UKSG

Start-up Story
New advances in open source infrastructure support: accelerated book digitization with Editoria

How can open source infrastructure support a modernized, accelerated book production workflow? The California Digital Library, the University of California Press and the Collaborative Knowledge Foundation collaborated to design a new platform – Editoria – to do exactly this, following a new user-driven design method to result in a simple, people-centric interface. This case study details the main problem facing publishers who are restrained by outdated, print-oriented production platforms, the 'reimagining' exercise and the iterative design process that has resulted in new technology which can be adopted, adapted and integrated by publishers.

CLARE DEAN

Independent
consultant and
Outreach Manager
for Metadata 2020

Introduction

The development of community-driven open source software is rapidly gaining momentum in the scholarly publishing industry. The joint development of innovative open source tools among those publishers working with the Collaborative Knowledge Foundation (Coko)[1] is providing the proof of concept needed for organizations to understand how the incorporation of this technology fits into their strategies for digital sustainability.

Coko staff observed that participants at 2018 meetings, including the Society for Scholarly Publishing Conference, the Library Publishing Forum and the Association for University Presses Annual Meeting, were increasingly interested in the idea of open source as a viable option for publishing infrastructure. Concerns about proprietary software and commercial ownership of scholarly infrastructure is inspiring greater curiosity about the prevalence of open source solutions on the wider web. This in turn has resulted in the realization that industries have been able to transform themselves through open source.[2]

The reimagining begins

Two years ago the California Digital Library (CDL) and the University of California Press (UCP), together with the Coko Foundation, discussed the hurdles of book publishing, particularly from the university press perspective. Cindy Fulton, Senior Editor at UCP, noted:

> 'A problem that has vexed us for quite a long time is that of using proprietary products from other companies like Microsoft Word, where we are constantly chasing the version control issue. We work from one version of Microsoft Word; we might have copy editors working on other versions, we might have authors working on other versions.'[3]

Key issues with current book publishing technology included:

- **document conversion challenges**
  Publishers historically lack the in-house expertise necessary to convert between .docx or latex and web-ready formats like HTML, XML, EPUB and PDF. Many engage offshore vendors who, in order to protect their business, use workflows and systems that are

'increasingly interested in the idea of open source as a viable option for publishing infrastructure'

opaque to publishers, representing a loss of control. Also, third parties in workflows sometimes introduce errors which require additional publisher-side quality checking.

- **_difficulties with offline collaboration between reviewers, authors, and production staff_**
  Moving .docx files between workflow participants as e-mail attachments, or using FTP technology, is time-consuming. It can also introduce version control issues.

- **_workflow organizational challenges_**
  Manual workflows and offline workarounds cost time within organizations.

- **_need for faster book digitization_**
  Open access monographs demand urgent dissemination. Locking urgent content in to a year-long workflow is not helpful to authors, researchers or anyone associated with scholarship.

- **_rising costs_**
  Vendor workflows and manual workarounds are costly. Vendors are successful when they protect their business through reduced transparency, but this costs publishers directly and indirectly.

  > 'When you think about the tools that are used for scholarly publishing or any kind of book publishing really you're looking at tools that for the most part were developed 35 or 40 years ago and continue to be used today to support those workflows.'

  Erich van Rijn, Assistant Director, Director of Publishing Operations, UCP

There was fatigue around vendor lock-in deals for rigid, flawed systems, and UCP and CDL realized they needed a more elegant and customizable solution. Together, they applied for a grant from the Mellon Foundation to support further investigation into solving these problems in a robust and scalable way. This reimagining of an ideal books workflow solution led to the birth of the Editoria book publishing platform. The partners envisioned that Editoria could be adopted in its original form or developed and customized as an individual modular component-based system, and would be configurable and interoperable with components in the wider technology ecosystem (including other Coko components). UCP and CDL's experience as monograph publishers meant that the team was able to develop Editoria with a deep understanding of the practical needs of users.

'reimagining of an ideal books workflow solution led to the birth of the Editoria book publishing platform'

The process: workflow sprints

The early work on Editoria at the beginning of 2017 was carried out in workflow sprints. Coko Co-Founder Adam Hyde adopted this methodology after observing that many software developers, while consulting system end-users to some extent, develop systems without consulting the vision and knowledge of those who use the software on a daily basis. Having recognized a need for a more user-led process of design, and after exploring a number of different design workflow methods, Adam adapted his own 'Book Sprint' methodology, a process through which a group is able to produce a book from start to finish in five days.

During the first workflow sprint, Cindy Fulton and Kate Warne from UCP:

- described and mapped their current books production workflow

- described their ideal workflow

- collectively modified this vision

- mapped and added layers of detail

- finalized their vision for a new system.

It was only after this point that software developers were involved to build the system and actualize the vision.

Editoria was built using the same modular 'PubSweet' framework as the rest of Coko's technology. Each segment is interchangeable, and an organization is able to configure variations of different components to suit their exact needs. Each organization builds its own custom platform, and then contributes the components (as code) back into the Coko community. This allows publishers either to adopt a technology already built in its entirety, or to adapt and create a modified version to suit their own specific needs. This iterative design process, in which components of the tool are modified and remodelled after testing, is a crucial aspect of the success of the developments to which it is applied.[4] Following this process, Editoria is not only an out-of-the-box solution (in the form in which it was built by UCP and CDL) but is also highly customizable, either by an organization alone, or in collaboration with the Coko Foundation.

> 'This iterative design process … is a crucial aspect of the success'

The solution: an end-to-end books workflow

Following development, Editoria was launched as a cloud-based end-to-end workflow, enabling publishers to:

- **convert documents efficiently**
  Converting from .docx to HTML for editing, allowing for rapid digitization in production, including instant render and export in a variety of formats such as EPUB and PDF and based on publisher-supplied CSS rules.

- **enable collaboration in the browser**
  Facilitates real-time collaboration between authors, editors, reviewers and production teams directly within the browser (see Figure 1). Eliminates delay associated with e-mailing or converting to FTP files in addition to version control concerns.

- **reduce or eliminate the need for vendors**
  Includes automation that reduces or eliminates the need for vendor intervention, including proprietary software-as-a-service and typesetting services.

Kate Warne, Managing Editor at UCP, explained what this means for publishers:

> 'Editoria can make your life easier if you are a production person. It will also require you to question some of your long-held practices and assumptions about the best ways to make books. It will provide opportunities you don't currently have to engage with the text quickly and make changes quickly to output the file formats you need without having to go to vendors; but it will require some adjustment.'
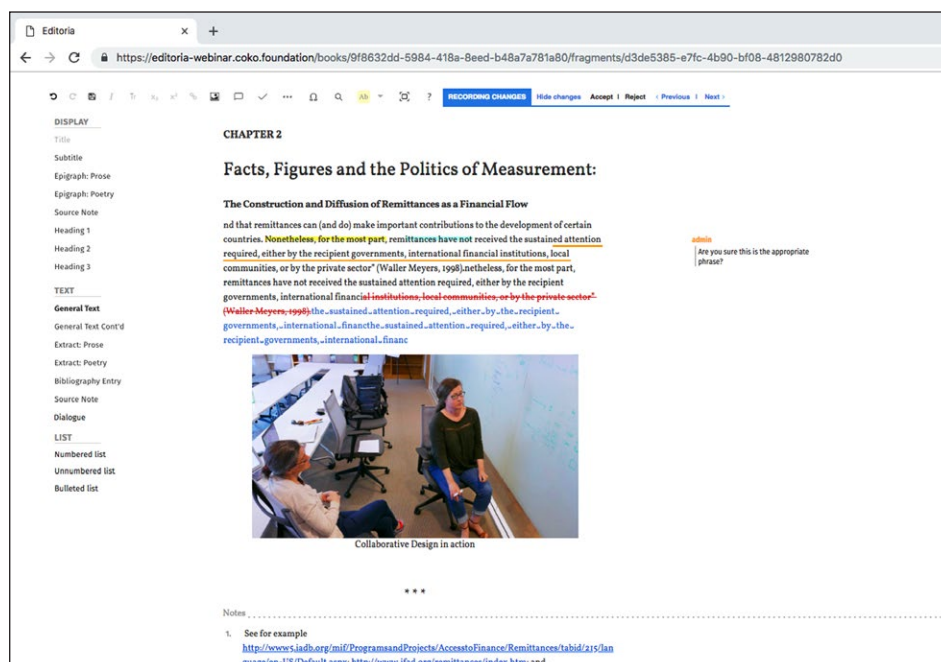

Figure 1. Editoria's interface (October 2018) showing the editor functionality, built on the Wax system

A community-driven solution

Before deploying a single line of code, Coko empowered Editoria's users to design their own software. Coko's collaborative product design process armed the end-users of the system with flip charts, markers, white boards – and undivided attention. These 'use-case specialists' tapped their deep understanding of book production to define the needs of end-users as the starting point – not the end point – of the technology build.

People-centric technology puts a relentless focus on simplicity. Rather than working with the Editoria team to map out every possible scenario, the shared goal was to solve common and simple problems without adding hardwired workflow complexity. This approach, combined with Coko's modular technology, means that as Editoria evolves, new tasks or functions can be added into the system without requiring a rebuild.

Supported by advisory board members at eight institutions, Editoria's continuing development is also informed by other publishing organizations interested in streamlining book production. Publishers, institutions and societies work with Editoria to support two main workflow categories.

'the needs of end-users as the starting point – not the end point – of the technology build'

Flat-collaborative workflow

In this collaborative authoring and editing workflow, workflow permissions take a 'back seat' and are managed socially. This workflow is in use by early adopter, Book Sprints.

'new tasks or functions can be added into the system without requiring a rebuild'

Post-acquisition workflow

In this workflow, .docx files from authors are loaded into Editoria for clean-up (styling, formatting, review) by press editors. Permissions are important in this workflow as each workflow participant plays a role and advancing a chapter's status is necessarily tied to these permissions. UCP, the American Theological Library Association and the University of North Carolina Press's Longleaf Services all employ the post-acquisition workflow.

While the adopting community will further develop and customize the solution, it is also available as an 'out-of-the-box' product as originally envisioned. After forking the openly available code on GitLab, Editoria's interface and workflow are customizable.

Conclusion: future sustainability through open source infrastructure

'Many incorrectly assume that open source software implies low quality, unsustainable, lacking in thought and effort, no reliable support, or worse. In fact, PubSweet and the platforms being built on top of it are quite the opposite: they represent close collaboration between publishing industry practitioners, high quality code and UX, and are a serious effort at tackling common industry problems by a variety of extremely experienced and talented professional developers, designers, project managers, data scientists, publishing staff, and others.'[5]

Open source infrastructure is relatively new to the publishing industry[6] and can seem risky at first. There are open source platforms and tools available, but regardless of the sophistication of the technology, publishers are not always sure how they can adopt these technologies or what resources will be needed to support their ongoing use. Coko and its partners have lowered the risk of open source by ensuring that platforms and tools are built for a wide range of uses, are fully reusable, and that there are ways to use the software without having in-house technical teams, such as through fully hosted 'turnkey' solutions.

'Coko and its partners have lowered the risk of open source'

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insight*s articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: **http://www.uksg.org/publications#aa**

Competing interests

Clare Dean is a consultant working with the Coko Foundation during the writing and publication of this article.

References

1. Collaborative Knowledge Foundation (Coko):
    https://coko.foundation/ (accessed 9 October 2018).

2. Hyde A, 6 September 2018, Guest Post: Open Source and Scholarly Publishing, The Scholarly Kitchen:
    https://scholarlykitchen.sspnet.org/2018/09/06/guest-post-open-source-and-scholarly-publishing/ (accessed 8 October 2018).

3. Quoted from a filmed interview (not included in resulting video clip) with Fulton C, Senior Editor, CUP, California, June 2018.

4. The details of the modular framework are further explained in Hyde A et al., *Pubsweet: How to Build a Publishing Platform*, published by the Coko Foundation using the Editoria platform, July 2018:
    https://coko.foundation/wp-content/uploads/2018/07/Coko_Pubsweet-screen.pdf (accessed 8 October 2018).

5. Hyde A, ref. 2.

6. Other examples of open source infrastructure used by publishers include BioOne's *Elementa: Science of the Anthropocene's* early adoption of the open source PLOS journal platform Ambra. (The journal is now being published by UCP using a different system.) There is also broad use of the open source annotation tool, Hypothesis, and web-service, WordPress. Some use COS infrastructure for preprints.

Clare Dean

Independent consultant and

Outreach Manager for Metadata 2020, US

E-mail: clare.e.dean.1@gmail.com

ORCID ID: http://orcid.org/0000-0003-3922-6077