
OPERATIONS PLAN

NDIIPP MetaArchive Project



2008-10-24

Summary

This document describes the operations of the MetaArchive Cooperative. The document includes information regarding our decentralized structure, our day-to-day workflows, and how we monitor and control the quality of our services.

About the Cooperative

Today, more than 93% of the world's information is produced as digital files, not print documents. Increasingly, cultural memory organizations are seeking ways to care for these new digital resources—from government websites to corporate emails and from scanned images to born-digital recordings. As evidenced by such catastrophic events as blackouts, fires, and hurricanes, as well as basic hardware and software failures, these organizations need to act now to begin providing long-term digital preservation services for our digital history or we risk losing them altogether.

The MetaArchive Cooperative is an independent, international membership association that provides collaborative and distributed digital preservation solutions for cultural memory organizations, including university libraries, government and historical research archives, museums, and other similar institutions (<http://www.metaarchive.org>). The central missions of the MetaArchive Cooperative are to support, promote, and extend the practice of distributed digital preservation; to serve as a catalyst and guide for other networks that seek to implement the distributed digital preservation methods it has developed; and to educate cultural memory organizations about distributed digital preservation.

To these ends, the Cooperative offers several important services to a wide range of cultural memory organizations. These include:

1. Distributed preservation for multi-format collections of digital objects
2. Training opportunities for groups/individuals interested in distributed digital preservation methods
3. Consulting services for groups/individuals interested in administering, participating in, or knowing more about distributed digital preservation strategies.

The MetaArchive Cooperative is, as its name suggests, a **Cooperative**—meaning it is an association managed and supported by its members with the purpose of supplying themselves with a needed service—distributed digital preservation—at a moderate price. All MetaArchive members share in the management of the Cooperative and, as such, have a shared interest in developing and sustaining its technical and organizational infrastructures. This sets the Cooperative apart from other digital preservation solutions, which operate on a vendor/client basis (e.g., Amazon S3, Portico, OCLC's

Digital Archive). The Cooperative, unlike these businesses, provides a way for cultural memory organizations to **own and control the process of digital preservation for themselves**. Instead, instead of administering a preservation service *for* organizations, MetaArchive encourages organizations to invest in their own infrastructures and build their own capabilities by participating in a decentralized preservation network.

Given that preservation is a core mission for most cultural memory organizations, the importance of maintaining both the knowledge of how their digital preservation solution works and control over that solution cannot be overstated. There is a marked difference between the aspirations of cultural memory organizations and commercial vendors. Most cultural memory organizations, whether university libraries, government and historical research archives, museums, and other similar institutions, seek to make our cultural resources available and accessible to either specific stakeholders or the general populace for such purposes as long-term record-keeping, study, and knowledge-building. Most commercial vendors, in contrast, seek financial gain. By outsourcing key missions (for example, in the academic world, the digitization of journals) to vendors, cultural memory organizations suffer in three ways:

1. Usually, they have limited knowledge of how the solution actually works and thus no power to perform it or even clearly assess it for themselves,
2. They become marginalized as organizations that contract for services rather than performing their own key missions, and
3. They (and their materials) are subject to the rising prices that vendors will charge in order to maximize their financial gain.

By empowering cultural memory organizations to administer and run their own digital preservation solution, MetaArchive encourages them to continue performing, rather than outsourcing to commercial vendors, a core mission of such organizations.

The MetaArchive Cooperative also differs significantly from most other digital preservation solutions in that we practice **distributed** digital preservation. Using a technical framework that is based on the LOCKSS (Lots of Copies Keep Stuff Safe) software developed at Stanford University (<http://lockss.org>), MetaArchive members' collections are ingested into a geographically distributed network where they are stored on secure file servers run by our members in multiple locations. These servers do not merely back up the materials, but rather provide a dynamic means of constantly checking each file and providing repairs whenever necessary.

The purpose of the MetaArchive Cooperative is to bring organizations together to jointly create what they cannot accomplish alone: **affordable** distributed digital preservation services. The methods we engage to preserve our digital files differ substantially from the processes university libraries, government and historical research archives, museums, and other similar institutions have used to preserve physical collections. What has not changed is our commitment to our central goal: to ensure that our collections will remain available and usable for future scholars, researchers, and citizens.

Stage of Development

The MetaArchive initiative was founded by a team of six academic research libraries in 2004 as a collaborative project with the Library of Congress to research, develop, and implement an organizational infrastructure and technical service for distributed digital preservation. Following the success of this initial endeavor, the initiative transitioned in 2006 to an independent, unincorporated, international membership association, the MetaArchive Cooperative, for the purpose of supporting,

promoting, and extending our collaborative approach to distributed digital preservation practices. We are an active preservation services provider with an expanding membership.

Services Overview

The Cooperative offers the following services to a wide range of cultural memory organizations:

Distributed preservation for multi-format collections of digital objects.

The Cooperative provides its members with preservation services via a decentralized, geographically dispersed organizational and technical framework. This framework was designed to mobilize group efforts between cooperating organizations that seek to actively preserve their digital collections. The MetaArchive Cooperative does not have a single location of operation; nor does it control production of its services in a centralized way. Instead, member organizations contribute to the service by providing secure servers (known as MetaArchive-LOCKSS caches) in their own organizational settings. The individual members, then, comprise the operation in a decentralized fashion. The MetaArchive Cooperative has no central operating location, nor does it own equipment or other assets. The success of the Cooperative depends on member organizations working together to invest in their own infrastructures and capabilities. The Cooperative is administratively served by the MetaArchive Services Group, headquartered in Atlanta, Georgia.

The MetaArchive Cooperative runs a network of “dark archives” (i.e., no access is provided to non-authorized individuals to the network’s contents) that is dispersed across its member organizations. This network performs bit-level preservation for all file types and “full preservation,” or preservation with migration, for supported file types. The open source LOCKSS (<http://lockss.org>) software, developed by Stanford University, serves as the foundation for MetaArchive’s distributed technical system.

The MetaArchive Cooperative’s network is comprised of content provider sites whose content is ingested and preserved by MetaArchive-LOCKSS caches, or servers that participate in the network. These MetaArchive-LOCKSS caches are administered by MetaArchive’s member organizations that contract with each other to preserve this content in a collaborative, distributed manner. The technical framework provides a systematic way for these MetaArchive-LOCKSS caches to constantly check in with one another for signs of file degradation or “bit rot,” as well as for a cache’s corruption or disappearance due to technical or physical catastrophe. If a problem with a file/cache is detected, the other MetaArchive-LOCKSS caches conduct a “poll” to determine which copy has been compromised. Once they reach quorum, the software can repair that damaged copy. Similarly, if an entire MetaArchive-LOCKSS cache experiences technical failure and withdraws/disappears from the network, the cache can be safely recreated via the other caches in the network.

Each of the network’s MetaArchive-LOCKSS caches is under separate administrative control, which guards both against natural/technical disaster and the threat of economic failure of any one cache. Any cache in the system can drop out — whether intentionally or due to technical or organizational failures — without jeopardizing either the network or its contents. Myriad testing scenarios have consistently demonstrated the stability of this preservation system’s technical infrastructure.

The Cooperative welcomes participation at three distinct membership levels relative to the goals and preservation needs of members. All membership terms are for three-year periods. The Cooperative also works with and recognizes a range of affiliate organizations. The levels, costs (as of September 2008), and responsibilities for each membership tier are as follows:

- A. **Sustaining Members** are integral to the research, development, and deployment work of the MetaArchive Cooperative and contribute their internal staff and resources to the effort. They

maintain a MetaArchive-LOCKSS cache and also work directly on programming and network testing tasks. Each of the first 15 Sustaining Members also have a representative from their organization serve on the MetaArchive Steering Committee. The fee for joining as a Sustaining Member is \$5,000/year for a three-year period (or \$12,000 if three year membership paid in full).

- B. **Preservation Members** engage in ongoing preservation activities and actively participate in the network. They maintain a MetaArchive-LOCKSS cache and make that cache available for testing purposes. The fee for joining as a Preservation Member is \$1,000/year for a three-year period.
- C. **Contributing Members** are primarily organizations that do not have a sophisticated technical infrastructure but that do have valuable collections that they wish to preserve. They join the Cooperative to achieve that goal without incurring additional technical and resource obligations. The fee for joining as a Contributing Member is \$300/year for a three-year period. Included in this fee is one plugin¹ for a collection that the Contributing Member submits for ingest and preservation in the MetaArchive Preservation Network. Additional plugins can be purchased from the Cooperative on an as-needed basis.
- D. **MetaArchive Affiliates** are those organizations, consortia, and collaborative groups that we have advised and/or that have collaborated with us on distributed digital preservation topics. Most MetaArchive Affiliates are either Members or administrators of other Private LOCKSS Networks, or engage in substantive digital preservation activities.

The Cooperative currently has eight Sustaining Members and more than six dozen organizations that have contacted us with an interest in membership across the three levels or in affiliating with us. Our network is actively preserving more than 100 collections for a cumulative content of 1 TB of data and is doubling in size each year. We anticipate that we will double our membership by the end of 2008, and plan to add a substantial number of new members each year for the foreseeable future.

Administrating training opportunities for groups/individuals interested in distributed digital preservation methods

As the Cooperative mission states, we are committed both to growing our own network and advancing other distributed digital preservation networks. To these ends, the Cooperative administrates workshops for diverse groups, including individual organizations that are interested in joining MetaArchive or other digital preservation networks as well as groups of organizations or consortia that are contemplating starting their own networks utilizing the organizational and technical framework that we have built.

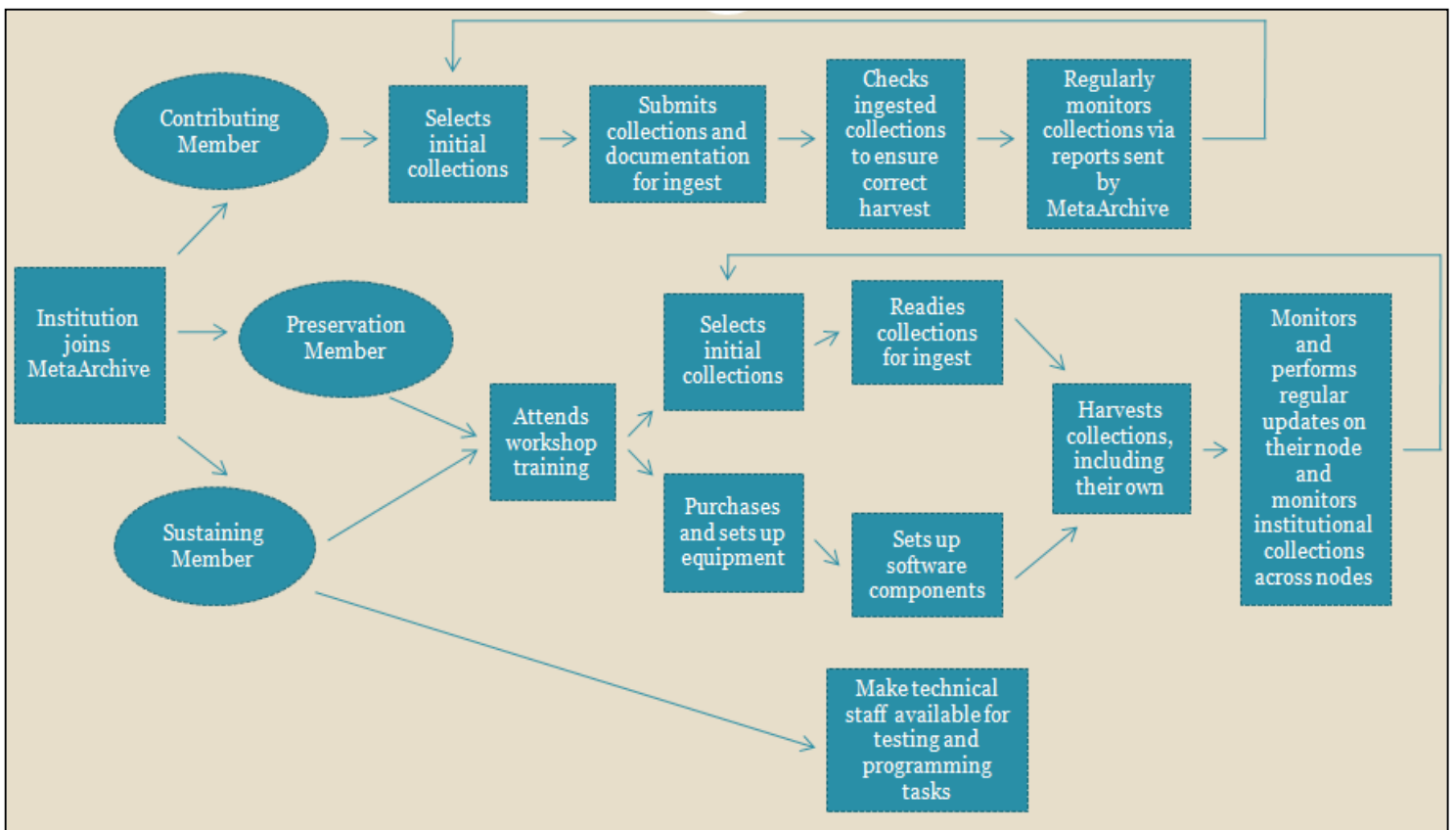
We are also writing a *Guide to Distributed Digital Preservation*. This book (which will be made available both as a freely downloadable pdf and as a printed book) will provide open documentation about running Private LOCKSS Networks to help both individual organizations and groups who seek more information and/or training materials to guide them as they join or establish a network.

¹ A plugin is a technical script that is required for ingest of each collection into the MetaArchive network. It defines the parameters of the collection, the rate at which it should be reharvested, and other important details that our network needs in order to ingest and preserve it.

Consulting services for groups/individuals interested in distributed digital preservation strategies.

The Cooperative provides consulting services for individual organizations and for groups of organizations that are interested in administrating, participating in, or knowing more about distributed digital preservation strategies. These include providing on-site presentations by MetaArchive Steering Committee and key staff members for preservation-oriented events; serving on Advisory Boards for projects and programs that are establishing networks; and providing systematic guidance to groups that are working to initiate their own networks.

Our basic workflow from a member organization’s perspective:



Expansion Plans

As previously mentioned, the Cooperative currently is comprised of eight sustaining members. More than six dozen organizations have contacted us with an interest in membership across the three levels, and we are currently working with four new organizations that intend to join us at various membership levels in the Fall of 2008.

There are two central ways in which we continue to increase our storage capacity and our geographic distribution within our network.

1. Each time that the Cooperative adds a new Sustaining or Preservation Member, it also adds a new MetaArchive-LOCKSS cache to the network.
2. Whenever a Sustaining, Preservation, or Contributing Member contributes more material than their initial space allotment allows (Sustaining: 40 GB; Preservation: 20 GB; and Contributing: 5GB), they pay a per-GB rate to cover the additional storage space required to ingest and distribute this content throughout the network (currently \$2/GB). This fee is used by the Cooperative to purchase additional GB of storage for the network. This additional storage space (in the form of new disks or servers) is then allocated to at least six² preservation caches in the network.

Suppliers, Risks, and Alternative Arrangements

As a decentralized business, the MetaArchive Cooperative only has two forms of “supplies:” software and servers. Because our infrastructure is not server-specific, there is no central vendor to whom we turn for our servers—each organization may choose to purchase a server that corresponds to the network’s Technical Specifications (documented in Appendix A of the Cooperative Charter and updated annually) from any vendor.

The MetaArchive Cooperative does, however, rely on its organizational members to provide the MetaArchive-LOCKSS caches that comprise the preservation network. In order to ensure the quality of our preservation services, the Cooperative demands that every file stored in the network is replicated across at least six distinct caches. This means that the network ideally must include at least six Sustaining and/or Preservation Members. In order to ensure that our network stays stable, we accept members in three-year terms and request that they provide us with a 90-day notice of renewal or resignation at the end of their membership terms. This provides us with a measure of insurance that we will be able to recruit a replacement member for any outgoing member prior to that outgoing member’s departure from the network.

In the unlikely case that our membership ever drops below this threshold, one or more organizations may bring up a second server as a temporary solution in order to satisfy the minimum number of caches required until a new member is recruited and in place.

MetaArchive also depends upon its Sustaining Members to provide the technical support needed to keep the decentralized business structure going. Although we have built upon the sound foundation of an open-source software package—LOCKSS—with a vibrant and growing user community, we do need technical staff members within the MetaArchive Cooperative’s membership who can provide ongoing maintenance and support for the network. To that end, our Sustaining Members agree to dedicate a portion of their technical staff time to support the MetaArchive Cooperative. If at any point this donated staff time from our Sustaining Members does not satisfy the needs of the Cooperative, the

² Six is the minimum number of caches in a MetaArchive network that host any given file/collection that is submitted for ingest and preservation

Steering Committee has the option to allocate funds from the Cooperative's membership fees to pay for staff or consultants to assist with technical duties.

Industry Standards and Compliance

The growing industry of preservation services will soon enjoy regulatory bodies that will certify specific businesses as compliant with the standards of our industry. To date, no such bodies exist, but there are several national and international guidelines emerging. To ensure the quality of the services we provide to our member organizations, the MetaArchive Cooperative is actively engaged with these emerging guidelines and works to fulfill the terms they provide.

The Cooperative was audited in 2007 by the NEDCC Digital Preservation Readiness Assessment team. This group was funded by the Institute for Museum and Library Services (IMLS) in 2006-2007 to develop and test an assessment tool for digital preservation readiness. The team completed their assessment of the MetaArchive Cooperative in March 2007, and the Cooperative is compliant with the valuable recommendations that this group provided to us.

Most importantly, the Cooperative is OAIS-compliant³ and complies with the guidelines provided by The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist developed by the RLG-NARA Digital Repository Certification Task Force (see <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>).

Quality Control Measures

In addition to our work to comply with emerging standards, the MetaArchive Cooperative fulfills its commitment to quality control in each of our three service areas.

For our preservation services, we provide quality control through 1) constant monitoring of the MetaArchive-LOCKSS caches, 2) regular network testing, including intentionally crashing and restoring test caches and test collections at six-month intervals; and 3) planning new development activities that will enrich our Cooperative services.

For our administered events, we ensure quality by conducting careful assessments of our workshops via pre- and post-event evaluations and making any necessary adjustments. For other training endeavors, including our documentation and the *Guide to Distributed Digital Preservation* book that we are currently writing, we engage peer review strategies, including administering "public comment" periods when appropriate.

With regard to our consulting services, we solicit and act upon feedback from participants in our on-site presentations, invite reviews by project/program managers when we serve on Advisory Boards for projects and programs that are establishing networks; and seek feedback via evaluative tools when we provide systematic guidance to groups that are working to initiate their own networks.

Notes on this document

This document was completed by the Steering Committee on 2008-11-18, edited by Matt Schultz in 2009-10, and approved by the Steering Committee in 2009-10.

³ See <http://public.ccsds.org/publications/archive/650x0b1.pdf> for details on the OAIS reference model.