2009

MetaArchive: A Cooperative Approach to Distributed Digital Preservation

Katherine Skinner
Emory University, katherine.skinner@emory.edu

Martin Halbert
Emory University, martin.halbert@emory.edu

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation
DOI: http://dx.doi.org/10.7771/2380-176X.2489

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Tracking E-journal Preservation from page 34

MetaArchive: A Cooperative Approach to Distributed Digital Preservation

by Katherine Skinner (Digital Projects Librarian, Emory Univ., 540 Asbury Cir., Atlanta, GA 30322; Phone: 404-783-2534; Fax: 404-727-0827) <katherine.skinner@emory.edu> http://metaarchive.org

and Martin Halbert (Director for Digital Innovations, Emory University, 540 Asbury Cir., Atlanta, GA 30322; Phone: 404-727-2204; Fax: 404-727-0827) <martin.halbert@emory.edu> http://metaarchive.org

What role will the Library take in digital preservation? On first glance, the question seems relatively easy to answer. As the library continues to transition from its centuries-long focus on print assets to a combination of print and digital resources, it will take an active role in the preservation of our digital cultural resources that is similar to that which it has long undertaken in the print realm.

Or will it?

Of late, many of us in the library field have become preoccupied with the concept of digital preservation — and rightly so. We wonder aloud about the forms that digital preservation will take, the amount it will cost, the rigor demanded in its implementation, and the feasibility of different organizational approaches to digital preservation.

But what does it mean to participate? How do we want to be involved? And what role(s) should we, as librarians and archivists, aspire to take in the realm of digital preservation?

Questions such as these led to the founding of the MetaArchive Cooperative, a collaborative network of institutions that have banded together to communally approach the challenges of preserving digital assets. The original six members founded this Cooperative due to their strong belief that libraries both could and should be actively engaged in the creation and maintenance of their own digital preservation solution. They knew that alone, none of these institutions were likely to create and maintain — much less sustain — a robust digital preservation infrastructure. But if they approached the issue as a group and built a shared infrastructure, they could accomplish together what no one institution had the resources to achieve in isolation.

The MetaArchive Cooperative: A Shared Digital Preservation Infrastructure

The MetaArchive Cooperative (http://MetaArchive.org) formed to enable cultural memory organizations to effectively and mutually preserve their archival digital assets for themselves. MetaArchive began in 2004 as one of the original eight initiatives contracted for the Library of Congress under the National Digital Information Infrastructure and Preservation Program (NDIIPP). The venture was led by Emory University in collaboration with Georgia Tech, University of Louisville, Virginia Tech, Auburn University, Florida State University, and the Library of Congress. The network established by this group was the first major effort to build and operate a private implementation of the open source LOCKSS (for Lots of Copies Keep Stuff Safe) software for digital preservation (http://www.lockss.org), an approach that has since been termed a Private LOCKSS Network, or PLN. The MetaArchive PLN is a distributed preservation infrastructure that meets the OAIS Reference Model standards for repositories.

Technically speaking, the foundation of the network is the open source LOCKSS software developed at Stanford University, which enables a group of LOCKSS caches, or node servers, to work together across geographical space to replicate and preserve content. MetaArchive is the only PLN in operation thus far that does not depend on the LOCKSS team to administer the network; we run a separate cache manager (coded in collaboration with the LOCKSS team) to monitor our network. The MetaArchive Cooperative has created and layered additional modules on top of the LOCKSS framework to provide our members with administrative tools, including a conspectus database and the cache manager. The conspectus database enables members to capture collection-level metadata for preservation decisions and actions, and the cache manager serves as a monitoring tool for network-wide tracking and troubleshooting activities. We are in the process of packaging these open source software components for use by other PLNs, and plan to release this software through SourceForge next year.

The organizational framework that we have constructed has been as integral to our success as the technological platform upon which we have built our preservation services. After running the network for three years, we transitioned from a sponsored-funding-supported project to an independent, membership association in 2007, a transition that has been greatly assisted through the support of the National Historical Publications and Records Commission. As part of this work, we founded a 501(c)3, the MetaArchive Services Group, to administer the Cooperative. All of the components of the network we run are owned and maintained by our member institutions. This decentralized apparatus enables the Cooperative and its services to be independent.
of each member — our members learn how to run and operate their own preservation node for the network, building their internal knowledge of the preservation process. They also are given opportunities to contribute to the software development efforts undertaken by the Cooperative.

The mission of the MetaArchive Cooperative is to support, promote, and extend our collaborative approach to distributed digital preservation practices. We have made our organizational model available to others as an example of how to create shared digital infrastructure. To this end, we not only run our own network, but also provide training and consulting assistance to other groups that wish to found similar preservation networks. We host workshops and make all of our documentation freely available to other collaborative projects and programs.

Unlike the public LOCKSS network, where participant libraries preserve journal content in which they all have a vested interest, the collections in the MetaArchive network are the unique holdings of each participant library and archive. In other words, MetaArchive’s members cannot rely on the incentive that drives participation in the public LOCKSS network — a shared body of content to which all subscribe and upon which all rely. Instead, the MetaArchive network requires a strong commitment between constituent institutions — each participates in order to preserve their own data in exchange for preserving other institutions’ data.

So what are the drivers in this PLN scenario? Topping the list are a strong sense of community engagement and a strong belief in the library’s cultural stewardship role. Our members share the conviction that libraries have a vested interest in preserving their own digital assets. Each has determined that they do not want to cede all of their digital preservation activities to external groups, and do want to participate in creating their own preservation solution. Building alone is a costly proposition, so these institutions have coupled their resources in order to achieve their preservation goals in a community-based effort.3

To enable this, MetaArchive formed as a cooperative, not a vendor. MetaArchive’s members do not pay for services, but rather make an investment to create and sustain their own preservation infrastructure. The Cooperative is more than a technical solution for preservation. It also functions as a learning environment in which members gain experience in developing and enacting a full preservation plan for their assets. Each member both contributes to and benefits from the expertise and the technical infrastructure developed by the overall community. In keeping with these principles, membership fees are kept at the absolute minimum required for the operation of the Cooperative, and range from $300 to $5K per year, together with a fee of $2 every 3 years per 1 GB of content contributed. These
other PLNs, which have largely opted to have LOCKSS manage and maintain their networks. We do benefit greatly from the LOCKSS team, both in terms of the regular updates they provide to the LOCKSS daemon and also in terms of the technical expertise they share with our central and distributed staff members, but we chose to build on an open source framework specifically because we believe this model offers the best odds for long-term sustainability. The overall LOCKSS community (including myriad PLNs) is already strong and it’s growing stronger. We believe that a solid open source development community could sustain and maintain the LOCKSS software if called upon to do so, and we have intentionally built a framework that relies only on this community, not on any one group within it.

Preservation and Institutionalization

Institutions form in order to address specific needs that are not already being met within the existing environment. This is to say that when dominant and traditional business practices (and libraries are a business, whether we think of ourselves as such or not) fail to meet community or market needs, it opens a space within which new institutions with new approaches may flourish. Witness Google, Elsevier’s journal services, and myriad other examples and exemplars that have already emerged to serve the information management and access needs of the digital age.

The library as an institution continues to serve many of the needs of its constituents — it is not in danger of persisting outright. However, it has not yet proven itself a serious contender in the digital realm. Scholars as well as the public are increasingly turning to companies such as Google to “to organize the world’s information and make it universally accessible and useful,” an access role that for centuries belonged primarily to the library field. To whom will these groups turn when they seek to preserve their digital assets, another core mission of the information science field? Will they turn to cultural memory organizations such as libraries and archives, or to corporations such as Amazon and Google? And should we, as cultural stewards, care so long as the preservation channel adequately provides for the needs of our institutions and our constituents? Is there a difference between commercially driven solutions and those created in the not-for-profit environment?

As libraries, our work is driven by the desire to maximize our stakeholders’ long-term access to materials, not by a desire to maximize profit for stockholders. This is a highly significant distinguishing factor and one that we cannot afford to take for granted. Cultural memory organizations are not, on the whole, profit-making enterprises. They are funded by tax dollars, foundations, and parent institutions whose constituents we serve. Our reason for being is to serve a public good — making our cultural assets, from books to datasets, accessible to the public for the long haul. If we cease to perform that function, instead outsourcing it to external parties, we are putting both our field and our cultural resources in precarious positions.

This is not to say that we should not outsource any of the digital preservation work we undertake. Just as is true in the print and physical artifacts domain, there will doubtless be portions of this work that are well suited to the work of external parties and too expensive to build in house. But, in order to know which parts are suited to in-house work and which parts should be outsourced, we must begin to explicitly engage in our own digital preservation solutions.

So from where will the successful approaches to digital preservation emerge? Sociologists tell us to watch the fringes, not the center, for seismic changes. Today, one of these fringes might well be the library. Or, in this case, many libraries, banding together in collaborative, cooperative ways to accomplish the preservation of their unique resources in a communally owned network environment that they run for themselves.