- Growing
- Stable but not static.

It Takes a Village: https://doi.org/10.7916/D89G70BS
Content + Open Source = Service?
Open Access and Open Source Are Not Enough!
We Need Five Opens

- Open Content - access, re-use, bibliodiversity
- Open Source - documented, re-usable
- Open Format - standards-based, facilitate use, migration
- Open Protocols - standard APIs, network protocols
- Open Community - diverse, representative, inclusive
“...systems, structures, and collaborative networks necessary to support the knowledge ecosystem have reached a level of complexity and scale that has overwhelmed the ability of existing methods...”
Content +
Open Source +
Open Standards +
= Sustainable Service?
Era of Rapid Change

“Library clients have come to demand instant access, online, to all forms of academic information content….These broad trends have been seen as presaging a fundamental reinvention of the academic library”

- The Evolving Role of Libraries in the Scholarly Ecosystem, 2012, Webster, Keith
Educopia -
Community Cultivation

Key Growth Areas:

● Vision
● Infrastructure
● Engagement
● Finances and HR
● Governance

- Community Cultivation – A Field Guide
  https://educopia.org/cultivation/
Content +
Open Source +
Open Standards +
Community-Based
= Sustainable Service

Is it enough?
“The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge.”

- **Sustain** - membership, governance, preservation
- **Open** - open standards, open source
- **Open content** - 6 million full view Ebooks out of 17 million
- **Community-Based** - 140 members
Coalition of Open Access Repositories

Vision - “A sustainable, inclusive, and trusted global knowledge commons based on a network of open access digital repositories”

- **Sustain** - international association, governance
- **Open** - documents, best practices, advocacy
- **Community-Based** - 140 members, 6 partners
Readium

“The fundamental goal of the Readium project is to produce a set of robust, performant, spec-compliant reading system toolkits that support digital publishing formats (e.g. EPUB, Web Publications etc.) and can be deployed in browsers or built into native apps on iOS, Android or the desktop.”

- **Sustain** - membership model, governance
- **Open** - open source, open standards, open protocols
- **Community-based** - 40 participating members
OpenEdition is run by the Centre for Open Electronic Publishing (Cléo), a public and non-profit initiative supported by prestigious research organizations. Its overall aim is to promote open access digital publishing.

- Sustain - membership, governance
- Open - ebooks, journals, 150K articles
- Community-Based - steering committee, scientific council, user committee
LYRASIS and DuraSpace Merger

“This merger will form a larger organization that is committed to advancing services, technologies and innovations for collections holding organizations worldwide, and consciously building communities to increase our collective impact and support programs that benefit the wider field.”

- **Sustain** - membership, governance, services
- **Open** - code, standards, practices
- **Community-Based** - 1,000+ members
LYRASIS and DuraSpace Merger

- Drive scalable change, new technologies and services
- Serving 1,000+ libraries, museums and archives
- Durable, persistent access to data and services

Fedora™

VIVO

collection space

ArchivesSpace

CASRAI

LYRASIS Islandora Hosting Services

SimplyE
Sustaining an Open Scholarly Ecosystem with Community-Based Open Source

New Challenge

The Academic Ebook
Can We Imagine a Better Academic Ebook?
Can We Imagine a Better Academic Ebook?

- A better academic ebook experience for faculty, students, researchers, society
- Open Technology Ebook Ecosystem - sustainable, community-based open source, standards
- Collaborate, at scale, at international level
What are Academic Ebooks?

- **Licensed Ebooks** - University presses, academic publishers (e.g. Columbia University Press, University of Michigan Press, MIT Press, NYU Press), including enhanced academic ebooks
- **Digitized Collections** - Library efforts to preserve and provide access to collections - aggregated at HathiTrust, Internet Archive, etc.
- **Open Access Academic Ebooks** - Knowledge Unlatched, University of Michigan Press, OAPen.org, etc.
Current State of Academic Ebooks at Columbia

- 3 million Ebooks at Columbia (see tinyurl.com/clio-online-books)
- Inconsistent user experience - multiple vendors, interfaces, technology
- Vendors developed proprietary web interfaces, proprietary mobile apps
- Licensed and open content is not available via one interface
- Content is fragmented across multiple formats, DRM and non-DRM

We need to do better!
Every dollar we spend on multiple, overlapping vendor technology platforms is a dollar not spent on building our collections.
Academic Ebook “Essentials”

1. Common Method of Access - simple, consistent, one login, mobile equal
2. Local Discoverability - direct link in catalog with high-quality metadata
3. Search - Download - Read - entire book in very few steps
4. Great Reading Experience - fast page turns, simple UI, search, TOC, index
5. Scholar Tools - citation, portable annotation
6. Library Branding, Reporting, Administration
7. Stop Proprietary Mobile App Proliferation

http://www.readersfirst.org/
Accelerating Towards Open

- $685K IMLS Grant for SimplyE Academic
- $1.8M IMLS Grant - NYPL for SimplyE
- Readium LCP - Proposed ISO Standard

2015

- $786K Mellon Grant for NYU Enhanced OA Monographs

2016

- Library Simplified: The Open Source Library Platform

2017

- OPDS 1.2 Release w/Library Lending Support

2018

- DPLA: $1.5M Sloan Grant to Move DPLA Exchange into Production
Alfred P. Sloan-Funded Feasibility Study

● Tasks:
  ○ **Survey:** Winter 2019 - collected survey data from 150+ libraries
  ○ **Research needs:** of faculty, students, libraries, publishers, aggregators
    ■ Conducted extensive conversations with academic community stakeholders
    ■ Identify use cases, stories, product and market scope
    ■ Meet with community participants to test and refine concepts
  ○ **Develop Proof of Concept:** as technology demonstrator, collect feedback
  ○ **Develop Technology & Community roadmap:** for SimplyE platform to meet academic use cases
  ○ **Develop Report & Plan:** Present recommendations, pilot plan, business plan, governance model

● **Co-Investigators:** Robert Miller, LYRASIS, Robert Cartolano, Columbia University Libraries
● **Steering Group:** LYRASIS, Columbia University, New York University, New York Public Library, Digital Public Library of America, MIT Press, Minitex/University of Minnesota
● **Project Consultant:** James English, Product Strategist, LYRASIS
Survey Results - Media Type & Access Support

Question: Ranked Very Important or Most Important

- 78% (101/131) - View Offline Entire Book PDF
- 73% (95/131) - View Offline Chapter-by-Chapter PDF
- 27% (34/131) - View Offline Entire Book EPUB
- 19% (24/131) - View Offline Chapter-by-Chapter EPUB
- 8% (9/131) - Listen (download/stream/sideload) Audio Books
- 2% (2/131) - View Offline Other
Survey Results - User Experience - Top Features

Question: Select five most frequently used features by your academic ebook readers:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy, paste and print text</td>
<td>81%</td>
</tr>
<tr>
<td>In-document keyword search</td>
<td>73%</td>
</tr>
<tr>
<td>Table of Contents navigation</td>
<td>66%</td>
</tr>
<tr>
<td>Citation creation and export into citation management system (e.g., Endnote, Zotero, etc.)</td>
<td>60%</td>
</tr>
</tbody>
</table>

Most important features that represent USE of ebook as a source of information.
Survey Results - Demographic Needs

Question: Select which groups most need simplified ebooks access. Check all that apply.

Opportunity: These results suggest that while digital natives are familiar with computers, the “tools of research” are not native to their digital upbringing as they were likely informed by “consumer” technology.
Survey Results - Device/Platform Requirements

**Question:** Indicated devices used for academic ebooks. Check all that apply.

<table>
<thead>
<tr>
<th>Device Type</th>
<th>Percentage</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop/Laptop - Macintosh and/or Windows</td>
<td>99%</td>
<td>126</td>
</tr>
<tr>
<td>Tablet - iOS or Android Tablet</td>
<td>93%</td>
<td>118</td>
</tr>
<tr>
<td>Smart Phone - iOS and Android</td>
<td>92%</td>
<td>117</td>
</tr>
<tr>
<td>ChromeBook / Netbook</td>
<td>57%</td>
<td>73</td>
</tr>
<tr>
<td>Kindle Paperwhite (or other e-ink device)</td>
<td>32%</td>
<td>41</td>
</tr>
</tbody>
</table>

**Opportunity:** Clearly, the traditional computer now shares its importance in academia in the new age of tablets and smartphones.
Survey Results - Community-Based Open Source

**Question:** Importance of open source applications as part of your technology/service infrastructure.

- Essential & Important: 62% (73)
- Neither important or unimportant: 28% (33)
- Don’t use open source software & Not important: 9% (11)

**Question:** How your institution implements open source solutions.

- A combination of externally and internally hosted platforms: 60% (69)
- Only hosted by external provider(s): 21% (24)
- Not applicable - we don’t use open source software platforms and applications: 16% (18)
Survey Results - Open Source Ebook Delivery

Question: Under what conditions would you use open source ebook delivery platform. Check all that apply

- It would improve my current ebook service experience (86%)
- It would help to change the status quo of academic ebook services (63%)
- It would reduce my cost of service (57%)
- It would augment my current systems (57%)
- It would meet my minimum/general needs (39%)
Develop Proof of Concept using SimplyE

- **Sustain** - explore with LYRASIS
- **Open Source** - open source - SimplyE mobile app, Library Simplified, Circulation Manager, Readium
- **Open Standards** - EPUB, PDF, OPDS, LCP (proposed ISO standard)
- **Community-based** - SimplyE Libraries, Reader’s First, library consortia
Vision: 1) Search
Vision: 2) Borrow/Download
Vision: 3) Read on Web, Tablet & Phone
Provide a Great Reading Experience
Synchronize Bookmarks & Reading Position on All Devices

Smart Phones

Tablets

Web Browser
Next Steps

- **Sustain**
  - Explore next phase of grant funding
  - Develop growth, sustainability and business plan

- **Open**
  - Enable and foster open technology ecosystem for academic ebooks
  - Improve academic reading environment leveraging proven open technologies
  - Contribute to existing open source efforts to support academic needs

- **Community-Based**
  - Develop community governance models and community participation
  - Develop service and membership models to foster adoption and engagement

- **Scholarly Ecosystem** - connect ebook providers to academic libraries
  - University Presses, Publishers
  - Digitized Collections
  - Open Access Ebooks
Next Steps: Develop Academic E-Reader Platform

Back End
Content host
- Local Collections
- Consortia (Minitex, Amigos)
- Ebook Providers (JSTOR, Muse)
- Open (e.g. HathiTrust, IA)

Middleware
- Open Source Library Simplified Circulation Manager
- Hosted at Local Institution, Consortium, LYRASIS or other Service Provider

Institution Authentication
- CAS/SAML, LDAP, ILS, etc. (AuthN/AuthZ)

Front end - Discovery, Reading
Users discover content in library catalog, landing pages, reading apps

- Library Discovery
  - Blacklight, VuFind, OPAC, etc.
  - OPDS URI
  - Pluggable components Discovery Interfaces

- Mobile Academic EReader
- OPDS 2.0
- OPDS or API
- LCP, RWPM, Annotation, Deep Search

Web Reader
- OPDS or API
Can We Sustain an Open Scholarly Ecosystem with Community-Based Open Source

Yes!

Let’s do it together!
Discussion