Elizabeth Bedford, Chloe Dufour, Corinne Guimont, Rachel Howard, and Shane Nackerud

Preserving What We Publish

Findings from the Library Publishing Coalition Preservation Task Force

Library publishing programs have the potential to be a critical component of the community-controlled infrastructure pushing the scholarly publishing landscape toward more open and equitable practices. However, multiple studies have demonstrated that long-term preservation is particularly problematic for open access publications,[1] and the products of library publishing programs are unfortunately not immune. While preservation is a significant challenge for small publishing programs generally, in many ways library publishers are in a better position to meet it than their non-library-affiliated peers. Libraries have long been centers for preservation and have invested in individuals, tools, and partnerships that are at the forefront of the preservation effort. Yet the expertise available in libraries often seems disconnected from the library publishing practitioners who could benefit from it.

The Library Publishing Coalition[2] (LPC) aims to support its members in overcoming institutional siloing and addressing resource scarcity to better steward the materials they create. In mid-2021, LPC charged a Preservation Task Force with investigating the preservation activities and challenges of library publishers, and recommending actions for LPC to take to strengthen practice in this area.[3] We took a multi-step approach to this work: first investigating the activities of library publishers through a combination of literature review, analysis of the Library Publishing Directory data,[4] and additional information gathering through surveys and focus groups; second, exploring the landscape of community-led preservation efforts through conversations with groups providing preservation services and support; and finally, preparing a gap analysis of library publisher needs in comparison with the resources commonly available.

LPC has committed to strengthening its support for preservation activities in line with our resulting report's recommendations. However, there are plenty of actions we can take as individual library publishers, the organizations that support them, and as communities of practice to move the needle on more fully preserving this important content.

Preservation in Library Publishing Programs

Our initial investigation into existing information about library publishing preservation was less productive than we had hoped. The Library Publishing Directory questions about preservation programs conflated several tools and services in a way that made it difficult to

Elizabeth Bedford is scholarly publishing outreach librarian at the University of Washington, email: ebedford@uw.edu. Chloe Dufour is business librarian at the University of Pittsburgh, email: chd84@pitt.edu. Corinne Guimont is digital scholarship coordinator and interim director of publishing at Virginia Tech, email: gcorinne@vt.edu. Rachel Howard is digital initiatives librarian at the University of Louisville, email: rachel.howard@louisville.edu. Shane Nackerud is director of affordable learning and open education at the University of Minnesota, email: snackeru@umn.edu.

get an accurate picture of publishers' practices. Further, our literature review revealed that research into library-publishing-specific preservation practices is extremely sparse. A few articles mention preservation as part of wider case studies, but it is seldom the focus and never detailed at a level that would allow a comprehensive understanding of the workflow. In the few cases where preservation is mentioned, many authors conflate deposit into an institutional repository (IR) with digital preservation. This is troubling, since IRs are most often designed as access platforms rather than preservation platforms, and they usually cannot facilitate the range of ongoing activities like format characterization, migration, normalization, virus scanning, or fixity checking that make up active preservation. LPC's annual Library Publishing Forum appears to be one of the few venues where library-publishing-specific preservation practices are explored, with many presentations offering case studies and more substantive work.[5]

The task force therefore concluded we needed to collect additional data to get a more detailed understanding of current practice. We designed a survey to collect information about respondents' programs, preservation activities, and challenges, which was distributed through the LPC listserv and the IFLA library publishers listserv, ultimately receiving 36 responses. We presented the survey results at a community meeting on October 20, 2022, which allowed us to directly engage with 20 LPC community members and dig deeper into the themes that we discovered in our initial findings. The survey and community conversation revealed a broad range of preservation activity, capacity, and experience.

Some survey respondents and conversation participants had comprehensive and sophisticated preservation workflows. Using a mix of external digital preservation services, community partnerships, and internally developed tools and workflows, these programs have a good handle on preservation of traditional publishing outputs. However, many called out digital humanities (DH) projects as being a primary challenge. They are grappling with complex formats where best practices are still developing and automated tools are nonexistent, resulting in preservation workflows that are even more resource-intensive.

On the other end of the spectrum, 40 percent of survey respondents said that their program did not include any preservation activities at all, and three quarters of respondents said that no one in their publishing program has digital preservation in their title or job description. While many respondents indicated that there was preservation expertise at their institution, only about a third reported that any library publishing content fell under that program's domain. This siloing leaves programs without the expertise or resources to develop robust preservation practices, which in many cases results in suboptimal workflows. There is tremendous opportunity here for baseline information and support for easily implemented workflows to make a significant difference.

A Broad Landscape of Community Resources

Working from a list provided by the LPC Board and staff and refined by task force members, we contacted seven community-led and -driven organizations and projects to set up hour-long conversations.[6] The organizations we spoke with serve overlapping niches, and many work together toward shared solutions in the digital preservation space. This was not meant to be a comprehensive list of preservation tools and vendors, but rather enabled the task force to get a better sense of the types of preservation support services available to library publishers.

Five of the organizations we spoke with focused on enabling or providing direct preservation services and showcased diverse governance structures and technology strategies. Three of the organizations, including Academic Preservation Trust (APTrust), the MetaArchive Cooperative, and Scholars Portal, are collaborative enterprises between memory organizations. Project JASPER (JournAlS are Preserved ForevER), which is sponsored by the Directory of Open Access Journals (DOAJ), directs DOAJ-indexed journal publishers to the external digital preservation service providers CLOCKSS,[7] the Public Knowledge Project Preservation Network,[8] and the Internet Archive.[9] Portico is a stand-alone not-for-profit organization offering preservation services by serving as a central hub for both libraries as access providers and publishers as content providers. The range of technology is similarly broad. Many of the providers we spoke with use LOCKSS[10] (Lots of Copies Keep Stuff Safe) systems, which are an open source distributed digital preservation technology that allows partner libraries to mirror and monitor each others' content. However, many have developed their own unique tools and workflows, some of which use open source software and some of which are built on proprietary systems.

Two of our conversations were with projects focused on creating information resources. The NASIG Digital Preservation Committee developed a broad-model digital preservation policy,[11] which can be tailored to suit different needs. Community-led Open Publication Infrastructures for Monographs (COPIM) primarily builds tools and platforms for open access book publishing, but one of their work packages is a toolkit of best practices for preservation of those publications (WP7: Archiving and Digital Preservation[12]). While the breadth and depth of available best practices and templates are impressive, the fact that so many projects sponsored by so many types of organizations (NASIG originated as serials access interest group, while COPIM is a partnership between libraries, publishers, universities, and community organizations) exist makes keeping tabs on available resources potentially overwhelming for individual library publishing programs.

Findings and Calls to Action

Breaking Down Silos and Gathering Information

The need for digital preservation of library published work has generally been demonstrated and seems to be clearly understood by library publishers. Despite this, digital preservation is often a secondary priority in library publishing programs, with little staffing or financial support given to this important task. Digital preservation programs are themselves often understaffed and underfunded and have therefore needed to set boundaries around taking on additional content. To be successful, programs must build a business case for these activities: aligning their work with the organizational mission, articulating why content should be included in the program's preservation policy, and quantifying the resources needed for varying levels of preservation. But as a first step, library publishers should take stock of their program's current practices and gather the information they need on how they might build toward better ones. This could include talking to their library's institutional repository and preservation departments about what those systems include and whether there is any capacity to expand services.

Similarly, library publishers can investigate their publishing platforms' ability to integrate with preservation services, since several publishing tools have integrated preservation into their project infrastructure. This includes tools like Open Journal Systems[13] (OJS), which

makes it very easy to send preservation data to CLOCKSS or Portico, and Janeway,[14] which also makes it easy to send preservation content to Portico.

The digital preservation community has done a tremendous job in creating resources and documentation, but somehow this information isn't getting to many library publishing practitioners. As articulated by a survey respondent, "Rather than a focus on reinventing the wheel, we'd prefer to see LPC help in the area of sharing information and resources funneling best practices that have already been established." In the community call, many attendees shared links and resources to existing best practices, demonstrating that there is not necessarily a need to create new guidelines but rather to create navigation pathways to take advantage of the wealth of existing resources.

Supporting Digital Humanities and Non-traditional Publications

While many of the existing best practices for digital publications may apply to digital humanities (DH) or nontraditional publications, not all of them do. These complex projects require a completely different approach to preservation and may even differ project to project. All the community organizations we spoke with agree on the need for additional attention to emerging preservation needs in digital humanities, 3D, and research data realms. Indeed, most of the service providers we interviewed only offer bit-level preservation rather than the ongoing format monitoring, migration, and emulation that would be ideal preservation practice. This is sufficient for traditional journals, but for multimodal journals this preserves the components but not their full functionality. Portico is aware of the need to preserve emerging formats and is looking into approaches, but clearly this is a developing area.

Forging More Direct Partnerships

Our community conversation participants emphasized that the fastest way to increase preservation of the materials library publishers produce would be to encourage preservation integration into the platforms they are already using. At minimum, it would be extremely helpful for publishing platform developers to increase transparency around the current functional preservation capacity of their software, including documentation on how to take advantage of existing capabilities. Ideally, platform developers would incorporate preservation best practices into their export capabilities, including the option of creating a complete submission information package, which consists of descriptive information of the contents and files to be used for long-term preservation. For those platforms where archival export options exist, developers could create plugins that allow linking with existing preservation services. For this to happen, publishing and preservation tool developers should come together to work in parallel and arrange for library publisher partners to assist with requirements gathering and testing.

Community models such as AP Trust, the Global LOCKSS Network, Scholars Portal, and MetaArchive Cooperative have been tremendously successful in expanding preservation of digital materials. How do we encourage institutions that are interested in creating new preservation communities, and how do we support existing collaborations? Our conversation with one of the preservation service providers revealed the alarming possibility that LOCKSS systems are potentially endangered by the trend of libraries investing in cloud-based infrastructure rather than in-house server hosting, since costs would be prohibitively

high to use cloud infrastructure as a LOCKSS node. We must make sure that preservation systems are part of the conversation when discussing institutional IT strategy.

Finally, significant expertise around the preservation of multimodal and complex formats exists outside of the library community. Subject-specific preservation services such as the Archaeology Data Service can be more sophisticated with these resources than are more traditional library-facing preservation services.[15] How can we better liaise with disciplinary organizations to learn from their experience and partner on finding solutions?

Library publishers are entrusted with disseminating a diverse range of materials, but our organizations' broader role as stewards of knowledge requires an additional commitment to long-term preservation. As library publishing programs mature, it is clear we must build a shared understanding of preservation best practices, technologies, and workflows to fully prioritize the preservation of our published works. There is a wealth of preservation expertise within libraries, cultural heritage organizations, and disciplinary communities, and by actively coordinating and learning from each other we can ensure the integrity and accessibility of our digital holdings for the long term. ✍

Notes

1.   See Mikael Laakso, Lisa Matthias, and Najko Jahn, "Open Is Not Forever: A Study of Vanished Open Access Journals," *Journal for the Association of Information Science and Technology* 72, no. 9 (September 2021): 1099–1112, https://doi.org/10.1002/asi.24460; and Mikael Laakso, "Open Access Books through Open Data Sources: Assessing Prevalence, Providers, and Preservation," *Journal of Documentation* 79 , no. 7 (2023): 157–77, https://doi.org/10.1108/JD-02-2023-0016.

2.   Library Publishing Coalition homepage, https://librarypublishing.org/.

3.   The LPC Preservation Task Force included chair Elizabeth Bedford of the University of Washington, Chloe Dufour of the University of Pittsburgh, Corinne Guimont of Virginia Tech, Rachel Howard of the University of Louisville, Amanda Hurford of Private Academic Library Network of Indiana (PALNI), and Shane Nackerud of the University of Minnesota, as well as guest members Jennifer Kemp of Crossref and Alicia Wise of CLOCKSS.

4.   "Library Publishing Directory," Library Publishing Coalition, accessed December 8, 2023, https://librarypublishing.org/lp-directory/.

5.   "2024 Library Publishing Forum," Library Publishing Coalition, accessed December 8, 2023, https://librarypublishing.org/forum/.

6.   The groups the Task Force interviewed were Academic Preservation Trust (APTrust), https://aptrust.org/; the MetaArchive Cooperative, https://metaarchive.org/; Scholars Portal, https://scholarsportal.info/; Project JASPER (JournAlS are Preserved ForevER), https://doaj.org/preservation/; Portico, https://www.portico.org/; NASIG Digital Preservation Committee, https://nasig.org/Digital-Preservation-Committee; and Community-led Open Publication Infrastructures for Monographs (COPIM), https://www.copim.ac.uk/.

7.   Digital Preservation Services—CLOCLSS homepage, https://clockss.org/.

8.   "PKP Preservation Network," Public Knowledge Project, Simon Fraser University, accessed December 8, 2023, https://pkp.sfu.ca/pkp-pn/.

9.   "Internet Archive Web and Data Services," Internet Archive, accessed December 8, 2023, https://webservices.archive.org/.

10. LOCKSS Program homepage, https://www.lockss.org/.

11. NASIG, "NASIG Model Digital Preservation Policy," March 2022, https://nasig.org/NASIG-model-digital-preservation-policy.

12. "WP7: Archiving and Digital Preservation," COPIM (Community-led Open Publication Infrastructures for Monographs), accessed December 8, 2023, https://www.copim.ac.uk/workpackage/wp7/.

13. "Open Journal Systems," Public Knowledge Project, Simon Fraser University, accessed December 8, 2023, https://pkp.sfu.ca/software/ojs/.

14. Janeway homepage, https://janeway.systems/.

15. See Archaeology Data Service, "Preservation Policy and Procedures," accessed December 8, 2023, https://archaeologydataservice.ac.uk/about/policies/ads-policies-and-procedures/.