MetaArchive: Final Report

Final project report for the National Historical Publications and Records Commission on the “MetaArchive, a Sustainable Distributed Digital Preservation Service”.

Project Period: January 2006 – December 2011
Report Submitted: March 13, 2012

Martin Halbert (Principal Investigator)
Cathy Hartman (Systems Director)
Katherine Skinner (Outreach Director)
Tyler Walters (Sustainability Director)
Matt Schultz (Project Manager)
# Table of Contents

- Executive Summary.......................................................................................................................... 3
- Project Highlights............................................................................................................................. 3
- About This Report: ............................................................................................................................. 5

## Section 1: Project Outcomes ........................................................................................................ 6
  - 1.1 Administrative Activities and Business Planning................................................................. 6
  - 1.2 Outreach Efforts ....................................................................................................................... 6
  - 1.3 Technical Framework ............................................................................................................... 7

## Section 2: Project Ambitions and Findings .................................................................................. 8
  - 2.1 Business Planning and Administrative Activities................................................................. 8
  - 2.2 Outreach Efforts ..................................................................................................................... 13
  - 2.3 Technical Framework ............................................................................................................. 16

## Section 3: Continuation and Long-Term Impact......................................................................... 18

## Section 4: Project Staffing ............................................................................................................ 19

## Section 5: Appendices
Executive Summary

The University of North Texas and its partner institutions, Educopia Institute, Georgia Institute for Technology, and the Chronopolis group of the San Diego SuperComputer Center are pleased to provide this final report to the National Historical Publications and Records Commission regarding the success of our work to establish a sustainable framework for the MetaArchive Cooperative distributed digital preservation network.

The NHPRC’s generous support has enabled the project team to produce a sustainable business model for the Cooperative’s operations and a technical bridge between LOCKSS and a data-grid preservation storage environment (iRODS-Chronopolis) for distributed digital preservation initiatives. The NHPRC’s support has also provided the MetaArchive Cooperative with the resources necessary to produce a stable technical and organizational infrastructure and model for the broader memory community.

We firmly believe that libraries and other memory organizations are making decisions today regarding their digital content that will determine their institutional viability in the future. We hope that our work will help to encourage other memory institutions to embrace a preservation model that actively engages them in the act of preserving their content in collaborative, memory-organization-owned networks.

Project Highlights

The project team is pleased to report that it has met or exceeded each of its project goals. We are also pleased to share that with this project, we have laid a solid foundation for future development work that will continue forward via several UNT-Educopia-Chronopolis collaborations in the next two years. We are providing a model for strategic alliance between different digital preservation approaches, one that we hope will further the development of the larger preservation field as well.

Highlights of the MetaArchive research and development project's accomplishments during the 2008-2011 project period include the following:

Administrative Activities and Business Planning

- We developed, formalized, and refined the MetaArchive Charter and Membership Agreement as the Cooperative’s core governance documents;
- We formalized the organizational structure of Educopia Institute and the MetaArchive Cooperative and implemented the administrative apparatus to host staff, and process billing, payroll, and employee benefits through Educopia Institute;
- We documented a Preservation Business Model that includes a Strategic Plan, Operations Plan, Management Plan, Financial Plan, Marketing Analysis, and Risk Assessment (See Appendix A);
• We diversified our funding streams for the Educopia Institute by formalizing Educopia Consulting as an additional program to aid in sustainability and began administering over one dozen consulting contracts;

• We conducted two rounds of bi-annual surveys with our member institutions to gather information concerning the technical staffing roles, responsibilities, and time investments that our members incur as part of their work with the Cooperative;

• We hosted four semi-annual and four annual Steering Committee meetings;

• We successfully adjusted MetaArchive member categories, benefits, responsibilities, and pricing levels to expand membership and to ensure the effectiveness and sustainability of Educopia’s program hosting; and

• We undertook an audit using the Trusted Repositories Audit and Certification methodology and updated and developed documentation, policies, and procedures in accordance with the recommendations of this audit.

**Outreach Efforts**

• We developed and formalized an Outreach Program Implementation Plan that committed to delivering at least five presentations and workshops, one peer-reviewed publication, and recruitment of at least 20 U.S. institutions;

• We delivered over 30 presentations and ten workshops;

• We drafted and completed with the MetaArchive membership *A Guide to Distributed Digital Preservation* that was published through Educopia in February 2010; published five peer-reviewed articles and published two ARL reports;

• We recruited 40 new institutions (including 30 participating institutions at our Collaborative Member sites) and integrated them into the organizational and technical framework of the Cooperative. We also renewed seven members for additional three-year membership terms;

• We formalized the relationship between the MetaArchive Cooperative and the Networked Digital Library of Theses and Dissertations (NDLTD) to engage their members in our preservation services at a reduced membership rate;

• We kicked off a targeted outreach campaign to U.S. LOCKSS Alliance/NDLTD members and furthered conversations with several prospective members.

**Technical Framework**

• We hosted a technical planning meeting in February 2009, to explore needs and possibilities for alignment of development efforts in various digital preservation efforts, particularly between Private LOCKSS Networks and other heterogeneous networks;
• We successfully completed two rounds of content exchange between the Chronopolis and MetaArchive-LOCKSS environments (See Appendices B, C, D & E);

• We completed the development and packaging of alternative data models and microservices to enhance MetaArchive’s ability to manage its digital objects and various aspects of the preservation network (See Appendix F);

• We developed with the Library of Congress, Chronopolis, University of North Texas, CDL, Penn State, and others a draft specification for replicating and exchanging BagIt-based content and collections. (See: https://github.com/acdha/restful-bag-server); and

• We began outlining the MetaArchive Contingency Plan (See Appendix G).

We look forward to sharing more detailed information with you about these project accomplishments in this report.

About This Report:

This report provides summary information about the goals, activities, and results of this project, focusing on January 2008-December 2011. Section One will provide a brief summary of the outcomes of the project work based on our project’s performance objectives. Section Two will offer a brief reiteration of our project’s original goals and ambitions and an in-depth discussion of our findings and development activities. Section Three will describe plans for the program’s continuation and its long-term impact. Section Four will provide a list of the staff who have worked on various aspects of this project during the project period.

Again, we greatly appreciate the NHPRC’s support of this important “Cooperative Network and Service Providers” initiative on distributed digital preservation.
Section 1: Project Outcomes

The project proposed nine major deliverables, each of which comprised a project goal and was associated with one of the project phases. Below, we briefly document the work completed toward the successful achievement of each of these deliverables. We then provide more in-depth information about the work undertaken in each of these project phases in Section Two.

1.1 Administrative Activities and Business Planning

**Outcome:** Timely submission of complete reports and three copies of grant products as specified in NHPRC guidelines and the grant notification letter.

**Project Results:** The project’s Principal Investigators have completed and submitted all project reports and products.

**Outcome:** Completion and dissemination of the formalized charter, the Preservation Business Model, and the administrative relationship between Educopia, Inc. and the MetaArchive Cooperative.

**Project Results:** The MetaArchive Cooperative Charter (and its affiliated documents, the Membership Agreement and Technical Specifications), the Preservation Business Model, and the administrative relationship between Educopia and MetaArchive have all been formalized and disseminated.

**Outcome:** Establishment of an effective cost mechanism to allow Educopia to provide ongoing services to the members of the MetaArchive Cooperative.

**Project Results:** The project team implemented, tested, and refined the cost mechanism for Educopia’s sustainability. Educopia has committed to ongoing support for the hosting of the MetaArchive program for the foreseeable future.

**Outcome:** Successful internal audit of the digital preservation model using TRAC, DRAMBORA, or the Nestor framework.

**Results to date:** The project team conducted a TRAC audit with an external evaluator in 2009 and published the results in 2010. The MetaArchive Cooperative was found to successfully conform to all 84 TRAC criteria. A checklist version of the final report is available for download at: [http://www.metaarchive.org/documentation](http://www.metaarchive.org/documentation).

1.2 Outreach Efforts

**Outcome:** Conduct five formal workshops targeted for five communities: archives, museums, libraries, state digitization groups, and state and local government agencies; evaluative results that reflect that appropriate customization of the workshops for each community and that the effectiveness of the workshops increased over time.

**Project Results:** Over the course of the program, we have hosted eleven successful workshops that reached our targeted communities. (Dallas, TX 2008; Atlanta, GA 2008; Pittsburgh, PA 2009; Houston, TX 2009; Austin, TX 2010; Boston, MA 2010; Arlington,
Outcome: Recruitment of at least 20 U.S. institutions from universities, state archives, museums with archival records, and historical societies.

Project Results: We have recruited a total of 40 new institutions (including 30 participating institutions at our Collaborative Member sites), 10 of which are U.S. institutions. (Rice University, Boston College, Clemson University, Folger Shakespeare Library, University of North Texas, University of South Carolina, Indiana State University, Penn State University, the Historically Black Colleges and Universities Library Alliance (with 22 participating consortial libraries), and Oregon State University).

Outcome: Publication of at least one peer-reviewed article and the book A Guide to Distributed Digital Preservation.

Project Results: We have published six articles, five of which were peer-reviewed (Against the Grain, Archiving 2009, Library Trends, Library Hi Tech, ACM), and we published A Guide to Distributed Digital Preservation in February 2010, which has received 1800+ unique downloads. Please see http://www.metaarchive.org/reading-room for more details.

1.3 Technical Framework

Outcome: Replication of the MetaArchive’s network structure by at least three other groups.

Project Results: More than 10 “Private LOCKSS Networks” are now in existence, all of which have based their work upon our model. We have worked closely with two of these (ADPNet and Data-PASS), and we continue to talk with new groups that are interested in running a PLN.

Outcome: Development of a functioning technical bridge between the MetaArchive LOCKSS virtual private network and grid-based systems.

Project Results: We have worked successfully with the Chronopolis team of the San Diego Supercomputer Center, using the BagIt tool (developed by the Library of Congress and the California Digital Library) to exchange content to and from the MetaArchive PLN and the SRB/iRODS instance run by Chronopolis. The successful technical outcomes of this process are documented in Appendices B, C, D & E.
Section 2: Project Ambitions and Findings

2.1 Business Planning and Administrative Activities

Over the course of this project, the MetaArchive Cooperative successfully transitioned from a successful University-based project to a thriving consortium hosted by a neutral (non-university) nonprofit, the Educopia Institute. The work we undertook to complete this major project goal included formalizing documentation for both MetaArchive (e.g., the Charter, Membership Agreement, and other key legal and descriptive components) and Educopia (e.g., the business plan, which highlights the role Educopia plays as a sustainable host institution with long-term aims to support the ongoing development of the MetaArchive community).

As part of this transition, we also moved the core staff from a member site into the Educopia Institute, building all of the necessary administrative apparatus to successfully cultivate these staff members. In the process, we became one of the pioneers of the distributed virtual organizational model (joining others in our community who made this transition in the same time frame, such as DuraSpace). Our core staff members, as employees of Educopia, work from home offices in North Carolina, Michigan, and New Jersey. We have now worked together for several years using daily videoconferencing and regular in-person meetings to ensure our sense of connectedness and shared vision for MetaArchive and Educopia. This transition of our core staff to a neutral nonprofit host has served to further stabilize the Cooperative such that it is not dependent on any single member institution for its central activities.

We also formalized our overall administrative structures for the Educopia Institute and the MetaArchive Cooperative with the assistance of our business consultant and our pro bono legal team at Kilpatrick and Stockton. This structure has since allowed us to efficiently handle billing, payment schedules, processing payments, and other important administrative tasks associated with running a service organization. Due to these efforts, we were able to apply for several grants and subcontracts during 2010, including a successful National Endowment for the Humanities grant that Educopia is hosting to study digital newspaper preservation and a successful subcontract on an Institute for Museum and Library Services award attained by the University of North Texas to study ETD preservation.

To ensure that MetaArchive does not depend solely upon membership fees for its ongoing success and upon grants and contracts for its research and development work, we also developed a second program, Educopia Consulting. Through this venture, MetaArchive staff members gain the opportunity to continue building their own skillsets beyond the LOCKSS-based network that we host, through ongoing engagement with many other preservation methodologies and services on behalf of our clients. This program has experienced great success, including consulting projects with the Minnesota Digital Library, the CLOCKSS network, the New Orleans Jazz and Heritage Foundation, and numerous university libraries. The financial success of this work has provided us with unrestricted funds that we are able to direct toward ongoing development work as needed.
As another essential part of our administrative work on behalf of the MetaArchive Cooperative, we also undertook several audits, including a formal TRAC audit conducted by an outside consultant. We are pleased to report that our audit findings have been overwhelmingly positive (for example, we were shown to successfully conform to all 84 TRAC criteria) and have helped us to continue enhancing and maturing our documentation and workflows for the preservation network.

More details on each of these developments are provided below and in the appendices of this document.

2.1.1 MetaArchive Charter and Membership Agreement

We undertook the first major revision of the MetaArchive Charter (and its affiliated documents, the Membership Agreement and Technical Specifications) in 2008, and began welcoming new members at the end of that year. The changes included the creation of a new “Affiliate” membership category, significant reworking of the Contributing Membership roles and responsibilities, and fine-tuning of numerous details within both documents. The Cooperative conducts annual reviews of the Charter and its appendices at the Steering Committee meeting each year. We have made significant changes to our membership categories since 2008, including the dissolution of the “Contributing Membership” and “Affiliate” membership categories and the addition of a “Collaborative Membership” category. These changes have enhanced our ability to serve a broader community, as evidenced by the 30 institutions that are now actively preserving content through MetaArchive under the “Collaborative” category that could not have joined us under the earlier categories. Our organizational model has been cited numerous times as a model for other initiatives, and several groups have turned to Educopia and MetaArchive for advice on how to create a sustainable infrastructure for successful digital library endeavors.

2.1.2 Educopia and MetaArchive Organizational Structure

Throughout the project, we have formalized and matured our administrative structures for the Educopia Institute and the MetaArchive Cooperative with the assistance of our business consultant and our pro bono legal team at Kilpatrick, Townsend, and Stockton. This structure has allowed us to efficiently handle billing, payment schedules, processing payments, and other important administrative tasks associated with running a service organization. The Cooperative gained its first full-time staff member at Educopia Institute in October 2009 and three additional key staff members in January 2010. In order to do so, the Educopia Institute engaged in a number of activities, including securing benefits, setting up payroll, creating internal processes for hiring and evaluating staff, and filing various state and federal reports. Formalizing this administrative structure has also enabled us to process memberships for incoming institutions, establish a billing schedule for our current and renewing members, and file annual tax returns and 5500 reports. Over the project period, we also worked with our Board to establish a unified and strong vision for our development. As part of this work, we completed our first Strategic Plan for the organization (see Appendix A). The level of maturity we have attained in a short three-year period is largely due to the focus this project allowed us to place upon the nonprofit’s administrative infrastructure. This strong foundation that we have built during this project will be essential to our ongoing success.
2.1.3 Preservation Business Model

In early 2008, we undertook a preliminary study of the Educopia/MetaArchive Cooperative business structure with the assistance of Dr. Dominic Thomas, a faculty member of the Goizueta Business School at Emory University. With a team of students, Dr. Thomas helped us to complete an initial industry and competitive analysis, a walkthrough audit, and a process model for our business. The study helped us to identify the documentation and procedural work that we need to undertake with our business consultant in the next two and a half years of the project, including clarifying the roles of Educopia and MetaArchive for potential funding entities. Building off this study, we developed the Preservation Business Model plan, which included both refinements of current documentation and the preparation of new documents geared toward both internal use and for distribution to prospective and current members on a requested basis. The plan included the refinement of a document set prepared during our concurrent work with the Library of Congress—namely a Strategic Plan, Financial Plan, Operations Plan, Management Plan, Marketing Plan, and Risk Assessment. As key deliverables, each of these documents are briefly described below and are available in full as Appendix A to this final project report:

- The **Strategic Plan** formalizes Educopia’s mission, vision, accomplishments, challenges, and goals. The Educopia Strategic Plan was ratified by the Educopia Institute Board of Directors and approved by the MetaArchive Steering Committee in early 2010.

- The **Financial Plan** covers expectations for Educopia’s growth and how this growth will impact the organization’s revenue streams and anticipated expenditures for its programs. It includes an overview of Educopia’s market and programs, its projected revenue, expenditures, and future operations.

- The **Operations Plan** documents Educopia’s organizational structure, as well as how the Institute monitors and controls the quality of its two current programs, Educopia Consulting and the MetaArchive Cooperative. The Operations Plan goes into specific detail about these programs’ workflows and activities from the perspective of a member or client and outlines the policies and procedures for expansion or changes for both programs. It also describes the Educopia’s compliance with industry standards and quality control measures.

- The **Management Plan** details the ownership structure of Educopia and its programs, the internal and external human resources required for operations, and the human resources Educopia anticipates that it will need in the future to ensure sustainability.

- The **Marketing Plan** analyzes Educopia’s primary markets of libraries and archives, but also addresses the urgency of addressing the needs of museums, state and federal government agencies, historical societies, and data research centers. The Marketing Plan outlines specific measures that will be undertaken to reach these markets effectively through Educopia’s competitive advantages.

- The **Risk Assessment** analyzes the key elements that impact Educopia’s performance, sustainability and accountability. It includes an overview of Educopia’s organizational attributes, governance and organizational viability, organizational structure and staffing, financial sustainability, and various liabilities.
2.1.4 Educopia Consulting

In keeping with Educopia’s mission and vision, Educopia Consulting was established to provide fee-based workshops, presentations, and other services to cultural memory organizations on the collaborative use of cyberinfrastructures to support scholarly creation, preservation, and dissemination. Since December 2009, Educopia Consulting has worked with Lyrasis, the Folger Shakespeare Library, the University of North Texas, the Minnesota Digital Library, the Indiana State Library, the New Orleans Jazz & Heritage Festival & Foundation, and the CLOCKSS Network to generate funding that serves as unrestricted funding that Educopia may use for ongoing support of the MetaArchive Cooperative.

2.1.5 Steering Committee Meetings

In April 2008, the Steering Committee held a semi-annual meeting in Atlanta, Georgia. At the meeting, the MetaArchive Steering Committee and their business and legal consultants pinpointed revisions for the charter and conceptual model and template, formed the Outreach Committee, and drafted an outline of the activities that will be undertaken by the Outreach Committee during program implementation.

In October, 2008, the Steering Committee met to finalize the Charter and Membership Agreement, review the Outreach Program Implementation Plan, review our technical development planning efforts, and review the administrative structure we had established for billing and payment.

In April 2009, the Steering Committee held a semi-annual meeting to further evaluate our outreach work, approve the market analysis and market plan, review the team’s progress in moving the Guide to Distributed Digital Preservation toward publication, and to be debriefed on the technical work completed between LOCKSS and SRB by the MetaArchive and Chronopolis teams.

In October 2009, the Steering Committee met to approve the 2009-2010 Charter, Membership Agreement, and Technical Specifications, approve the semi-final version of the TRAC self audit, update committee assignments (Content, Preservation, Technical Committees), and propose the next steps on an affiliate and consortial membership.

In April 2010, the Steering Committee held a semi-annual meeting to approve the Strategic Plan, gather feedback on the technical exchange mechanism, perform a New Member Packet update and outline any necessary additional materials, review our Outreach Program Implementation Plan, and perform and initial GDDP impact assessment.

In October 2010, the Steering Committee Meeting met in Boston, MA to approve the 2010-2011 Charter, Membership Agreement, and Technical Specifications, review and approve progress on the NHPRC grant deliverables, progress on technical developments, as well as the development of the final Preservation Business Model documents. Of particular importance for recruitment purposes were adjustments to pricing and membership levels, as well as the formation of strategies that are now documented thoroughly in the Educopia Marketing Plan. These include targeted marketing campaigns to LOCKSS Alliance and Networked Digital Library of Theses & Dissertations members to leverage their existing fees & discounts to encourage joining
the Cooperative.

In April 2011, the Steering Committee held a semi-annual meeting by videoconference to review and approve progress on all of the NHPRC grant deliverables, including the technical development work, the continued development of the Preservation Business Model documents, and the process and timeline for the targeted outreach campaign.

In December 2011, the Steering Committee met in Arlington, VA to approve the 2011-2012 Charter, Membership Agreement, and Technical Specifications, review the list of prospective members that expressed interest during our targeted outreach campaign, and perform a final review and approval of all NHPRC grant deliverables, including the complete Preservation Business Model.

2.1.6 Membership Adjustments

The MetaArchive Cooperative continues to attract a high degree of interest from coordinated groups of institutions that collaboratively produce and maintain a single, centralized repository. Beginning in early 2010 the MetaArchive Steering Committee began formally exploring the possibility of adding a new membership category to include such collaborative groups. At that point, the Cooperative had extended membership opportunities only to individual institutions that possessed the resources and technical staff to host the necessary infrastructure. The ideal scenario given the MetaArchive’s participatory approach, would be that the “Member” would be the collaborative’s central organization, not the individual institutions that participate in that collaborative. That “Member” would act as all other members do, namely hosting and maintaining a MetaArchive-LOCKSS cache for the network and writing plugins for its jointly maintained collection(s).

After much discussion and review at the December 2010 annual Steering Committee meeting, the MetaArchive Steering Committee was pleased to finalize and announce the creation of the “Collaborative Member” category. The Steering Committee has set this “Collaborative Member” category pricing at $2,500/year plus a small fee per year for each participating institution. This membership category has provided a low-barrier of entry for small institutions in need of digital preservation services—a key institutional type that has long been of concern to the MetaArchive’s membership.

In addition, to further stimulate recruitment and ensure sustainability of the MetaArchive Cooperative and the related services of Educopia, the MetaArchive Steering Committee approved at its 2010 annual meeting the following adjustments to the MetaArchive membership levels and pricing:

1. Increased the space fees from $2/GB/3-year term to $1/GB/year in order to encompass both at-cost space and administrative overhead incurred with content ingest and monitoring;
2. Increased the Sustaining Membership fee from $5,000/year to $5,500/year and reduced the break in pricing for institutions that pay for all three years during the first 6 months of their contract to 10% (it was previously 20%);
3. Increased the Preservation Member pricing to $3000/year in order to bring it in line with other membership options;
4. Eliminated the limit for Steering Committee members (which was at 15 members), effectively freeing the Cooperative to continue expanding its top category of membership in the coming year; and

5. Allowing one nonvoting representative from the Preservation members and one nonvoting representative from the Collaborative members to participate at the annual Steering Committee meeting.

Each Steering Committee representative agreed that the above mentioned changes to representation on the Steering Committee and the new Collaborative member category would be instrumental in improving recruitment and that the modest increases to space charges and membership fees would be highly effective in underwriting operational costs for Educopia's programs. No adjustments were made to membership levels and pricing at the 2011 annual Steering Committee meeting.

2.1.7 Audits and Assessments

Beginning in May 2009, we undertook a DRAMBORA assessment and a Trustworthy Repositories Audit and Certification (TRAC) audit. Both of these preservation assessment tools have been deemed of great value to those engaging in digital preservation services. Particularly helpful was the TRAC self audit, for which the MetaArchive contracted with an external evaluator. Led by University of Michigan graduate Matt Schultz, individual interviews and a comprehensive documentation review were conducted between July and August 2009. Initial results were prepared by Schultz in collaboration with the MetaArchive Cooperative in September 2009 and were evaluated by the Steering Committee at the annual meeting in Houston, Texas, in October 2009. The MetaArchive was found to conform to all 84 criteria. A checklist version of the final report is available for download at: http://www.metaarchive.org/documentation.

Beginning in January 2010, the MetaArchive Cooperative central staff worked with members (particularly those serving on the Preservation and Content Committees) to develop new supportive documentation and formulate necessary policies and procedures in accordance with the self-assessment's findings and recommendations. These included an Ingest Procedures document, an Authorizations and Authentication Policy, an Access Policy and a Content Policy. Each of these documents were formally reviewed and approved by the Steering Committee at the 2011 annual Steering Committee meeting.

2.2 Outreach Efforts

A key goal of this project was to increase our visibility within the community and to attract new members who shared our general philosophy: that libraries and other memory institutions should wherever possible, embed both the technical infrastructure and knowledge regarding digital preservation within their own organizations.

The project's outreach efforts over the course of the project have been tremendously successful. Both the core staff and the membership of the MetaArchive Cooperative made presentations at many of the digital library field's major events, including CNI, iPRES, DLF,
SAA, and ALA. We also delivered workshops to a broad range of participants, including administrators and technical staff members of academic libraries, state government agencies, state libraries, county libraries, museums, historical societies, and nonprofit organizations.

We also produced a number of publications, including *A Guide to Distributed Digital Preservation* (an edited volume authored by our membership), articles in leading journals including *Library Hi Tech* and *Library Trends*, an ARL Report (“New Roles for New Times,” by Tyler Walters and Katherine Skinner), and an ARL SPEC Survey (#325, by Gail McMillan, Matt Schultz, and Katherine Skinner).

We also engaged with other digital preservation and collaborative digital library groups on many initiatives. We forged strategic alliances with the Chronopolis team at SDSC, the Networked Digital Library of Theses and Dissertations, and the Data-PASS preservation initiative of ICPSR.

At the end of the project term, we embarked on the first phase of a concentrated outreach effort that included a mailing campaign to all LOCKSS member institutions. We hope to see these institutions choose to leverage their investment in LOCKSS (through their existing memberships in the LOCKSS Alliance for e-journal preservation) through working with MetaArchive to preserve their unique holdings in our community-governed network.

Further details about our outreach efforts are documented below and in Appendix A of this report.

### 2.2.1. Outreach Program Implementation Plan

At the beginning of the project the PIs worked with the Outreach Committee to create an Outreach Program Implementation Plan. Several sections from the original project proposal provide context for this plan. The plan included four main areas of activity: presentations, workshops, writing, and consulting. The Cooperative benefited from having this plan to help structure the presentations, workshops, and writing projects that the core staff and distributed network of active members conducted over the project period.

### 2.2.2. Presentations and Workshops

To promote awareness of the MetaArchive Cooperative and the importance of collaboration in digital preservation, the central staff and membership delivered over 30 presentations and conducted eleven workshops. For a full list of presentations and workshops delivered, please see [http://www.metaarchive.org/events](http://www.metaarchive.org/events).

### 2.2.3. Publications

We published *A Guide to Distributed Digital Preservation* in February 2010. This book has since played a major role in the field, including helping to foster the development of several other Private LOCKSS Networks. Other distributed preservation groups, including the DICE group at Chapel Hill, have cited the relevance and importance of the publication to their own work. Authored by our membership, the book focuses on the broad issue of engaging in collaborative networks in order to preserve digital collections, and uses the PLN example as a concrete case study of what Distributed Digital Preservation principles look like in action. To date, this publication has received 1,800+
unique downloads from around the world. Please see http://www.metaarchive.org/reading-room for more details.

2.2.4. Recruitment and Renewals
Between January 2008 and December 2011, 40 new institutions (including 30 participating institutions at our Collaborative Member sites) joined MetaArchive (Rice University, Boston College, Clemson University, Pontifícia Universidade Católica do Rio de Janeiro, Folger Shakespeare Library, University of North Texas, University of South Carolina, Indiana State University, Penn State University, Consorci de Biblioteques Universitàries de Catalunya (with eight participating consortial libraries), the Historically Black Colleges and Universities Library Alliance (with 22 participating consortial libraries), and Oregon State University). Seven members also renewed their memberships for a second three-year term (Georgia Tech, Auburn University, University of Louisville, Hull University, Florida State University, Boston College, and Virginia Tech). We also worked with more than 20 prospective members, many who came forward as a result of our targeted outreach campaign (see below), to explore their membership options. We are excited about the expansion of our network and feel certain that we have laid a solid foundation for future growth during this project.

2.2.5. MetaArchive and NDLTD
In late 2008, we entered into an alliance with the Networked Digital Library of Theses and Dissertations (NDLTD) to engage their members in our preservation services at a reduced membership rate (5% discount). This alliance immediately contributed to the recruitment of two Sustaining Members (Rice University and Boston College), one Contributing Member (Pontifícia Universidade Católica do Rio de Janeiro), and one Collaborative Member (HBCU Library Alliance), and remains a core offering for our targeted outreach campaign (see Appendix A). The NDLTD also continue to partner with the MetaArchive on grant related research and development. The NDLTD is currently a principal project partner on our IMLS-funded Lifecycle Management for ETDs project, which seeks to study, document, and engage the academic community on the curation and preservation needs of ETDs.

2.2.6. Outreach Campaign
As mentioned in our Marketing Plan (Preservation Business Model), Educopia has largely relied upon event outreach and publications to promote the MetaArchive Cooperative. This year, it pursued lightweight and targeted marketing strategies to expand its membership more robustly. Outreach began in June 2011 and was initially aimed at both U.S. and international institutions that are already members of the LOCKSS Alliance. This group of institutions is already paying LOCKSS Alliance fees, and via joining MetaArchive, they may leverage their membership for multiple purposes. Another core group of interest to MetaArchive this year is the Networked Digital Library of Theses and Dissertations. All of its members are managing digital theses and dissertations, and many of them are going “e-only” with their ETD collections. This type of content is of great importance, not just to the libraries, but also to their parent institutions, as this is a key component of each university’s research output. As such, we believe that this genre will be a leading “tipping point” marker for university library engagement in preservation activities. We have undertaken a successful pilot project with the NDLTD for the past two years. Educopia will build upon this relationship this year by actively recruiting new members from this group. It will also continue
to offer a joint-membership discount with NDLTD of 5% to all who are members in good standing in both groups.

2.3 Technical Framework

This project sought to establish a foundation for AIP exchange and interoperability between two of the world’s leading digital preservation approaches, iRODS (through the Chronopolis service) and LOCKSS (through the MetaArchive service). Our partnership with the Chronopolis team has been strong and the two groups have shared a great deal of technical and organizational work over the course of this project that has enhanced each project’s ongoing sustainability.

The cornerstone of the project’s technical accomplishments has been the successful exchange of preserved content to and from the Chronopolis and MetaArchive environments. Such content exchanges between repositories are fraught and complex, as documented in other endeavors including the now infamous “Library of Congress Archive Ingest and Handling Test” (AIHT) study and the FCLA’s “Towards Interoperable Preservation Repositories” (TIPR) project. To assess the options for this work, we drew together a group of experts including Tom Habing (HANDS project), Bill Underwood (GA Tech Research Institute, NARA-based work with iRODS), and Cliff Lynch (CNI) to advise on the scope of this project undertaking.

We started the exchange tests in 2008 using BagIt to exchange content between Chronopolis’s SRB (predecessor to iRODS) and MetaArchive’s LOCKSS network. By 2011, we successfully conducted exchanges between MetaArchive’s LOCKSS network and Chronopolis’s iRODS system. In so doing, we bridged the two preservation environments at the levels of ingest and archival storage. This lays the foundation for future collaborations between Chronopolis and MetaArchive, and by extension, fosters such collaborations between a wide variety of other iRODS and LOCKSS-based groups.

Following the success of these exchanges, we turned our development energies toward developing and packaging alternative data models and microservices that could be layered with LOCKSS-based or other preservation frameworks. We also drafted an open-source replication specification for exchanging BagIt-based content and collections.

More information about each of these accomplishments is documented below and in the appendices of this report.

2.3.1. MetaArchive-Chronopolis Data Transfer

Beginning in late 2008 we began testing the first round of exchange mechanisms for data transfers between Chronopolis and their SRB client and LOCKSS using the BagIt tool (developed by the CDL and Library of Congress). Upon transfer of the project’s no-cost extension to University of North Texas in early 2010, we were able to work with one of their software developers to build upon the work done to date and extend it in coordination with the Chronopolis team, as they moved their infrastructure from SRB to iRODS. By January 2011 we concluded all activities and were able to report that we had: 1) successfully completed the first and second phases of content exchange with the Chronopolis team at the San Diego Supercomputer Center; 2) completed documentation for the retrieval,
validation, and packaging of content on the LOCKSS file system for contingency planning purposes; and 3) completion of the exchange of content from the Chronopolis environment with the MetaArchive-LOCKSS environment through the use of standard LOCKSS harvesting protocols. With this set of work we effectively bridged these two heterogeneous environments at the level of ingest and archival storage. The Logical Data Record from the MetaArchive perspective is an archival unit (or AU), which represents a grouping of files that can be routinely processed and verified for fixity by the LOCKSS software. This transfer process involved the packaging of 246 archival units of 1.3 terabytes of master TIFF images from the Folger Shakespeare Library (used with permission), and 1 archival unit of less than 1 terabyte of multi-format research data files from Data-PASS (used with permission). Please see Appendices B, C, D & E for a full description of the MetaArchive-Chronopolis data transfer process—including details on content validation.

2.3.2. Data Management & Microservices Development
Beginning in late 2010, due to the successful completion of our data transfer initiatives described above, the MetaArchive and UNT project staff began drafting a workplan for the development of a micro-services approach to data management that could be layered on top of our LOCKSS-based network. Beginning in early 2011, the Project Manager and technical staff began experimenting with PREMIS-derived data models and microservices (format validation, fixity checking, virus checking, etc.) that could enhance the MetaArchive’s ability to track and validate digital objects at various stages of their lifecycle. To date, the software developer at UNT has crafted and tested a suite of Python-based modules that can facilitate the recording of outputs from various preservation microservices into PREMIS Event records. These can then be queried via an AtomPub feed and managed through a Django web administration user interface. These Python and Django components have been successfully installed on a MetaArchive development server and are being inventoried for various software library dependencies. Several stages of the development have been documented (See Appendix F).

2.3.3. Replication Services Development
Educopia hosted a planning meeting in 2010 to investigate opportunities to collaborate across technical teams to develop a common open-source replication specification for exchanging BagIt-based content and collection. Development of such a replication specification would enable Educopia and its MetaArchive technical framework to interact in more flexible ways with both content producers and service providers to ensure a broader range of digital preservation options. Educopia fostered development activities throughout 2011 between the Library of Congress, Penn State, Chronopolis, CDL, and several other groups. To date we have reached a near-final specification that can be viewed at: https://github.com/acdha/restful-bag-server. The consensus of the principal collaborators is that the current specification should be routinely promoted and encouraged on the Digital Curation list and described on the Educopia Institute website so that groups can have time to evaluate their current infrastructures and explore integrations as they make sense. The MetaArchive Cooperative may have a near-term use case for this replication mechanism in other related project research and development.

2.3.5. MetaArchive Contingency Plan Outline
Now that we have successfully tested, finalized, and documented exchange mechanisms for data transfers between the Chronopolis SRB/iRODS clients and the MetaArchive-LOCKSS network using BagIt and various custom scripts, we have begun outlining the formal
MetaArchive Contingency Plan to demonstrate sustainability in any anticipated succession scenarios. This document or set of documents will be produced in coordination with our MetaArchive Preservation Committee and Chronopolis throughout 2012 and will adhere to community best practices for disaster planning and continuity of business operations encouraged by such organizations as ICPSR and NIST (See Appendix G for a semifinal draft outline).

Section 3: Continuation and Long-Term Impact

We are very pleased to report that today, the MetaArchive Cooperative is a stable program with a sound sustainability plan. Its host institution helps to provide a secure foundation for the program through a diversified revenue stream comprised of membership and storage fees, consulting income, and grant funding. The MetaArchive membership continues to operate as a true cooperative, with every institution contributing staff time and infrastructure as part of its commitment to the overall network. The network demonstrably scales well, both in terms of space (via additional caches at new member sites) and in terms of content management practices. We plan to build upon our initial success through an expansion of our preservation offerings in 2013 to include optional micro-services we are developing for our membership. We are also studying the lifecycle management and preservation issues that are specific to two genres of material, digital newspapers and electronic theses and dissertations. On these projects and in our work more generally, we are developing and deepening our strategic alliances with other preservation practitioners, including Chronopolis, DPOE (Library of Congress), NDSA, DuraSpace, Data-PASS, and the LOCKSS Alliance.

We believe that the practice of digital preservation is still emerging. Judging by surveys we have conducted and our experiences with prospective members, the memory community currently is just beginning to plan toward the preservation of their digital assets. Most institutions are not actively preserving content at this time. Even within some of the field’s exciting new developments, including the formation of the ARL-library driven “Digital Preservation Network” (DPN), more institutions are actively pursuing alliances with each other to study and address preservation issues in the future rather than actively engaging with existing preservation services. Over the next 3-5 years, we expect there will be a “tipping point,” in which a larger number of institutions (particularly research libraries and other large content stewards) will begin to engage in preservation activities.

The maturation of the field will come with time. Our challenge (which we share with all of the other emerging preservation services, including FCLA DAITSS, UC3 Merritt, DuraCloud, Chronopolis, and vendor-based offerings such as Tessella, Rosetta, and OCLC Digital Archive) is to stage our growth appropriately so that we will be ready to meet the future demand. Our ability to scale and grow quickly is facilitated by our distributed model. So long as we have the ability to orient and involve new members appropriately as they join, both our organizational and technical model can quickly scale to meet whatever demand we do encounter in the coming years.

The generous support provided by the National Historical Publications and Records Commission for the MetaArchive’s transition period has provided the project team with the necessary space, time, and financial resources to properly plan and implement our
administrative infrastructure. Most businesses that fail do so because they did not adequately study and design their organizational processes. As a direct result of this project, both Educopia and MetaArchive are strong, stable, and poised for future growth.

Section 4: Project Staffing

The official project staff included dedicated personnel at the University of North Texas, Educopia Institute, University of Louisville, Auburn University, Florida State University, Georgia Tech, Virginia Tech, and the San Diego Supercomputer Center’s Chronopolis group.

In addition to those explicitly named below, the MetaArchive Cooperative’s Steering Committee and membership played pivotal roles throughout the project work to forward our outreach, documentation, and network development. We appreciate all of the time and energy that this group invested over the course of the project period, and we applaud the results of their work.

Those that have worked intensively on this project throughout the 2008-2011 period include the following:

**University of North Texas**
- Martin Halbert (PI)
- Cathy Hartman
- Mark Phillips
- Kurt Nordstrom

**Educopia**
- Katherine Skinner
- Matt Schultz
- Monika Mevenkamp
- Bill Robbins

**Chronopolis (SDSC)**
- David Minor
- Don Sutton

**Auburn University**
- Aaron Trehub
- Beth Nicol

**Florida State University**
- Becca Bichel

**Georgia Tech**
- Tyler Walters
- Kent Woynowski

**LOCKSS**
- Vicky Reich
- Tom Lipkiss

**University of Louisville**
- Carrie Daniels
- Dwayne Buttler

**Virginia Tech**
- Gail McMillan