

Service Model for Wisconsin Statewide Digital Archival Storage

Wisconsin's public library systems, under the auspices of the Wisconsin Public Library Consortium (WPLC) and in partnership with Recollection Wisconsin, are working together to offer safe storage of archival copies of digitized materials and metadata as a sustainable long-term service, effectively in perpetuity.

The Digital Archival Storage System may ultimately be used for a variety of purposes. The initial use case, and the service model defined in this document, is **digital archival storage** - stable, centralized storage for digitized and born-digital content created or collected by Wisconsin libraries and cultural heritage institutions. This is a **dark archive** for the purposes of disaster recovery. Stored data is available only to depositors in order to replace the submitter's local files when necessary. The system does not support any public access, searching capability, or server backup functionality.

Libraries and cultural heritage institutions should consider the statewide Digital Archival Storage System as one component of their local digital preservation strategy. Following the best practices defined by the National Digital Stewardship Alliance and others, institutions should store at least one additional copy of all files in a location other than the statewide system.¹ Making more than one copy of digital materials and utilizing more than one type of storage solution mitigates a variety of digital preservation risks.

DEFINITIONS

Service administrators: The Digital Archives Backup Committee of the Wisconsin Public Library Consortium (WPLC) is the service administrator for the Wisconsin Statewide Digital Archival Storage program. The service administrator is responsible for approving depositors' use of the program.

Depositor: A **depositor** is any party authorized to deposit data into the digital archival storage system. At the launch of the service, eligible depositors include public library systems, collaborations of multiple public library systems, resource libraries, and the Recollection Wisconsin consortium.

Submitter: A **submitter** is any library or cultural heritage institution that partners with a depositor to provide data for deposit into the digital archival storage system. Each depositor is responsible for determining eligible submitters (for instance, a public library system's member libraries, or Recollection Wisconsin Content Partners).

¹ <https://ndsa.org/publications/levels-of-digital-preservation/>

SERVICE MODEL ELEMENTS

- I. Content selection
- II. Content preparation
- III. Content ingest
- IV. Content monitoring and tracking
- V. Content retrieval and removal

I. CONTENT SELECTION

All data deposited in the system must meet the following requirements:

- 1) Content aligns with the [Recollection Wisconsin Collection Policy](#). Specifically, the following types of content are out of scope:
 - a) Data-only records, such as birth, death or other genealogy indexes
 - b) Finding aids or EADs
 - c) Institutional repository content, such as student dissertations, theses, or research data
- 2) Content that is being preserved as part of a records management program (electronic records, meeting minutes, etc.) is out of scope unless it meets the parameters of the RW Collection Policy. Deposited items are not subject to retention schedules, and are assumed to be stored indefinitely.
- 3) [Restricted/sensitive](#) content (HIPAA, personally identifiable information, etc.) is prohibited.
- 4) Encrypted files are prohibited.
- 5) Files are accompanied by item-level metadata in an xls, txt, csv, or xml file.

Note: All file types are accepted as long as they meet the above criteria, but it is strongly recommended to provide the highest-quality files available in uncompressed, non-proprietary standard formats (e.g. TIF, WAV).

Roles and responsibilities:

- 1) Each depositor is responsible for ensuring that content from their submitters meets the above criteria.
- 2) Each depositor is responsible for working with their submitters to determine ownership, copyright, and intellectual property issues and for confirming their right to preserve content prior to submitting it for storage.
- 3) In the event of a dispute, the Digital Archives Backup Collaboration Steering Committee of the Wisconsin Public Library Consortium will make decisions about content eligibility. The WPLC Committee reserves the right to reject content for

any reason including the constraints of the storage space, as monitored and advised by SCLS and LEAN WI Technology staff.

- 4) The WPLC Committee reserves the right to require removal of content. The depositor will remove the content in a timely manner following a removal request.

II. CONTENT PREPARATION

In order to facilitate ongoing data authentication and tracking as well as future data recovery, all content must be packaged according to the [BagIt File Packaging Format](#) prior to deposit. According to the Library of Congress, “bags are based on the concept of “bag it and tag it,” where a digital collection is packed into a directory (the bag) along with a machine-readable manifest file (the tag) that lists the contents.”²

A bag is the “archival unit” that can be retrieved from the storage system in the future. Therefore, bags and their contents should be organized in a useful and logical way based on the needs of the depositor and the submitter. For example, a bag might be created for each of a submitter’s individual digital collections or projects.

Content should be virus-checked prior to running a tool to create the bag and its manifest. Depositors may use any available tool to create bags, as long as the bags conform to the parameters outlined below. DART (Digital Archivist’s Resource Tool), developed and maintained by APTrust, is recommended.³

To launch the service, Recollection Wisconsin project managers will provide training and assistance to depositors in bagging content using DART. After the initial service launch, each depositor is responsible for either a) bagging content for their submitters or b) providing guidance and support to their submitters to bag their own content.

Each submitted bag should:

- 1) Be uncompressed and unzipped.
- 2) Include a BagIt Profile that documents the depositor name and the submitter name.
- 3) Use the naming convention **depositor_submitter_foldername** (e.g. SCLS_PDS_LocalHistoryPhotos).

² <https://www.digitalpreservation.gov/series/challenge/data-transfer-tools.html>

³ <https://aptrust.github.io/dart-docs/>

After it has been deposited into the system, a bag is considered “closed.” Additional content should be placed in a new bag, not added to an existing bag. If a bag needs to be modified, it is recommended to create and submit a new, replacement bag and remove the old bag.

III. CONTENT INGEST

The currently-used ECS system stores data as objects within S3 buckets. The digital archival storage environment within the ECS will be configured as follows:

- a) One bucket for each depositor (bucket owner = primary contact at depositor institution).
 - i) One folder within the bucket for each submitter (individual library or cultural heritage institution).
 - 1) One sub-folder for each bag, where each bag represents a project, collection, or other logical grouping defined by the submitter.

Notes: Retention period for buckets will be set to infinite. An [“object lock”](#) function will be enabled at the bucket level using the S3 API. This activates WORM (write once read many) capabilities, which prevents accidental overwriting or deletion. It is recommended to use “legal hold” and “compliance mode” settings in Object Lock.

Roles and responsibilities:

- 1) Depositors are responsible for transferring complete, validated bags into the appropriate bucket and folder. Data transfers will be made by connecting to the ECS system using a VPN connection and an S3 browser.
- 2) SCLS Technology staff will create buckets, manage bucket settings, and manage user roles and permissions.
- 3) SCLS Technology staff will provide access credentials and documentation for connecting to the ECS.
- 4) SCLS and LEAN WI Technology staff will provide assistance to depositors if needed for troubleshooting issues with connecting to the ECS or with data transfers.

IV. CONTENT MONITORING AND TRACKING

The functions of the currently used storage system include automatic validation and healing. "For data integrity, ECS utilizes checksums [md5]. Checksums are created during write operations and are stored with the data. On reads checksums are calculated and

compared with the stored version. A background task scans and verifies checksum information proactively."⁴

Beyond the automatic monitoring described above, each depositor is responsible for monitoring content while it is at rest in the storage system. For instance, a depositor might choose to manually check their data on a recurring basis by downloading a selection of bags and running a validation.

To support ongoing management of the deposited data, depositors should log their submissions in a central location. A depositor might maintain a spreadsheet listing each bag, its submitter, and when it was deposited, or copies of all bag manifest txt files could be saved in a central location.

V. CONTENT RETRIEVAL AND REMOVAL

Retrieving bags from the system may be necessary in the event of a disaster or other local data loss. Each depositor is responsible for managing retrieval requests from their submitters. Bags can be downloaded by depositors using the S3 browser over the VPN connection, and then delivered to the submitter using whatever local method is preferred. It is strongly recommended that the depositor runs a virus check on downloaded bags before delivering them to submitters.

DELL ECS SYSTEM

The current infrastructure available for digital archival storage consists of two identical Dell ECS platforms. As described by Dell, “ECS (Elastic Cloud Storage) is a modern software-defined object storage platform designed for both traditional and next-generation workloads that provides organizations with an on-premise alternative to public cloud solutions.”⁵

The primary ECS instance is housed at South Central Library System (SCLS) in Madison, Wisconsin and is managed and maintained by SCLS Technology staff. The replication instance is located at the Chippewa Valley Technical College’s Regional DataCenter in Eau Claire, Wisconsin and is managed and maintained by LEAN WI Technology staff.

4

<https://www.delltechnologies.com/asset/en-us/products/storage/industry-market/h14071-ecs-architectural-guide-wp.pdf>

⁵ <https://www.dell.com/en-us/dt/learn/data-storage/ecs.htm>