

Driving a vision of publisher efficiency through collaboration

Kristen Ratan*

Collaborative Knowledge Foundation 2973, 16th St, Suite 300, San Francisco, CA 94111, USA

Abstract. The publishing world is weighed down by print-centric legacy workflows, siloed technology, manual processes, and vendor lock-in. It is easy for publishers to feel that meeting demands to be experimental in terms of business models, editorial processes and content types while also becoming more efficient and reducing costs is nearby impossible. Funders, institutions and researchers are asking for more transparency and openness in the publishing process and, while publishers may agree in principle, getting there poses significant operational and technological challenges.

Other industries have reinvented themselves by collaborating on shared infrastructure, acknowledging that the foundational layers of their operations are not where the competitive advantages lie. By embracing a collaborative mindset, scholarly publishers can achieve more while simultaneously reducing cost and time-to-publication. Community-built open source is one such method of building new, shared infrastructure. The Collaborative Knowledge (Coko) Foundation is leading projects to create new platform technologies and workflows that can significantly streamline publishing, giving publishers opportunities to address challenges and industry demands. Importantly, we consider what it takes for publishers to truly collaborate; and how an adjustment of mindset and structure is needed to maximize the effectiveness of a new kind of open source for scholarly communications.

Keywords: Open source, collaboration, transparency, reproducibility, scholarly communication, publishing

1. Introduction

Slow, expensive, incomplete, static, and closed - most of us in publishing are familiar with the problems our community faces surrounding research communication. The industry's legacy of print-based workflows and siloed systems have stymied agility, efficiency, innovation, and economic sustainability for the broader scholarly communication community. Larger publishers are constantly faced with cumbersome systems and interoperability challenges as they merge and acquire new companies, respond to content agility and storage demands, and introduce new products and services. Smaller publishers struggle to compete with the service offerings and economies of scale of the larger publishers. Maintaining the infrastructure silos is massively costly, keeping publishing costs artificially high and prohibiting editorial and business model innovation.

Meanwhile, many technology industries are moving to open and shared infrastructure as a strategic way to accelerate the development of new user-facing products and services. If you are not reinventing the basic components of the infrastructure you can focus on rapidly developing new features and better functionality. It's time to ask: how can the scholarly communications industry de-silo, regroup, and work

*E-mail: kristen@coko.foundation.

together to share common structures and costs? How can we work smarter to create flexible solutions for publishers that meet the demands of the industry efficiently while making way for innovation? Can shared open infrastructure increase the reliability and speed of publishing in a sustainable way? And does a shift to open source increase the utility of and ownership over a technology solution?

2. A legacy of silos

Today the publishing industry relies heavily on a limited selection of outdated technology silos that were born in the print era. This has resulted in legacy thinking about publishing workflows that prevents publishers from being able to take advantage of the modern web. For example, publishers currently use a manuscript submission system that is wholly separate from the web delivery or “hosting” system, with a gap in-between that is typically filled with a combination of homegrown solutions and production vendors. These two separate systems result in two separate content databases, two different user databases, and different access control systems, reporting systems, and more. These duplications not only increase costs and maintenance efforts, but also prohibits the development of innovative ways to connect author and reviewer behaviors with reader behaviors.

The output of most publishing processes is a static PDF and semi-static HTML page. In an era where web users are communicating in real time with dynamic content across media sharing platforms, this largely immutable article container feels out of date and limits the use and reuse of the scholarly content. The full impact of the legacy architecture and processes is widespread and difficult to quantify, from the volume of manual work required to handle editorial and production processes to the upkeep costs of the technology silos. But perhaps the most troubling consequence is largely hidden - the inability to evolve and push the boundaries of research communication.

3. Digital-era: Reproducibility, openness and speed

Publishers face weighty demands to transform as the industry shifts further away from print. A 2017 Imbue Partners study¹ reported that publishers face challenges to improve storage, metadata, content agility, discoverability, and collaboration; these changes are motivated by customer demands, the need for creation of new revenue streams, and new product opportunities.

“Their organizations expect digitization to address ever-changing and often indeterminate customer preferences, emerging competitive threats, new technology capabilities, and shrinking budgets.”² There three trends in scholarly communication are have become increasingly important: Transparency, Openness and speed.

Transparency has become a major concern in research communication, and there is a growing call for increased integrity surrounding research outputs. That is, outputs should be connected with their corresponding elements (data, protocols, code, and any other materials needed for others to reproduce or reuse the work), and these elements should also be publicly available as part of the scholarly record.

¹Imbue Partners ‘Industry Leaders’ Perspectives on the Digital Transformation Journey in Publishing’, 2017 (last accessed 12 June 2018).

²Imbue Partners ‘Industry Leaders’ Perspectives on the Digital Transformation Journey in Publishing’, 2017, p. 2, (last accessed 12 June 2018).

Better transparency also leads to better reproducibility since more context around research workflows allows others to better recreate conditions and patterns that generated the reported outputs.

Openness is related to transparency and reproducibility, but focuses more on the accessibility of the shared research outputs. Often this discussion is limited to open access of research publications, but also applies to the other elements mentioned. Regarding open access, there are more open access policies in place every year, from funder policies (e.g., the European Commission's Horizon 2020 Research and Innovation Program, the UK's Wellcome Trust, the U.S. National Institutes of Health, and private funders such as the Bill and Melinda Gates Foundation) to institutional policies (e.g. Harvard University and the University of California). Similarly data sharing policies are also becoming more commonplace, but are less common than open access policies. Policies around sharing other research outputs are even less prevalent; making these elements accessible has not been incentivized or encouraged historically.

Speed is the third trend in current research communications. The community is recognizing the importance of decreasing time-to-publication and increasing the accessibility of knowledge objects as early as possible in the interests of expedited discovery. The recent interest in preprints signals a growing concern with long publication timelines and perhaps an over-reliance on peer review as the sole gating mechanism for scholarly outputs.

4. The opportunity: Collaboration and collective action

In general, scholarly communications is becoming more aware of and more adept at evaluation and benchmarking. One example is Crossref's *Participation Reports*, an open dashboard showing the completeness of every publisher's metadata; a second is the *Make Data Count* project³, which focuses on transforming social and technical infrastructure to elevate data publication as a key research output. Publishers and service providers are gradually breaking down barriers and sharing more information. This exposure is inclining them towards collaboration to solve common problems (Manuscript Exchange Common Approach or 'MECA'⁴, FORCE11⁵, and Metadata 2020⁶ are all great examples of such collaborative efforts).

Moving forward, collaborative projects and cross-organization partnerships like these could be used to address the biggest challenges in advancing and improving the scholarly communication infrastructure. The Collaborative Knowledge Foundation (Coko) is working with eLife, the University of California Press, and Hindawi to build an open source manuscript submission and peer review system that is truly digital first, with all work done in the browser using real-time collaboration tools. This is a community-led project in which the infrastructure is not a key differentiator of publishers, but rather something that is shared, reducing costs and enabling innovation at the interface and services levels.

The partnership among Coko, eLife, UC Press, and Hindawi is one of the rare instances of publishers collaborating to build shared infrastructure. The technology created is based on Coko's PubSweet framework, a platform-building toolkit with a robust architecture and interoperable components. Using PubSweet, publishers, solution providers, and developers can build custom platforms and solutions, from authoring and collaboration tools to editing and production systems, workflow management solutions,

³<https://makedatacount.org/> (last accessed 12 June 2018)

⁴<https://www.manuscriptexchange.org/> (last accessed 12 June 2018)

⁵<https://www.force11.org> (last accessed 12 June 2018)

⁶<http://www.metadata2020.org/> (last accessed 12 June 2018)

etc. The full suite or individual modules can be adopted depending on the needs of each publisher. Components can also be added to customizable workflows. The set of PubSweet components used for journal workflows is called xPub. Components that make up a given workflow may vary, but could include an author dashboard, a peer review assignment and management component, and an interface for commenting on and reviewing manuscripts, and platforms.

5. A vision for the future

The first steps towards addressing these extensive challenges for the publishing arm of scholarly communications is to take a step back, regroup, and reconsider:

How might we replace print workflows with relevant digital approaches?

How can we increase efficiency and speed through automation?

How can we publish all of the research outputs, including data, code, and protocols?

How can we broaden access to all?

When Adam Hyde and I founded the Coko Foundation together with our advisors, these challenges were front and center in our minds. The ultimate goal is to leverage the best of the open web to enable rapid, in-browser editorial and production workflow, and to transform the output - now static, with supporting data, methods and other document ingredients inaccessible - to a constellation of connected research objects. While Coko is building technologies to meet this vision, we envision a diverse landscape with many tools and service providers contributing flexible, nimble, interoperable, and cost-efficient offerings to support an open research communication system.

5.1. *Collaboration*

True collaboration is difficult to achieve. It requires the creation of a completely new mindset and the disregard of ego and agenda to come together and work to develop something new. We are seeing a new generation of scholars naturally embrace this way of working. These students and early career researchers naturally gravitate towards shared tools and real-time collaboration tools and refuse older systems and interfaces that isolate their work. At Coko, we not only wanted to create new browser-based workflow tools, but also to infuse collaboration in the process of invention. Only through this method will we be able to create solutions truly owned by the communities we serve.

To develop these tools, Coko employs a workflow and product design process that focuses first on the needs of each of the stakeholders in the process, from authors and reviewers to in-house editorial and production staff. The resulting journal and book production platforms that we are building have easy-to-use dashboards and intelligent signaling methods to enable streamlined workflows. Each publisher can design their own workflow and have a custom solution without paying for custom software.

5.2. *Efficiency*

Platforms and tools contain much of the same infrastructure - databases, permissions systems, and transactional reporting are ubiquitous. Reproducing these components over every publisher platform is enormously costly, time-consuming, and error-prone. Modern web architecture no longer works this way due to the inefficiency of doing so. In our community, this persistent inefficiency is prohibiting publishers from focusing on business model innovation, and often drives smaller publishers and societies either out of

business, or forces them to partner with commercial presses to reduce infrastructure costs when they would rather remain independent. Through our work building open source technology solutions, publishers can share the costs of infrastructure, positioning themselves to truly innovate and deliver exemplary value to the scholarly communication community. This also enables publishers to take back control of the technology infrastructure as open source solutions are owned by the community.

6. On open source reliability

Publishers often have questions as to how open source solutions will improve upon the current systems in publishing. Concerns occasionally arise that open source is in some way less reliable than closed source.

As Coko CoFounder Adam Hyde has pointed out in his blog⁷, we all use open source technologies every day:

- The internet: TCP/IP (governing protocols), BIND (DNS resolver)
- The Web: 75% adoption of OS browsers (Chrome, Firefox), 50% of sites delivered by Apache web server, 70% of sites use WordPress, Joomla or Drupal as a CMS⁸
- Computers: Android + Apple (built on OS BSD Kernel) have 2X adoption of Windows⁹
- Phones: Android, iOS (built on OS Darwin operating system) dominate¹⁰
- Cloud hosting: OpenStack is run by 50% of the Fortune 100 companies¹¹

Perhaps more exciting is to imagine the ways in which fully open infrastructure places ownership and control back in the hands of scholars, institutions, and the societies and associations that are allied with them and working to meet their needs.

7. Summary

Coko believes that no one platform can solve all the problems. We need an ecosystem of tools and software, creating modular and interoperable systems by the communities that they serve. Coko aims to provide the facility for publishers to move from closed and linear workflows to collaborative webspaces, and from proprietary platform silos to an open source ecosystem.

Coko is in the midst of facilitating the collaboration needed in the community to co-develop open source solutions for publishers. Coko powers the development of these solutions, and publishers collaborate to develop and customize. But the real innovation is coming from those publishers who are daring enough to reimagine their future and contribute to building the foundation upon which it will grow.

⁷<https://www.adamhyde.net/open-source-successes/> (last accessed 12 June 2018).

⁸<https://www.netmarketshare.com/browser-market-share.aspx> (last accessed 12 June 2018).

⁹https://en.wikipedia.org/wiki/Usage_share_of_operating_systems (last accessed 12 June 2018).

¹⁰https://en.wikipedia.org/wiki/Mobile_operating_system (last accessed 12 June 2018).

¹¹<https://www.openstack.org/assets/reports/OpenStack-AnnualReport2017.pdf> (last accessed 12 June 2018).

Acknowledgements

The Collaborative Knowledge Foundation is supported by the Gordon and Betty Moore Foundation, the Laura and John Arnold Foundation, and by a grant to Adam Hyde from the Shuttleworth Foundation.

About the Author

Kristen Ratan is the Co-Founder of the Collaborative Knowledge Foundation (Coko). She has twenty years of experience developing new technology, leading strategic innovations, and building community in the publishing industry. Kristen was most recently the Publisher at the Public Library of Science (PLOS). She has held Board positions with the Society for Scholarly Publishing and CrossRef.

References

- [1] Partners Imbue, 'Industry Leaders' Perspectives on the Digital Transformation Journey in Publishing' 2017, last accessed 12 June 2018.
- [2] D. Mietchen, R. Mounce and L. Penev, Publishing the research process, *Research Ideas and Outcomes* **1** (2015), e7547. doi:[10.3897/rio.1.e7547](https://doi.org/10.3897/rio.1.e7547), last accessed 12 June 2018.
- [3] G. Bilder, J. Lin and C. Neylon, Principles for Open Scholarly Infrastructure-v1, 2015, retrieved 23 March 2018, doi: [10.6084/m9.figshare.1314859](https://doi.org/10.6084/m9.figshare.1314859), last accessed 12 June 2018.
- [4] J. Priem and B. Hemminger, Decoupling the Scholarly Journal, *Frontiers in Computational Neuroscience* **6**(19) (2012). doi:[10.3389/fncom.2012.00019](https://doi.org/10.3389/fncom.2012.00019).
- [5] R. Vale, Accelerating scientific publication in biology, *Proceedings of the National Academy of Sciences* **112**: (2015), 13439–13446. doi:[10.1073/pnas.1511912112](https://doi.org/10.1073/pnas.1511912112).