Michigan Digital Preservation Network (MDPN) LOCKSS Node Hosts - FAQs & Technical Requirements

What is MDPN?

[MDPN](#) is a membership-based organization dedicated to providing affordable and easy-to-use digital preservation services to cultural memory institutions, research organizations, and others that maintain electronic records or collections. The MDPN welcomes organizations of all sizes, technical capacity, and financial ability across Michigan. By concentrating resources and knowledge toward a single effort, the MDPN reduces redundant efforts in the state and builds capacity for more organizations to participate.

The MDPN was created by a 2019 Memorandum of Understanding between the Library of Michigan, Grand Valley State University, Michigan State University, the Midwest Collaborative for Library Services, and Western Michigan University. MCLS serves as the organizational home and fiscal agent for MDPN.

What is LOCKSS?

[LOCKSS](#) (Lots of Copies Keeps Stuff Safe) is an open source software for digital preservation based at Stanford University Libraries since 1999. LOCKSS software runs across networks made up of multiple locally-managed servers often called "nodes," where content is replicated and stored. Nodes collect content for preservation from a secure staging server using a Web crawler. Each node stores the same content as the other nodes on the network, so that if content on one node becomes corrupted or otherwise unusable, redundant copies of that content can be retrieved and replicated from other nodes in the LOCKSS network.

What are the technical principles behind the LOCKSS software?

LOCKSS preserves content on a continual basis through its Poll-and-Repair Protocol, also called the Library Content Audit Protocol (LCAP). Through LCAP, mutually-distrusting LOCKSS nodes operating in a network conduct randomized polls to confer on the integrity values of co-preserved content. The system uses cryptographic nonces to force re-hashing and prevent caching and replay of previously-computed integrity values.

Landslide agreement on integrity values provides reassurance that the preserved contact is intact and unchanged across the conferring nodes. Landslide disagreement on integrity values indicates that the copy of data held by minority dissenting node(s) is likely corrupted, and a repair from the source or another node is automatically prompted. In cases where a plurality of votes cluster around varying integrity values, no repair actions are undertaken and an alert is raised for the discrepancy to be investigated further.

Over time, a node's participation in polls and the results establish its reputation in the network, weighting it as more likely to be responsive and trustworthy for the co-preservation of the specific content in question. Multiple copies of data are read regularly but randomly, and repaired as needed.

This is visualized through the diagram below:

Why do you need node hosts?

The LOCKSS Poll-and-Repair Protocol requires at least three nodes be connected to the network in order to run a valid poll. The other reason to have multiple nodes is to help mitigate risk – having more copies in more locations spread across different geographic areas and electrical grids builds network redundancy and helps to protect the network against threats from natural disasters, power outages, or intentionally malicious attacks such as ransomware.

Is LOCKSS secure?

Yes! In a 2014 Trustworthy Repositories Audit and Certification (TRAC) of the CLOCKSS Network by the Center for Research Libraries, the LOCKSS software received a perfect score in the "Technologies, Technical Infrastructure, and Security" category. The LOCKSS software maintained this rating in a TRAC recertification of CLOCKSS in 2018.

Are there other networks that use LOCKSS?

Yes! There are dozens of networks that rely on LOCKSS for their preservation needs. Some of them are statewide or regional, like the Alabama Digital Preservation Network, while others are international like MetaArchive or CLOCKSS. Many of these networks have been using LOCKSS to preserve electronic journals, research data, institutional records, and cultural memory materials successfully for over a decade. This is in addition to the many individual members of the Global LOCKSS Network (GLN), including a number of universities here in Michigan.

Why did MDPN choose LOCKSS over other preservation strategies?
- Cost - LOCKSS networks are low-cost to maintain beyond staff time and periodic hardware refreshes. There are no expensive annual service agreements or contracts.
- Control – LOCKSS allows MDPN members to maintain local control over their collections and we at the MDPN likewise are able to maintain local control over the network.
- Values – LOCKSS is an open source, collaborative solution for digital preservation, which aligns with the MDPN's values of access, collaboration, and equity.
- Sustainability – the LOCKSS Program has been offering robust preservation services to a diverse group of international users since 1999 and is well-supported by Stanford University Libraries.
- Security – In a [2014 Trustworthy Repositories Audit and Certification (TRAC) of the CLOCKSS Network](#) by the Center for Research Libraries, the LOCKSS software received a perfect score in the "Technologies, Technical Infrastructure, and Security" category. The LOCKSS software maintained this rating in a TRAC recertification of CLOCKSS in 2018.
- Community - there is a robust community of users around the globe who are eager to share insight, resources, and experience in order to keep these community-run networks available as an alternative to corporate service providers.

What is required to host a LOCKSS node for MDPN?

Generally, managing a LOCKSS node requires having sufficient system expertise with the Linux environment to manage the install, and regular ongoing maintenance such as applying upgrades to the operating system.

The installation is based on application virtualization using Docker containers, and is well-documented and easy-to-deploy. Management of the node - which is primarily limited to harvesting new archival collections when posted to the network - uses a web based interface and only requires a few steps.

What responsibilities do node hosts have?

Node hosts are expected to set up and manage the LOCKSS hardware (WMU and LOCKSS staff will be available to assist with this), to install and perform basic management of a Linux system, to manage collection loading as needed, and attend to regular meetings to assist in the management of the network. Node hosts will also be required to periodically update the LOCKSS system as new versions are released. Other LOCKSS networks dedicate anywhere from .2-.8 FTE for the above activities.

Node hosts will also be required to assign a representative of their institution to the MDPN Technical Committee, which works cooperatively to solve community issues and maintain network operations during monthly meetings.

What are the network needs for access?

The LOCKSS server does not need to be open for public access. It will require a few connections allowed through a local firewall:

- Ability to host a LOCKSS node on a stable public IP address. This may be behind a NAT-based firewall.
- Connection to the props server at MCLS - standard Web connection (port TBD)
- Connection to the other nodes in the network using the LOCKSS protocol (port 9729)
- Allowed incoming connections to a limited IP range at WMU, for support needs (via ssh or equivalent)
- Allowed incoming connections to the LOCKSS IP range at Stanford, for support needs (via ssh or equivalent)

What are the bandwidth requirements?

Bandwidth needs are dependent on the collections. LOCKSS servers download the identified content locally, which may require high bandwidth, but this can be done during off hours. After the initial download, LOCKSS conducts checks against the content for fixity, but only takes a limited amount of bandwidth.

What hardware do nodes require, who does the purchasing?

Hardware will be provided by MCLS with funds from the Institute for Museum and Library Services. To assure that the hardware will integrate within your existing environment, node hosts will coordinate hardware selection with their existing technology providers. Purchase orders will then be sent to MCLS for payment. Storage capacity will be between 10-30TB, to be determined after consultations with LOCKSS.

Specific hardware requirements from LOCKSS can be found in the [LOCKSS 2.0 alpha5 System Manual](#).

What version of LOCKSS will nodes run?

MDPN uses LOCKSS 2.0, otherwise known as LAAWS (LOCKSS Architected as a Web Service), which is currently in alpha5 release.

According to our partners at the LOCKSS Program, alpha6 will be released in November 2022 and beta1 in January 2023. Nodes should expect that some initial upgrades may be required to get onto production versions as they become available. As noted previous, application upgrades are rather simple due to the Docker environment.

Can it be virtualized?

Yes - according to the LOCKSS Program, "There are no special concerns for running a LOCKSS system on a virtual machine; the LOCKSS Program and many LOCKSS Alliance members already do so." Western Michigan University is currently running nodes in this fashion using VMWare technologies. Other specifics can be investigated with LOCKSS, but generally seems to work well.

Who provides support?

Support will be shared between MDPN and our partners at Western Michigan University and the LOCKSS Program.

MDPN will manage membership, communication between nodes, and will coordinate training and provide documentation, as available. MDPN will also manage the network's props server.

Western Michigan University will provide support for connecting to the network.

LOCKSS will provide support for issues related to the LOCKSS software.

Support requests will be submitted via [Request Tracker](#) (RT) and routed to the correct contact as outlined above; node hosts will be provided with a login upon hardware setup.

What benefits can my organization receive by serving the community as node hosts?

Node hosts are eligible for a number of tangible and intangible benefits, including subsidized storage allocations in the network, discounted annual membership fees, dedicated representation on the MDPN's Technical Committee, free staff training and support resources from the network and our partners, and the knowledge that you are contributing to a project that will benefit the entire cultural memory community across the state and keep Michigan's cultural memory materials safe, accessible, and usable for the long-term future.

What is the timeline?

- Hardware selection / purchasing: Fall 2022
- Hardware setup, testing: Winter 2023
- Software installation: Winter 2023