

Effectiveness and Durability of Digital Preservation & Curation Services

Ithaka S+R requests \$240,102 for a 16-month research project to conduct eight case studies that will examine and assess how digital preservation and curation systems (DPCS) are developed, deployed, and sustained, identifying key success factors and impediments to the long-term success of these systems. This National Digital Infrastructures and Initiatives (NDII) project will result in best practices and lessons learned to guide the design choices of future systems, offer alternative sustainability models, compare organizational characteristics of open source (OS) or commercial products, and inform the end-users' selections and grantmaker's policies.

Statement of National Need

During the last two decades, cultural heritage institutions have participated in the development of community-based DPCS, such as DuraSpace, Samvera/Hyku, and Mukurtu, to foster cooperation among cultural institutions that share a common vision of safeguarding and increasing access to our cultural heritage. These investments have led to a scholarly communications sector that is brimming with small, independent start-ups and membership organizations. Libraries, archives, and museums have a growing dependency on these digital platforms to support the curation, discovery, and long-term management of digital content. Yet, as we witness the recent organizational challenges faced by services such as Digital Preservation Network and Digital Public Library of America, we are reminded of the importance of creating durable services with clear visions and value-propositions that are aligned with the marketplace and available resources. As participants at a recent NEH/Mellon workshop recognized, there is a crisis in the sustainability of key infrastructure for the humanities, including systems responsible for preservation.

As DPCS and their usage proliferate, it is critical to understand how their deployment and long-term maintenance needs are factored into the initial development and adoption stages in order to ensure their durability and effective use of funding. Yet only limited research and knowledge exist on the organizational, structural, cultural, and economic enablers and barriers that make the ongoing success of these preservation systems viable. Building on the recent OS software studies that outline best practices for sustainability, offer strategies and metrics for assessing and managing community projects, and map out the scholarly communication infrastructure,¹ our work will focus on DPCS, and compare the organizational characteristics of OS or commercial initiatives, offer lessons learned through case studies, and propose alternative sustainability models for long-term maintenance and development.

Objectives, Research Questions, Project Design

Ithaka S+R proposes a NDII project that will investigate how DPCS are designed, implemented, and assessed in order to identify key success factors and impediments related to the development, deployment, and sustainability of these systems. Given our broad experience on issues related to collaboration, sustainability, funding models, funders' practices, and OS software for heritage institutions, Ithaka S+R is uniquely qualified to undertake this work.² Oya Y. Rieger, Senior Advisor will lead our team for Ithaka S+R and Roger Schonfeld, Director of Libraries, Scholarly Communication, and Museums at Ithaka S+R. We will form an advisory committee with representation from the leadership of various cultural heritage organizations, DPCS development projects, and sustainability/business experts.

The project will seek to answer the following research questions that pertain to community-based DPCS:

1. What business approaches³ are used to plan and implement community-based DPCS and how do they compare to strategies followed by commercial entities offering similar products?

¹ Examples of related work include: [National Digital Platform at Three, It Takes a Village: Open Source Sustainability](#), [Community Cultivation: A Field Guide](#), [Community Health Analytics Open Source Software](#), [Understanding What Constitutes a Vibrant Open Source Community](#), [Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race?](#)

² Ithaka S+R conducts research and provides convening and advisory services. Please see our work on [funding and sustainability](#) and [cross-institutional collaboration and open source software](#). See also Schonfeld's [relevant pieces](#) at the Scholarly Kitchen.

³ Business approaches include utilization of business plans, governance models, usability studies, market research, risk assessment and mitigation, and agility and flexibility to navigate and adapt to change in time of operational challenges, changing priorities, evolving leadership, and shifting funding streams.

2. How are long-term maintenance and development requirements factored in business life cycle planning beyond the initial investment stage?
3. What strategies are deployed to accommodate different needs and resources of cultural institutions that serve user communities with diverse geographic, cultural, and socioeconomic backgrounds?

Our research approach will include desk research on OS and commercial DPCS initiatives and interviews with individuals from leadership, governance, staff, and funders. Data analysis will be conducted through grounded theory with an inductive approach to data coding. We will utilize NVivo qualitative data analysis software. These cases studies will be the basis for best practices to promote sustainability while providing efficient access to digital content and collections at scale for all types of users, as well as advance realistic and sustainable approaches for the stewardship of new and complex content types and emerging digital formats

September - October 2020. Review reports, project wikis, and social media about DPCS. Convene advisory committee. Select eight DPCS⁴ for case studies and arrange in-person and virtual interviews.

November 2020 - February 2021. Conduct interviews with leadership, governing boards, and staff involved in developing and maintaining DPCS about initiation and development processes. Seek input from users and non-users⁵ about usability, functionality, and reliability of DPCS examined.

March 2021 - April 2021: Analyze case study findings. Review findings and recommendations with advisory committee.

May 2021 - August 2021: Plan and launch a series of publications, webinars, and blog posts, and interviews with project participants to start sharing findings.

September 2021 - December 2021. Publish final report. Present findings broadly.

Statement of National Impact

The proposed research will provide important insights and guidance on best practices to promote sustainability while providing efficient access to digital content and collections at scale for all types of users to the institutions that invest in, the individuals who build and maintain, and the community that benefits from DPCS.

Furthermore, it will advance realistic and sustainable approaches for the stewardship of new and complex content types and emerging digital formats help them. To support these outcomes, we will produce the following deliverables:

1. A major final report along with a series of webinars, blog posts, presentations, and interviews throughout the project, all distributed on a CC-BY-NC basis, targeting audiences who design, develop, select, use, assess, and fund DPCS.
2. Customized briefings and workshop for IMLS (and other invited funders) on how to apply the outcomes of this research to support the long-term success of initiatives that start with grants and gifts.
3. Convening of participants of the case studies, members of the relevant higher education community, and interested funders and policy makers to facilitate a community-based discussion of the research findings, implications, and alternative models (to be held at the end of research, prior to the final report).

Budget Summary

We estimate that this project will cost a total of \$240,102 over a period of approximately 16 months, including \$198,075 in personnel costs, \$15,200 in travel, \$5,000 in convening expenses, and \$21,827 in indirect costs.

⁴ Initial case study candidates include Archivematica, Chronopolis, ContentDM, DPN (defunct), DuraSpace, Dryad, EPrints, figshare, Greenstone, HathiTrust, Invenio, LOCKSS, Mukurtu, Portico, Rosetta, Samvera/Hyku, and Zenodo. Although our focus will be on open source and community-build systems, we will include two commercial products in our case studies to allow comparison.

⁵ The term non-users denote groups and individuals who choose not to use the service, stopped using the service, preferred a comparable service, or are unaware of the service's existence.