Appendix A: MetaArchive Technical Specifications

This document provides an overview of the current recommendations and requirements for administering a preservation cache for the MetaArchive Cooperative, including staffing and hardware. This document was last revised on November 4, 2010.

1. Skills Recommendations

Our Member institutions have identified three key roles that they assign to their local staff members in order to effectively run their caches and prepare their content for ingest into the network: cache administration, plugin development, and data wrangling. It is possible that a single technical staff member may be responsible for one or more of these roles. For example, the staff assigned to plugin development and data wrangling may be the same person (if not they should, if possible, work closely together). The MetaArchive staff provides training for these roles annually. The anticipated time commitments, skill-sets, and common tasks, as based on current Member experiences, are documented below.

1.1 Cache Administration

Time required	Between 2-10 hrs/mo (average 5 hrs/mo)
Required skills	Basic level administration of UNIX-based platforms; ability to run and maintain servers, proxy servers and firewalls
Helpful skills	Knowledge or experience in digital libraries or library IT
Common tasks	Installing a MetaArchive-LOCKSS cache; assisting with content ingest; performing updates for the cache; monitoring the cache; documenting procedures

1.2 Plugin Development

Time required	Between 2-24 hrs on 1st plugin (average 13 hrs) Between 1-5 hrs on additional plugins (average 3 hrs)
Required skills	Familiarity with XML; familiarity with file structuring on widely used platforms (Windows/Unix/Linux); understanding of regular expressions; solid understanding of web technologies (e.g., browsers and plugins)
Helpful skills	Familiarity with metadata standards; programming experience
Common tasks	Writing/testing plugins

1.3 Data Wrangling

Time required	Between 15-40 hrs per collection (depends on existing repository solution)
Required skills	Familiarity with file structuring on widely used platforms (Windows/Unix/Linux); basic understanding of web technologies (e.g., web servers)
Helpful skills	Experience with re-formatting digital content and media; experience with archival appraisal and selection methods; familiarity with metadata standards and cataloging
Common tasks	Creating manifest pages; re-naming and re-sizing files; preparing web servers to deliver content; creating collection level metadata

2. Operational Requirements

2.1. Preparing the Technical Environment

- Member system administrators (or designated technical staff members) should have ready access and authorizations to access their MetaArchive-LOCKSS caches to the fullest extent possible.

- Member system administrators (or designated technical staff members) should have the ability to effectively coordinate with staff members that are responsible for configuring institutional firewalls to allow MetaArchive-LOCKSS caches to participate in the MetaArchive preservation network.

2.2. Necessary Cost Expenditures

- Designated Members must purchase hardware that meets the specifications below to operate a MetaArchive-LOCKSS cache.

- Member institutions must be prepared to adequately staff the necessary roles (see Skills Recommendations above) to implement and maintain a MetaArchive-LOCKSS cache throughout the period of their membership.

3. Support and Equipment Life Cycles

3.1. Member Obligations

- Members agree to purchase and maintain the necessary technical hardware (as described below) required to operate a MetaArchive-LOCKSS cache throughout their membership period.

- Members also agree to update their technical hardware on a three-year cycle using the current MetaArchive cache specifications. This ensures that all of the MetaArchive network's equipment is replaced in a manner consistent with industry best practices. This rolling cycle also enables the Cooperative to avoid network-wide uniformity of technical hardware.

3.2. Replacement Option

- In the case of catastrophic circumstances, members have the ability to request technical and financial assistance with the restoration of a preservation site's servers, software, and collections by the MetaArchive Cooperative. These requests will be reviewed and, at the discretion of the Steering Committee, either approved or denied.

4. Technical Specifications

Required	Recommended	Notes
Machine architecture capable of running an operating system with a Sun or Sun-compatible JVM	Intel Core i7 processor and Intel Core i7 Extreme Edition compatible	(Quad Core Processors based on this architecture is the current standard for new LOCKSS caches)
	Rack-mountable server chassis	Not a hard requirement, but standard for most PLNs.
At least 6 GB of RAM	6 or more GB of RAM	At a lower level, excessive swapping may occur.
16 TB, No RAID disk storage		(This (16TB) is the space configured on the latest generation of servers and is the minimum for a new production cache.)
RPM-based Linux Distribution	CentOS is recommended	Red Hat Enterprise Server is another option.
Java Virtual Machine version 1.6 or later		LOCKSS software requires a Java Virtual Machine; any version starting from 1.6 should run the LOCKSS daemon.
LOCKSS software		The Cooperative provides a link to the current version that is in use. All caches run the same version of the LOCKSS daemon and we synchronize any upgrades.
LOCKSS caches should be physically secure, accessible only to appropriate staff members, and climate controlled		Temperature of 40-80°F and humidity of 10-80%.
A firewall should be used to block access to all unused ports		Any services not required for the functionality and maintenance of a LOCKSS cache should be protected by a firewall. Appropriate ports to leave open include ports used by: LOCKSS network; 2) the Cooperative to communicate with the node itself; and 3) those for secure remote access. For Unix machines, ssh should be considered an appropriate method of remote access. Telnet and VNC are not considered secure methods.
User accounts should be kept to a minimum		Only the system administrators who need to maintain the server should have user accounts. These accounts should have strong passwords.
There should be no direct remote administrative access		In the case of SSH, this constitutes disabling root logins.
Security patches should be applied promptly	Always follow local site security update policies.	For users of the RPM-based Linux distributions, this can be achieved by periodically running yum or up2date for the latest software updates. In general this is handled by the creation of a cron job when the server goes through the Kickstart mechanism. Synchronized OS upgrades will be performed throughout the year in coordination with the MetaArchive system administrator.