

# MetaArchive Technical Specifications

## Introduction

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 by the Technical Committee in June 2021, with minor updates in March 2022.

### [Introduction](#)

#### [1. Skills Recommendations](#)

##### [1.2 Cache Administration](#)

##### [1.2 Plugin Development for Content Ingest](#)

##### [1.3 Data Wrangling](#)

#### [2. Operational Requirements](#)

##### [2.1 Preparing the Technical Environment](#)

##### [2.2 Necessary Cost Expenditures](#)

#### [3. Support and Equipment Life Cycles](#)

##### [3.1 Member Obligations](#)

##### [3.2 Replacement Option](#)

#### [4. Technical Infrastructure](#)

##### [4.1 Servers](#)

##### [4.2 Storage](#)

###### [4.2.1 Network Storage](#)

###### [4.2.2 Local Storage](#)

##### [4.3 Security](#)

##### [4.4 Networking](#)

## 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 MetaArchive Preservation 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 member assigned to plugin development and data wrangling may be the same person (if not they should, if possible, work closely together). The MetaArchive Central staff provides training for these roles as requested. The anticipated time

commitments, skill-sets, and common tasks, as based on current Member experiences, are documented below.

## 1.2 Cache Administration

MetaArchive recommends that *at least two people* from each member organization have SSH access to their LOCKSS cache.

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

## 1.2 Plugin Development for Content Ingest

Members that will be [packaging and ingesting their content as BagIt bags](#) will be able to use a generic bagit plugin that will require no additional development time or activity. Members who wish to directly integrate digital repositories and the MetaArchive LOCKSS network will need to write [custom LOCKSS plugins](#).

<b>Time required</b>	Between 2-25 hrs on 1st plugin (average 15 hrs) Between 1-6 hrs on additional plugins (average 3 hrs)
<b>Required skills</b>	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).
<b>Helpful skills</b>	Familiarity with metadata standards; programming experience.
<b>Common tasks</b>	Writing/testing plugins.

## 1.3 Data Wrangling

<b>Time required</b>	Between 15-40 hrs per collection (depends on existing repository solution).
<b>Required skills</b>	Familiarity with file structuring on widely used platforms (Windows/Unix/Linux); basic understanding of web technologies (e.g., web servers).

<b>Helpful skills</b>	Experience with re-formatting digital content and media; experience with archival appraisal and selection methods; familiarity with metadata standards and cataloging; programming experience; database management; experience using tools to create bagit bag packages.
<b>Common tasks</b>	Creating manifest pages; renaming and resizing 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

Members opting to host caches must purchase hardware, or utilize local IT services for server hosting, that meet 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 opting to host caches must agree to purchase and maintain the necessary technical hardware (as described below, either physical or virtual) required to operate a MetaArchive LOCKSS cache throughout their membership period.

The above Members also agree to update their technical hardware on a five-year cycle using the current MetaArchive Technical Specifications. This ensures that all of the MetaArchive Preservation Network's infrastructure 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. The five-year cycle applies to both physical and virtual servers.

MetaArchive members hosting caches agree to either repurpose their technical hardware for the MetaArchive test network when it has reached its five-year cycle (as long as it still functions) or to host a virtual server for the same purpose. Storage requirements for test caches are 500 GB. Caches re-purposed for the MetaArchive test network will not be subjected to any recovery actions in the event of disk failures. If a member has retired more than one cache, only its most recently retired cache should be part of the test network; older caches may be repurposed as needed by the member for other functions.

## 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 caches, software, and collections by the MetaArchive Cooperative. These requests will be reviewed and, at the discretion of the Leadership Team, in coordination with the Membership, either approved or denied.

# 4. Technical Infrastructure

## 4.1 Servers

Members may use either virtual or physical server hardware. All LOCKSS caches should be physically secured (physical servers and virtual server infrastructure), accessible only to appropriate staff members, and appropriately climate controlled. MetaArchive members are recommended to work with Central Staff and LOCKSS Support on deploying new caches.

- 4 processors for virtual servers; Intel Core i7 or better processor with at least 4 cores if physical server
- 8 GB RAM
- RHEL 7, CENTOS 7, ; RHEL 8, Rocky Linux 8 (minimal install group recommended)
  - CENTOS 8 is **NOT** recommended.
- Root-level (administrative) access is required to execute the [MetaArchive integration script](#)

## 4.2 Storage

Members may use network or locally attached storage to provision 40 TB of RAID protected storage. If the network reaches full capacity, members may be asked to provision additional RAID protected storage.

### 4.2.1 Network Storage

40 TB. While 40 TB is required, members may choose to scale up to that amount over time; when doing this, members must provision 12 TB to start and increase over time to reach 40 TB.

### 4.2.2 Local Storage

40 TB RAID protected storage (RAID6 recommended). RAID6 requires two hard drives for parity. For example, if using 8-TB hard drives, seven are required (56 TB raw) to meet 40 TB requirement.

## 4.3 Security

Only the system administrators who need to maintain the server should have user accounts; these accounts should have strong passwords. It's recommended to disable root logins if possible. Security patches should be applied promptly, following local site security update policies. One way this can be achieved is by using yum-cron.

## 4.4 Networking

LOCKSS caches in the MetaArchive network require a dedicated IP address that is routable from the Internet and can accept inbound connections from the internet. NAT is supported.

A firewall should be used to block access to all unused ports. Some ports are required to be open for communication in the MetaArchive network. LOCKSS requires certain inbound TCP ports to be open, outbound connections to all ports should be allowed. The Conspectus IP address needs to be allowed to access each member caches' LOCKSS user interface. See the [MetaArchive Technical Specifications, Firewall and Port Settings](#) for full details.