Enhancing Services to Preserve New Forms of Scholarship

Kate Wittenberg
Portico, United States
kate.wittenberg@ithaka.org

David Millman
New York University, United States
dsm@nyu.edu
https://orcid.org/0000-0001-5028-1101

Karen Hanson
Portico, United States
karen.hanson@ithaka.org
https://orcid.org/0000-0002-9354-8328

Craig Van Dyck
CLOCKSS, United States
cvandyck@clockss.org
https://orcid.org/0000-0001-6277-6013

Susan Doerr
University of Minnesota Press, United States
doer0012@umn.edu

Abstract – The advance in technologies for publishing digital scholarship has outpaced the development of technologies for reliably preserving it. Authors and publishers are creating increasingly sophisticated products without realizing that some of their enhancement choices might put preservability--and valuable scholarship--at risk. The poster describes the in-progress work and findings of a collaboration between preservation organizations, libraries, and publishers that are creating enhanced digital publications. The work aims to identify what can be effectively preserved with existing technologies, and to produce a recommended set of practices to help authors and publishers prioritize and plan their enhanced digital products for maximum preservability

Keywords – monographs, publishing, collaboration, preservation

Conference Topics – Collaboration: a Necessity, an Opportunity or a Luxury?; Building Capacity, Capability and Community.

I. Background

The advance in technologies for publishing digital scholarship has outpaced the development of technologies for reliably preserving it. Authors and publishers are creating increasingly sophisticated products without realizing that some of their enhancement choices might put preservability--and valuable scholarship--at risk.

As scholarly communication becomes more complex, publishers are creating digital products that are more than mere proxies for a print original; they offer new types of content and user experience. Features of these new forms of scholarship might include audio and video content embedded within text; interactive, high-resolution images, maps, and visualizations; annotations that can be shared, saved, and further annotated, privately or publicly; non-linear paths of engagement; and custom-built digital publications. Specific examples of this form of scholarship can be found in the Library of Open Access Titles of the University of Minnesota Press [1], which uses the open source Manifold platform, Fulcrum of the University of Michigan Press [2], the digital projects of the Stanford University Press [3], and Open Square of the NYU Press [4].

A single publication on these platforms can potentially include many of the innovations mentioned. Each of these innovations presents preservation challenges; their combination creates an even greater challenge: the need to maintain multiple formats and the connections among them.

While there has been remarkable growth within the digital preservation community in developing shared strategies, practices, and tools, this knowledge has matured around a limited scope of file formats--text, audio, video, and image. With the

iPRES 2019

exception of Web archiving, efforts to preserve complex digital objects have been fewer and smaller in scale. Preservation workflows at scale were designed for simpler objects.

II. PROJECT SCOPE

The poster presents the in–progress work and findings of a project funded by the Andrew W. Mellon Foundation. The *Enhancing Services to Preserve New Forms of Scholarship* project consists of two key sets of activities that will be performed over 18 months, beginning in April 2019.

A. *Identify what can be preserved*

The first activity is to determine the degree to which the enhanced forms of scholarship *can* be preserved using existing technologies. This will focus on scalable solutions that can be incorporated into the workflows of the participating preservation institutions. The ideal outcome would be a seamless transfer of enhanced publications to the preservation organizations with sufficient actionable information for recreating the user experience, as well as ensuring that their component bits can be preserved.

B. *Develop strategies and guidelines*

Iteratively and in parallel with the first activity, the project partners will develop a set of guidelines and current best practices identifying the classes of materials and workflows more likely to lead to successful preservation. The guidelines will describe a set of conventions that can serve as a model for other publishers and, potentially, create an environment in which digital scholarship has a better chance of being effectively and sustainably preserved at scale

III. PARTNERS

The project involves collaboration among a group of university presses, libraries, and preservation services who are looking at innovative, effective, and sustainable approaches to preservation of these new forms of scholarship.

The publishing organizations include NYU Press, Michigan Publishing, the University of Minnesota Press, Stanford University Press, and Vega Academic Publishing at Wayne State University.

The preservation service organizations include CLOCKSS, Portico, and the libraries of the University of Michigan, Duke University, and NYU. The team working on the Emulation as a Service Infrastructure (EaaSI) project at Yale University will be consulted for materials demanding either a particular "original experience" or their original software context.

IV. FORMATS AND FEATURES

After considering the composition of a variety of enhanced works, a list of formats and common features was identified for analysis. First, the project will include 3 major formats:

- EPUB3s with embedded multimedia
- Self-contained HTML5 projects that are not dependent on remote resources
- Web-based publications dependent on external content, code, or systems.

In addition, several features that are typical of these works will be given separate attention:

- Works with supporting files -- specifically video, audio, software, code, or data -- that are linked or embedded, local or remote.
- Works with dynamic features more complex than hyperlinks that require content to be retrieved from a remote server based on user interactions. These features could have: (a) a finite number of states e.g. click play to stream video, hover-over for a pop-up; or (b) an open-ended number of states – it's likely that a replica of the live service would be required for the full experience e.g. full text search engine, map navigation, live feed.

Analysis of each work will include consideration of versioning support and the ownership and rights of supporting material. Portico and CLOCKSS will receive or collect representative samples and attempt to ingest them into the archive. The results will contribute to the final guidelines documentation.

iPRES
20I9

REFERENCES

[1]   Library of Open-Access Titles, University of Minnesota
       Press. https://manifold.umn.edu
[2]   Fulcrum, University of Michigan Press.
        https://www.fulcrum.org
[3]   Stanford Digital Projects, Stanford University Press.
       https://www.sup.org/digital
[4]   Open Square, NYU Press. https://opensquare.nyu.edu

iPRES
2019