

Alabama Digital Preservation Network

New Member Planning Packet

This document is designed to assist you in planning your membership in the Alabama Digital Preservation Network (ADPNet). ADPNet is a distributed digital preservation network for locally created digital content. It represents a low-cost digital preservation solution for academic institutions, state agencies, and cultural heritage organizations in Alabama. Any Alabama institution with digital collections that are freely available to the public is invited to join ADPNet and use the network as a dark archive for digital materials. ADPNet is not a web service, searchable repository, or digital library for digital content.

ADPNet has four categories of membership, each with specific fees and responsibilities, based on the aggregate size of content to be preserved.

- Anchor: the highest category of membership for organizations with large amounts of digital objects to be preserved and the necessary infrastructure to host at least one server node.
- Host: the next category of membership for organizations with moderate amounts of digital objects to be preserved and the necessary infrastructure to host one server node.
- Participant (large): a category for organizations with any amount of digital objects to be preserved, but without the necessary infrastructure or desire to host a node.
- Participant (small): a category for organizations with a small amount of digital objects to be preserved and no ability to host a node.

The fees and space allotments assigned to each of these categories are designed to ease the transition from one category to the next as an organization's digital preservation needs grow over time.

Rationale: Why do this, why not just backups? Backups are fine for what they do, and highly recommended for insuring the availability of production services. Backups, however, are one-off sequential copies that, once verified, are not monitored for reliability. Think of backups as a short-term insurance policy for things that get deleted, to retrieve recent earlier versions of an item, or for recovery from server failure.

Digital preservation is a different approach; it is not a production service like a web server or a CONTENTdm implementation. It entails creating digital archival versions of things, such as non-proprietary loss-less file formats, producing preservation metadata that includes information helpful for managing archival files and reconstructing relationships well into the future, and ensuring files do not change over time due to “bit rot,” neglect, or nefarious purposes.

Like backups, digital preservation is furthered by having multiple copies of the “bits” distributed across time and space so that if one copy is destroyed, others remain. Having additional copies alone, even if they are in the cloud (e.g., Amazon’s EC2), however, does not constitute digital preservation. The key protective element is ensuring that the files are constantly being checked for availability and change. To preserve something effectively requires a mechanism that

monitors the health and well-being of the materials and reports out any anomalies. The software underlying ADPNet, LOCKSS (Lots of Copies Keep Stuff Safe), provides the system of checking across multiple copies of the data (minimum six copies).

There are other digital preservation options, but none with the proven record of LOCKSS and the cost-effectiveness of ADPNet.

More details about ADPNet can be found on the web site: <http://adpn.org/>.

A digital preservation mechanism can address a variety of potential failures ranging from relatively mundane and limited circumstances to severe disasters that wipe out an entire organization. On the lighter end of this range, if a server fails or specific disks crash that contain the archival formats, these materials can be recovered from the preservation network. While the same can be said for backups, if an error occurred in the files and went unnoticed for longer than the retention schedule for backups, the organization would have lost a copy of the original. Access to a preservation network ensures the availability of these files over the long term.

In the case of a serious disaster in which large portions of an institutional infrastructure are destroyed, access to materials stored in an off-site preservation network will reduce the overall expense of rebuilding the no-longer-extant digital library or service.

Fees, allocations, responsibilities

Anchor:

Base annual membership fee: \$5,000

Base local data allotment: 1.5 TB

Allotment increments to base allocation: 500 GB (min.) @ \$350/each (annually)

Governance representation: 1 seat/vote on the Steering Committee, 2 seats/votes on the Technical Committee.

Responsibilities:

- Purchase/allocate and host at least one server with sufficient capacity to accommodate the anticipated network holdings for the coming year (may require an initial investment in a server for new members at this level).
- Agree to expand server capacity as needed. (Capacity is reviewed annually in the summer to determine expansion needs; the goal is to cover future server and storage expenses through the ADPNet membership fees.)
- Provide system administration for the server(s) according to network policies (0.25 FTE).
- Guide the development of the network (time commitment ~2 hours/month).
- Make material available for ingest.
- Serve as an onramp for Participant category members (harvest content, procure content, and/or provide assistance with technical issues).
- Follow the policies and procedures of ADPNet.

Host:

Base annual membership fee: \$2,500

Base local data allotment: 500 GB

Allotment increments to base allocation: 100 GB (min.) @ \$250/each (annually)

Cannot exceed 1 TB, must move to Anchor category.

Governance representation: 1 seat/vote on the Steering Committee, 1 seat/vote on the Technical Committee.

Responsibilities:

- Purchase/allocate and host one server with sufficient capacity to accommodate the anticipated network holdings for the coming year (may require an initial investment in a server for new members at this level).
- Agree to expand server capacity as needed. (Capacity is reviewed annually in the summer to determine expansion needs; the goal is to cover future server and storage expenses through the ADPNet membership fees.)
- Provide system administration for the server according to network policies (0.25 FTE).
- Guide the development of the network (time commitment ~2 hours/month).
- Make material available for ingest.
- Serve as an onramp for Participant category members (harvest content, procure content, and/or provide assistance with technical issues).
- Follow the policies and procedures of ADPNet.

Participant (Large):

Base annual membership fee: \$800

Base local data allotment: 1.5 GB

Allotment increments to base allocation: 500 MB (min.) @ \$25/each (annually)

Governance representation: 1 elected representative/vote per five category members (or portion thereof) to the Steering Committee.

Responsibilities:

- Make material available for ingest by staging it on an accessible web server.
- Participate in writing plug-ins for content ingestion.
- Follow the policies and procedures of ADPNet.

Participant (Small):

Base annual membership fee: \$300

Base local data allotment: 500 MB

Allotment increments to base allocation: 100 MB (min.) @ \$50/each (annually)

Cannot exceed 1.5 GB local data, must move to Participant (Large) category.

Governance representation: 1 elected representative/vote per five category members (or portion thereof) to the Steering Committee.

Responsibilities:

- Make material available for ingest.
- Follow the policies and procedures of ADPNet.

Example 1:

Organization A has 645 MB of digital content to preserve and is not able to host a server. Given the size of the content and the desire not to host, the Participant (Small) category would be appropriate. The annual ADPNet fee for Organization A would be \$400 (\$300 for the base allotment, and \$100 for two increments of 100 MB each). If Organization A exceeds 700 MB in the network, they would need to purchase another increment of 100 MB (or more, depending on the total size of their content).

Example 2:

Organization B has 33.7 GB of digital content to preserve and isn't sure whether or not they wish to host a server. Thus, a comparison of Participant (Large) and Host options is in order. If Organization B selects the Participant (Large) option the annual ADPNet fee would be \$2,425 (\$800 for the base allotment, and \$1,625 for 65 increments of 500 MB each). The annual ADPNet fee for the Host option would be \$2,500, and the base allotment would provide much room for growth. The decision would be made based on whether or not the organization valued hosting or not having to host a server more.

Example 3:

Organization C has 2.3 TB of digital content to preserve. The most appropriate option would likely be Anchor. At this category the annual ADPNet fee would be \$5,700 (\$5,000 for the base allotment, and \$700 for two increments of 500 GB each).

An organization may also wish to consider the level of engagement with the ADPNet organization and operations when weighing the decision of which category.

What you need to do:

1. Gather material: systematically determine what digital items are to be preserved.
2. Size the material: calculate the size in kilobytes, megabytes, or gigabytes.
3. Stage stuff: make files available to ADPNet servers via a web server.
4. Write plug-ins: create the files that describe what is ready for ingest.
5. Create html manifest pages that link your content to be harvested and include the LOCKSS permission statement.

ADPNet is not a mechanism for digitizing collections or making them available for public access. Membership in ADPNet assumes that the organization has already completed these tasks.

Understanding the process of harvest/ingest....combination of work in the member organization, LOCKSS staff, and existing ADPNet membership.

What to expect in the first year:

1. The planning process, as outlined above, will consume some portion of the first year depending on the size and complexity of the digital collections.
2. Harvesting will begin at a slow pace to ensure that all of the pieces of the process are working as expected.
3. A small to moderate amount of files from the organization will be ingested into the network during this period. The exact amount will depend on what materials are readied by the organization and the level of harvesting activity already underway in the network.
4. It is highly likely that an organization will begin the first year at a lower level category of membership than perhaps anticipated. In this way the organization is not paying for an amount of preservation activity that cannot be delivered.

Planning your preservation space needs

In order to arrive at a reasonable estimate of your storage needs, add together the file sizes of all your digital content and associated metadata.

1024 bytes = 1 kilobyte

1024 kilobytes = 1 megabyte

1024 megabytes = 1 gigabyte

1024 gigabytes = 1 terabyte

Planning for file types

ADPNet can accommodate most file types. Knowing this information beforehand supports better planning. Currently ADPNet has plug-ins for the following file types:

- Bitmap image files: bmp
- Cascading style sheets: css
- Graphics Interchange format (a type of bitmap image): gif
- Plain text: txt
- Extensible markup language: xml
- Waveform audio file format: wav
- Free Lossless Audio Codec: flac
- Joint Photographic Experts Group (JPEG images): jpg or jpeg
- JPEG 2000 image compression standard and coding system: jp2
- Icon image format containing 2 bitmaps: ico
- Portable Network Graphics bitmapped images : png
- Tagged Image File Format: tiff
- JavaScript source code: js
- Portable Document Format: pdf
- Gnu-zipped tar file archive: gzip or gz
- Compressed file archive: zip

Plug-ins are descriptions for various file types that will be harvested into the network. The files types listed above already have plug-ins; other file types will require the creation of additional plug-ins. Anchor and Host members do and Participant members may engage in writing plug-ins.