STATEWIDE DIGITAL ARCHIVAL STORAGE

PROJECT HANDBOOK

WPLC
Wisconsin Public Library Consortium

November 2024

TABLE OF CONTENTS

HOW TO USE THIS HANDBOOK

This handbook contains all of the background information, step-by-step instructions, templates, forms, and definitions you will need for your library system to participate in the Statewide Digital Archival Storage program.

Background information comprises the first several pages. Once you've reviewed the background information, begin following the steps in the Project Preparation section (page 8). Below is an overview of the training process to participate in the Statewide Digital Archival Storage program.

Review this handbook and return signed forms. Set up first training session with project manager.

Attend first training session. Determine if you will be uploading digital files or if you'd like the project manager to upload.

If uploading yourself, install software and organize files as directed. If not uploading yourself, transfer copies of digital files to project manager as directed.

If applicable, attend second training session and upload files.

If you're viewing this manual in PDF, clicking the home icon on any page will take you back to the Table of Contents for easy navigation.

PROJECT INFORMATION

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 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.

The Statewide Digital Archival Storage Project is governed by the WPLC Technology Collaboration Steering Committee and managed by the WPLC Digital Archives Backup Collaboration Workgroup.

PROJECT HISTORY

In 2019, Wisconsin public library systems entered into a collaboration to support two major initiatives: the Backup Collaboration Project and the Digitization Archives Storage Platform Project. These two initiatives are independent but interrelated in that they both rely on infrastructure housed in two data centers (South Central Library System and LEAN Wisconsin). Startup funding for both projects has come primarily from LSTA funds distributed by the Department of Public Instruction. Also in 2019, Recollection Wisconsin established a Storage Working Group to investigate models for providing a cost-effective shared managed storage environment for contributing institutions. Working Group members were Ann Hanlon (UW-Milwaukee), Paul Hedges (WHS), Judy Pinger (Milwaukee Public Library), Scott Prater (UW-Madison), and Vicki Teal Lovely (SCLS).

In 2022, the governance for both projects was moved under the auspices of the Wisconsin Public Library Consortium. Also in 2022, WiLS (Wisconsin Library Services) was brought on as the project manager to onboard participating library systems to the digital archival storage servers.

Planning and documentation for the onboarding process began in 2023; onboarding began and continues in 2024.

DEFINITIONS

Roles and responsibilities

Service Administrator
The Digital Archives Backup Collaboration Workgroup 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.

Storage Manager
Each depositor organization will designate a Storage Manager, who is responsible for monitoring and maintaining data within the shared digital archival storage system, ensuring data integrity and accessibility. This role is typically filled by an IT, database, or digitization specialist.

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).

Bag: a digital package of files and associated metadata; the digital equivalent of placing physical archival documents in an acid-free box and making a record of the box's contents.

Software

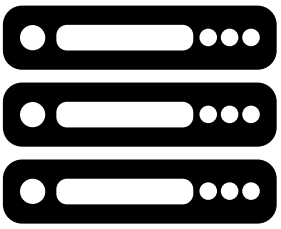
DART: Digital Archivist's Resource Tool
Depositors will use this tool to "bag" digital content and metadata.

ECS (Elastic Cloud Storage) / S3 Browser
ECS is the Dell object storage platform where digital content will be stored. The S3 browser is the navigational tool to manage and operate each depositor's access to ECS.

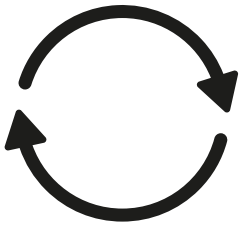
VPN
Virtual private network, used to securely establish a connection between your computer and a remote server. You'll use a VPN to access the Elastic Cloud Storage (ECS) and S3 Browser.

SERVICE OVERVIEW

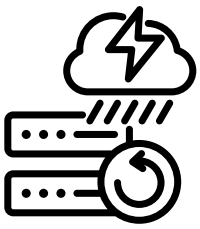
Depositor (storage manager) uploads files



WPLC (with SCLS and LEAN WI) maintain servers



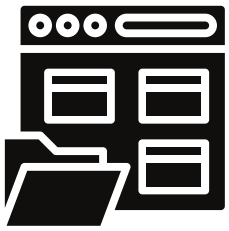
Service administrator and depositor both maintain files and add new files



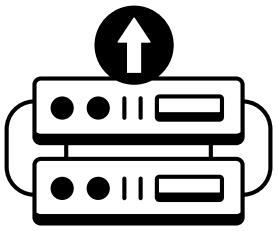
Files available to replace in case of local loss

UPLOAD PROCESS OVERVIEW

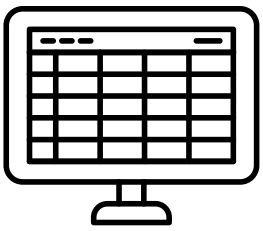
Depositor's storage manager prepares and organizes files



Depositor bags files and metadata using DART



Bag is uploaded to storage server



Recordkeeping: files, metadata, organization

DIGITAL PRESERVATION

Although the Digital Archives Backup Collaboration was not initiated as a digital preservation solution, the language and goals of **digital preservation** are key to informing how this project is approached. The American Library Association defines digital preservation as the ongoing activities that "combine policies, strategies and actions that ensure access to digital content over time."

Digital Preservation ensures that digital assets are **protected, interpretable and authenticated over long periods of time**. Storing[1] data unchanged on secured disks ensures authenticity and guards against unplanned, unrecorded changes and corruption; however, it does nothing to make sure that data will be readable in the future on next-generation applications, computers, or meaningful to future users. Nor will a disk array protect any rights and licenses over the preserved materials.

Digital Preservation **requires an institutional commitment** to keep interpretable copies of digital resources in perpetuity, in the face of disaster, institutional change, and long-term changes in the world at large over time. It requires significant ongoing investment in technology and continuous curation to make sure that the resources being preserved are not lost to future generations. It also requires technical staff intervention to retrieve the preserved files, then ready them for further distribution.

Digital preservation is an **ongoing activity**, rather than an end state: something is never preserved, but is continually being preserved. Although the goal of digital preservation is to ensure perpetual availability and usability of digital resources, the rolling horizon of the those engaged in digital preservation is to maintain accessibility over the next three-to-five years, with the assumption that sustainable institutional commitment to the program writ large will continue indefinitely.

1. "Definitions of Digital Preservation", ALCTS Preservation and Reformatting Section, Working Group on Defining Digital Preservation, https://www.ala.org/alcts/resources/preserv/defdigpres0408

DIGITAL PRESERVATION

Backups are not Digital Preservation

Backups are a necessary component of a digital preservation system, but they are not the same as digital preservation. Backups are designed to ensure continuity of operations: a backup returns you to your current state in the event of data loss. Backups are rarely held for longer than a year, and backups do nothing to guarantee the long-term interpretability of digital resources. Backup technologies age rapidly, and as older backup media and hardware are deprecated, the content stored on them are increasingly vulnerable to decay and loss.

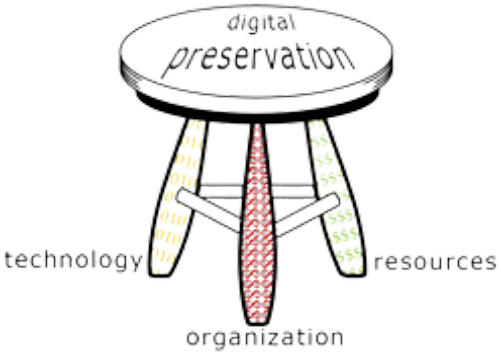
A complete digital preservation plan will include backups of stored data as a matter of course, but will go beyond the technical infrastructure to ensure ongoing institutional commitment to guarantee usability of stored digital assets for the medium and long-term.

A **Digital Preservation Framework** is a document that is both a statement of purpose and a contract that commits an entity or entities to a digital preservation program. Once the need for a digital preservation program has been agreed upon, the first step towards building that program is to articulate the entity's concrete vision, goals, scope, and commitments to digital preservation. Most institutions with a digital preservation program publicly publish their Digital Preservation Framework document.

There are many models for the digital preservation framework. One of the most widely-used is the template developed by Nancy McGovern and Anne Kenney in 2007. The structure of this template in turn is based on McGovern and Kenney's Three-legged Stool model to describe the areas of concern of a digital preservation program.[2] *(see next page)*

2. McGovern, Nancy. "Version 2.0 Digital Preservation Policy Framework: Outline." ICPSR, October 2007. http://www.icpsr.umich.edu/files/ICPSR/curation/preservation/policies/dp-policy-outline.pdf

DIGITAL PRESERVATION



The three-legged stool of digital preservation [3]

Organizational Infrastructure includes the policies, procedures, practices, people–the elements that any programmatic area needs to thrive, but specialized to address digital preservation requirements. It addresses this key development question:

What are the requirements and parameters for the organization's digital preservation program?

Technological Infrastructure consists of the requisite equipment, software, hardware, a secure environment, and skills to establish and maintain the digital preservation program. It anticipates and responds wisely to changing technology. It addresses this key development question:

How will the organization meet defined digital preservation requirements?

Resources Framework addresses the requisite startup, ongoing, and contingency funding to enable and sustain the digital preservation program. It addresses this key development question:

What resources will it take to develop and maintain the organization's digital preservation program?

3. Anne R. Kenney and Nancy Y. McGovern, "The Five Organizational Stages of Digital Preservation," in Digital Libraries: A Vision for the Twenty-first Century, a festschrift to honor Wendy Lougee, 2003. Available from the University of Michigan Scholarly Monograph Series http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11.

PROJECT PREPARATION

Before you get started, there are a few preparatory steps to take to set you up for success.

Designate a single person to act as your storage manager.
Each depositor (e.g. public library system or collaboration of multiple systems) will be provided with access credentials for a single user who is authorized to upload, access, and remove files from the depositor's designated bucket in the storage system. Contact the WPLC project manager at wplc-info@wils.org to request, update, or transfer those credentials. Note this information on your participation agreement.

Identify a processing workstation for your storage manager to use.
This workstation should have a fast, reliable Internet connection and at least 200GB of space available for temporary loading and processing of files. The workstation will occasionally be bogged down by long uploading processes, so it should be a computer that can be left alone during those times. It's also likely to be faster if connected to the internet directly (rather than via wi-fi).
If you don't have 200 GB available, you can use an external hard drive if absolutely necessary, but this is not recommended as it increases the chances for file corruption.

On that workstation, install three pieces of software. You'll receive credentials from the SCLS technology team for the VPN and the S3 browser.

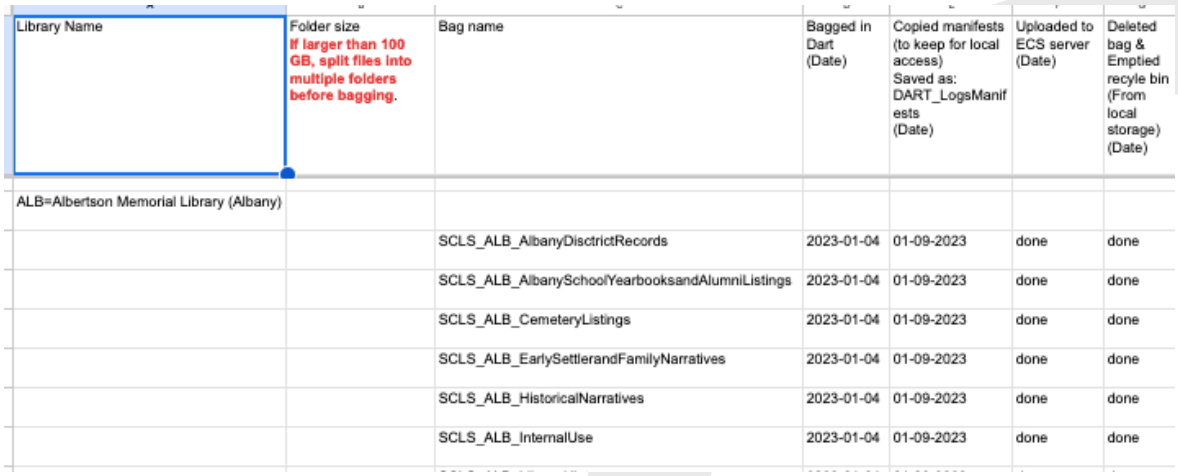
- Digital Archives Repository Tool (DART) (see p. 18)
- VPN (see Appendix B)
- S3 browser (see Appendix B)

PROJECT DOCUMENTATION

Set up your storage log and folders.
A key piece of a well-managed digital storage program is maintaining up-to-date documentation of the files you have stored, who they belong to, and where they are located. Each depositor using the storage system is responsible for creating and updating their own documentation.

We recommend a two-part approach to documentation:

- **Digital Archival Storage Log.** Use a log to record information about each bag (package of files and related metadata) you move into storage. The Digital Archival Storage Log template can be downloaded and modified to meet your needs. This spreadsheet contains several different tabs for tracking files at different stages.
- **Folder for storage manifests.** Each bag you create will come with several text files, known as manifests, that contain important information about what's in the bag. Designate a location outside of the Dell ECS storage system where you'll save copies of these manifests for future reference.



Library Name	Folder size If larger than 100 GB, split files into multiple folders before bagging.	Bag name	Bagged in Dart (Date)	Copied manifests (to keep for local access) Saved as: DART_LogsManifests (Date)	Uploaded to ECS server (Date)	Deleted bag & Emptied recyle bin (From local storage) (Date)
ALB=Albertson Memorial Library (Albany)						
		SCLS_ALB_AlbanyDisctrictRecords	2023-01-04	01-09-2023	done	done
		SCLS_ALB_AlbanySchoolYearbooksandAlumniListings	2023-01-04	01-09-2023	done	done
		SCLS_ALB_CemeteryListings	2023-01-04	01-09-2023	done	done
		SCLS_ALB_EarlySettlerandFamilyNarratives	2023-01-04	01-09-2023	done	done
		SCLS_ALB_HistoricalNarratives	2023-01-04	01-09-2023	done	done
		SCLS_ALB_InternalUse	2023-01-04	01-09-2023	done	done

CHOOSE FILES FOR STORAGE

General Requirements

All content placed in the storage system should align with the Recollection Wisconsin Collection Policy. Formats may include, but are not limited to, photographs, postcards, maps, letters, diaries, articles, books, artifacts, artwork, audio, film, and video. Content is not required to be made available online through Recollection Wisconsin at the time it is stored, but it should support Recollection Wisconsin's overall mission to provide access to digital collections from Wisconsin cultural heritage organizations.

Ineligible Content

The following types of content are out of scope:

- Content that is not included in the Recollection Wisconsin Collection Policy, specifically:
  - Data-only records, such as birth, death or other genealogy indexes
  - Finding aids or EADs
  - Institutional repository content, such as student dissertations, theses, or research data
- Content that is being preserved as part of a records management program (electronic records, meeting minutes, etc.). All content placed in the storage system is assumed to be stored indefinitely, and should not be subject to retention schedules.
- Restricted or sensitive content (HIPAA, personally identifiable information, etc.)
- Encrypted files

Depositors are responsible for communicating these requirements to their member libraries or other submitters. Questions or concerns about content eligibility will be reviewed by the Digital Archives Backup Collaboration Work Group of the Wisconsin Public Library Consortium. The WPLC Work Group reserves the right to reject content or require its removal from the storage system.

CHOOSE FILES FOR STORAGE: CONSIDERATIONS

What are the local needs and priorities?
If you're a public library system working with member libraries, what are those libraries' expectations and needs? For example, a grant-funded digitization project might require a specific approach to storage and preservation.

Are the files "done" and ready to store?
A digitization project should be as "done" as possible before moving it into the dark archive. The storage system is not intended to be an active backup - while a project is in progress, be sure to back up your files regularly in a separate location, so you don't lose your work in the event of a hardware failure, accidental deletion, etc.

What is the risk if files are lost? Consider where your files fall on the Data Criticality Scale:

Data Criticality		Examples
1	Only digital and we hold the only copy - if we lose it, it's gone forever.	Oral history interviews recorded digitally "Digital donations" (i.e. materials loaned for scanning by a patron, patron keeps the originals)
2	We have a digital copy, but physical versions are at high risk.	Digitized cassettes, VHS tapes, or other analog audiovisual media
3	We have a digital copy, but physical versions reside elsewhere.	Collaborative projects with partners, i.e. a local historical society
4	We have a digital copy and digital copies reside elsewhere.	Institutional partnerships, i.e. materials digitized for public libraries by University of Wisconsin Digital Collections, or historical newspapers scanned from microfilm by the Wisconsin Historical Society
5	We have a digital copy and still hold the original physical item.	Materials that can be readily located and re-scanned, if necessary.

CHOOSE FILES FOR STORAGE: METADATA

A file, such as a spreadsheet, csv, txt, or xml file, containing descriptive metadata for each item should be included with **every bag**. This is so that when a bag is retrieved in the future, the relevant metadata can be retrieved along with the digital files. The only time a bag will not have a metadata file is if metadata was not available at the time the materials were bagged.

Tip: For CONTENTdm users, TXT and XML metadata files downloaded from CONTENTdm should be packaged in every bag. See instructions for exporting metadata from CONTENTdm here.

Whenever possible, content should be accompanied by item-level metadata in a csv file. You can download a template here to modify for your own use. Alternative but non-preferred formats include xls, txt, or xml file.

TYPES & FORMAT OF FILES TO STORE

Primary files
All file formats are accepted as long as they meet the criteria outlined above. The general recommendation is to store the highest-quality files you have available, in uncompressed, non-proprietary standard formats. Consider these your primary files (also known as archival files, preservation files, or master files).

In many cases, those primary files will be TIFF (for images and text), WAV (for audio), or MOV (for video). A good overview of recommended file formats and specifications is available from the UC-Santa Cruz Libraries' Digital Initiatives Department.

Primary files and access files
Typically, access files, also known as derivatives, are copies that have been reduced or compressed for sharing online. In some cases, you might choose to store both primary files and access files. For example, if your primary and access files are already stored together locally, it might be more efficient to keep them together instead of reorganizing. Or, if your access files are cropped, enhanced, or otherwise edited, you might want to retain those alongside your unedited primary files. For more on primary vs. access files, see Digitize: Bronze Level in the Recollection Wisconsin Digital Readiness Toolkit.

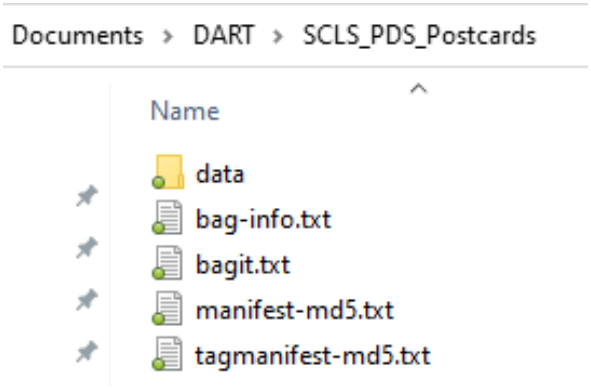
Access files only
It's possible that you might not have primary files available for storing -- maybe they were created long ago and can't be located, or maybe they never existed. In that case, store the best version you have available to you. It's better to store a less-than-ideal file than to lose it entirely.

Unreadable, obsolete, or proprietary files
You may find that some of your files cannot be opened because they were created in formats that are now obsolete, incompatible with your current system, or dependent on a particular piece of software to read. These unreadable, obsolete, or proprietary files can still be stored! Again, it's better to store a less-than-ideal file than to lose it entirely. In the future, they can be converted on a case-by-case basis as time, resources, and tools allow. Indicate on your digital archival storage log if a file format is obsolete.

It is recommended to note if files are primary files or access files on your metadata spreadsheet, as well as if the files are unreadable, obsolete or proprietary.

PREPARE FILES FOR STORAGE

All files must be packaged into **bags** before they are placed in storage. A bag is a folder structured in a specific way (the BagIt file packaging format). See the image to the right for an example of a bag. Every bag includes a sub-directory, named "data," that contains the files you will store, plus four text files that carry important information about the bag and its contents. You will bag up files and their metadata, then load that bag into the storage system.



Creating a bag is, more or less, the digital equivalent of placing physical archival documents in an acid-free box and making a record of the box's contents. Packaging files into bags before moving them into storage will make it much easier to locate, check, move, and retrieve those files in the future.

This section describes the steps to organize your files so they're ready for bagging. The process of creating the actual bags is described starting on page 18.

This preparation step can be one of the most time-consuming parts of this process, but it's also one of the most important. The better organized your files are before they go into storage, the easier it will be to locate a particular file needed in the future. The time invested in preparation now is a gift to your future self and colleagues. Keep in mind that if quality control of file naming, formats, and organization can be handled consistently during digitization, it will eliminate the need for a lot of preparation.

GENERAL RECOMMENDATIONS FOR BAGS

Each bag should be around 100 GB in size, more or less. This is for ease of uploading and file management, so a little over 100 GB is okay. If a submitter has files totaling more than 100 GB, split into multiple bags.

Group content in bags in a useful and logical way. Remember that the bag is the unit that you will put in storage and what you will get out of storage -- you can't pull individual stored files or folders out of a bag.
For example, a bag might be created for each of a library's individual digital collections or projects. If you are handling content for lots of different submitters, you might want to keep each submitter's files in a separate bag.

Bag a copy of your primary files. The Dell ECS system is a supplement to your local storage system, not a replacement for it. Don't remove your files from local storage in order to put them in shared storage - copy them, so they can live in both places. Remember: LOCKSS - Lots of Copies Keep Stuff Safe. Preparing copies of the files also reduces the potential for introducing errors or modifications to the original primary files.

Have enough working space for preparing files. If possible, the depositor should arrange for around 200 GB of free space on the administrator's local computer to serve as processing space. Each submitter's files can be removed from this space after they are bagged and moved into storage.
Tip: It is not recommended to use external hard drives for this purpose, as they have a short life and increase the risk of bit rot and other invisible damage to files.

FILE PREPARATION

Overview of File Preparation Steps:

1. Copy files onto your local workstation or network
2. Review file formats/types to include. Check for duplicates, invalid formats, etc.
3. Determine folder structure/organization
4. Generate item-level metadata file
5. Check for viruses

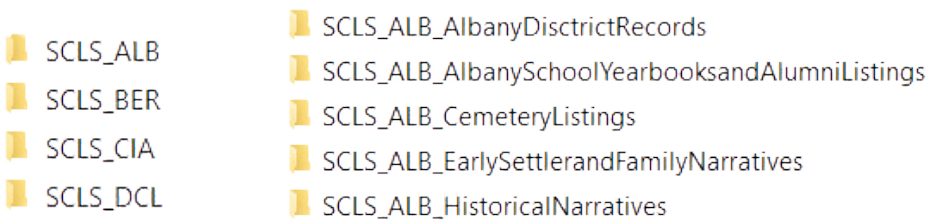
Folder Size and Structure

- Use standard directory organization with folders, sub-folders, sub-sub folders as appropriate.
- Bags should be about 100 GB, which means you should track the file size of each folder in your digital archival storage log so you know how many folders can be bagged together. If a single folder is larger than 100 GB, separate out into smaller folders.
- Tip: You can bag much smaller folders individually, for instance if you want a single collection to be in a bag, but it's more time-consuming to upload folders that way - and will be more time consuming in the future to retrieve them. This decision will depend on the submitter, collection size, and future use recommendations.

Folder Naming

- Each folder should be named so that both the depositor and the submitter are identified. You may want to use a standard code or acronym, for example: SCLS_PDS or RW_DCHS. Add this standardized identifier as the prefix to all folders and sub-folders.
- Do not use spaces in folder or file names. The only punctuation can be underscore (_) or hyphen (-).
- File names within folders do not need the identifier.

Advanced Renamer is a good tool for bulk renaming folders on Windows computers. Mac computers have this capability built in.

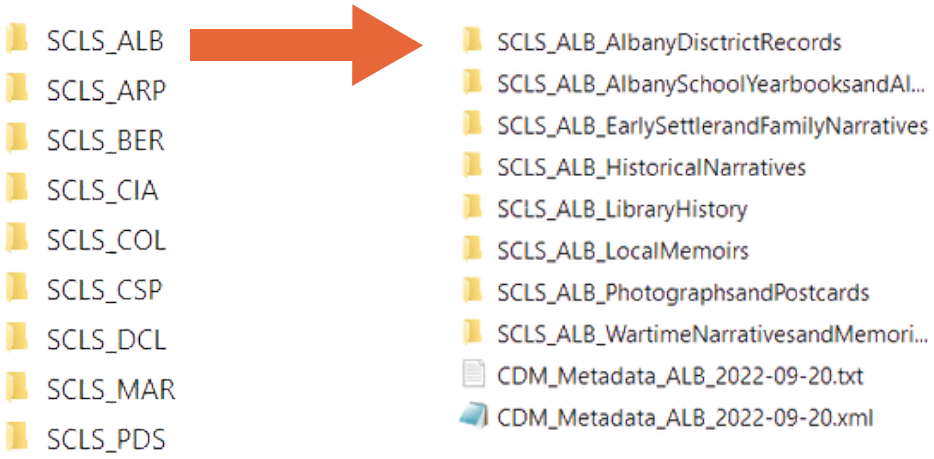


This is an example of how your folders and subfolders could be named and organized.

FILE PREPARATION

Prepared Directory Example

After completing the above steps, the directory should resemble the image below before proceeding to use DART (more on DART on page 18). The SCLS_ALB folder includes eight sub-folders organized by project (with the identifier SCLS_ALB in every folder name) and two files (txt and xml) containing descriptive metadata for all items in each project.



Virus Check

As a final step before creating bags, scan all folders for viruses. Use the anti-virus software your organization subscribes to, or use free anti-virus software such as ClamAV or AVG. If a virus is found, contact your IT department or follow your organization's protocols for quarantine or repair.

PACKAGE FILES FOR STORAGE

Recommended Packaging Tool: DART
All files must be packaged into bags before they are placed in storage. The Digital Archivist's Resource Tool (DART), maintained by APTrust, is a free, simple open-source tool for creating bags.

Steps for getting started with DART:

1. Install DART on the computer where you will be creating bags and uploading them to the storage system. The current version, available for Windows, Mac, and Linux, is available here: https://aptrust.github.io/dart-docs/download/
2. Also save or bookmark the DART documentation page for future reference.
3. Configure a BagIt profile within DART. This generates some key information that will be applied to all of the bags you create. See Appendix A: Configuring a BagIt Profile for more information.

After you've installed DART and configured an institutional BagIt profile, you'll use DART to generate a bag for each of the folders you established in the file preparation step. Bags created using DART are temporarily saved on the computer where you're running DART before they are moved into the storage system. That means that you must have enough space available on your computer to hold the bag. We recommend keeping each bag around 100GB in size, so be sure to have at least 200GB of working space on the computer where you're creating the bags.

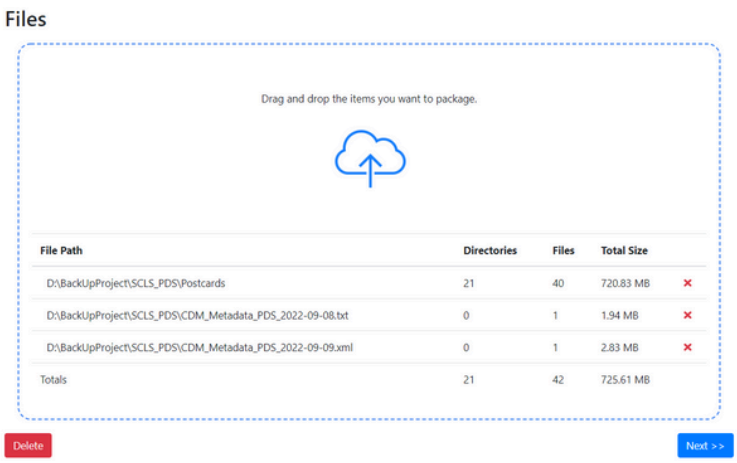
Tip: **Do not configure DART to place bags on an external hard drive,** as this will increase the potential for corruption or errors.

Tip: When you are using DART, it may be helpful to temporarily disable any local software that runs automatic backups or syncing of files (like Carbonite or Dropbox). The software may initiate creating a back-up and you will not be able to delete the bag until the backup is completed. Not disabling the software may slow down the workflow.

CREATE AND CONFIGURE A BAG

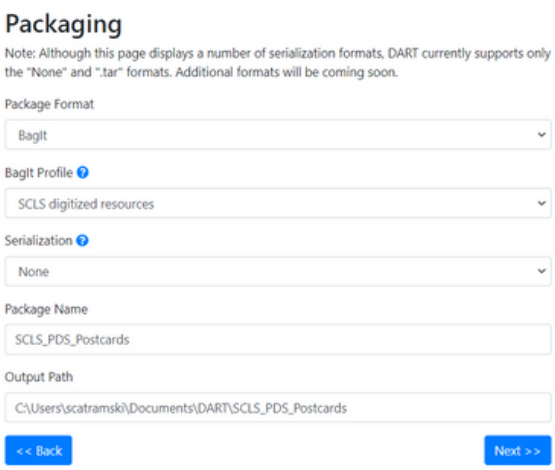
To create a bag in DART:

- Click **Jobs >> New**.
- Drag and drop the items you want to package. Include folder(s) plus associated metadata files.
- **Click Next**.

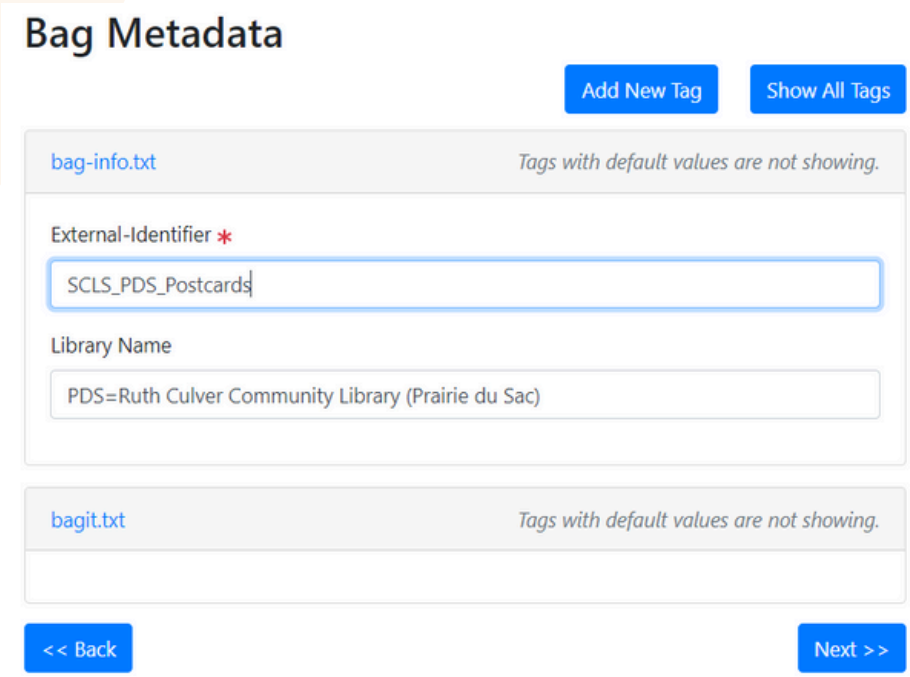


In **Jobs >> Packaging**, set up the following configuration:

- Package Format = BagIt
- BagIt Profile = Name of the BagIt profile you set up earlier (see Appendix A)
  - Example: *SCLS Digitized Resources*
- Serialization = Leave blank
- Package name = Your chosen file prefix plus file or folder name. This will become the name assigned to your bag.
  - Example: *SCLS_PDS_Postcards*
- Output Path = This will automatically populate with the drive location set in your profile, plus the package name. This is where your bag will be placed after it is created.
  - Example: *C:\Users\scatramski\Documents\DART\SCLS_PDS_Postcards*
- Click **Next**.



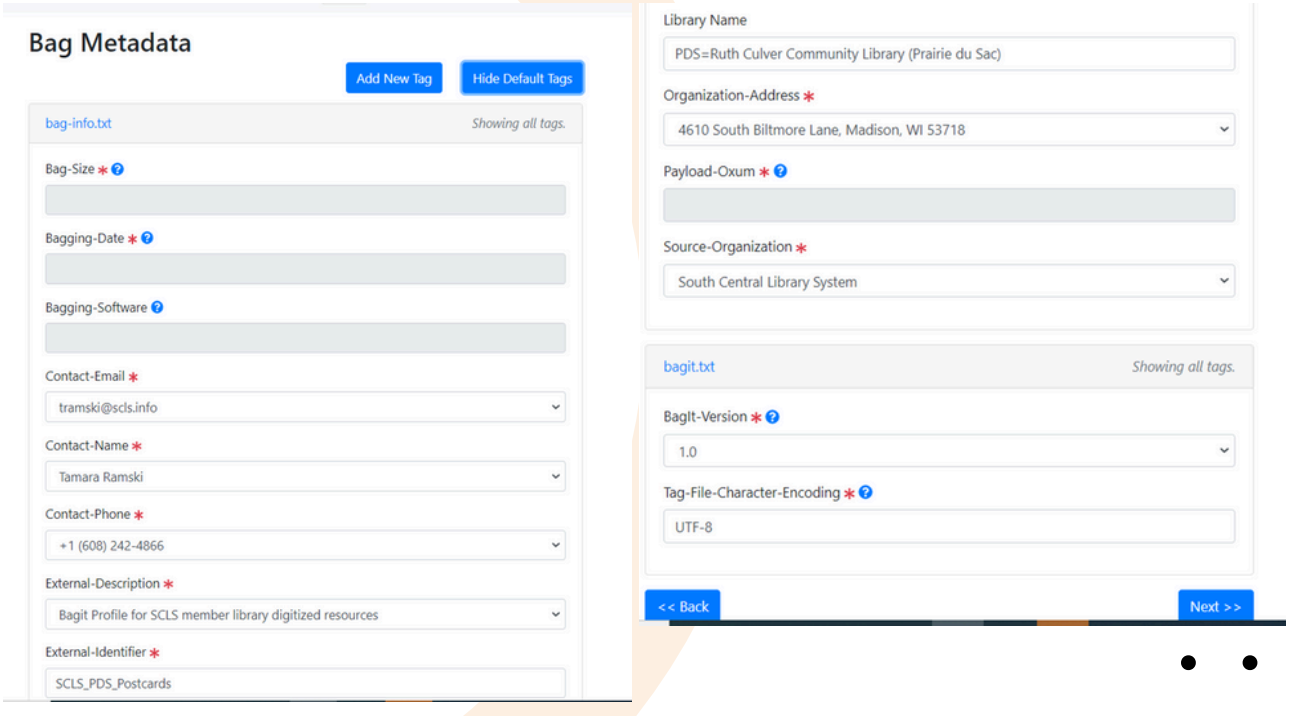
- In **Jobs >> Bag Metadata**, use the following configuration:
  - External-Identifier = Should be identical to the Package Name on the previous screen.
    - *Example: SCLS_PDS_Postcards*
  - Library Name and Address = Add the official name, address, or any other basic details that will help identify the individual library or other submitting institution(s) associated with the bag.
    - Example: *PDS=Ruth Culver Community Library (Prairie du Sac)*
      Tip: If you'll be creating multiple bags for each submitter, create a list of library names to copy/paste into this field in order to save time and improve consistency.
- Click **Next**.



Alternatively, if you did not create a BagIt profile, or you need to change any information in the profile, you can input more detailed metadata for the bag at this step.

To input more detailed metadata or to edit metadata, In **Jobs >> Bag Metadata**, use the following configuration:

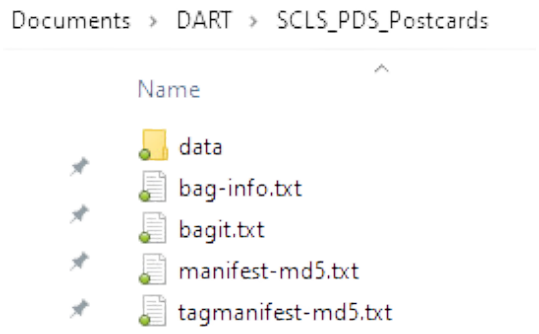
- In the upper right, click **Show All Tags.**
- **Bag Size, Bagging Date, and Bag Software** = Leave blank. These will automatically populate later in the process.
- **Contact Email, Contact Name, and Contact Phone** = Enter the submitter's name (library or content partner staff), email address, and phone number. This is in case someone needs to retrieve contact information for this particular bag in the future.
- **Bag-Count** = Leave blank
- **Bag-Group-Identifier** = Leave blank
- **External description** = Should automatically populate.
- **External-Identifier** = Use the package name here that you set up earlier.
  - Example: *SCLS_PDS_Postcards*
- **Library Name and Address** = Add the official name, address, or any other basic details that will help identify the individual library or other submitting institution(s) associated with the bag.
- **BagIt Version** = 1.0
- **Tag-File-Character-Encoding** = UTF-8
- Click **Next**.
- In **Upload Targets**, click **Next**.
- Click **Run Job**.



VERIFY THE BAG

- Locate the saved bag on your system and open it.
- Do a visual check to confirm that the following folders and files were created:
- data folder - this is where your files for storage are gathered
  - bag-info.txt
  - bagit.txt
  - manifest-md5.txt
  - tagmanifest-md5.txt
- Open the data folder and visually verify that it contains what you intended to bag. There is no need here to do a close examination of each file, just glance in the folder to confirm the number of files and folders, the file types, file names, etc. appear correct.
- Open the bag-info.txt document and verify the following:
  - Packet Size: In MB or GB, should match the total size of the files you bagged - check against the DART log in your project spreadsheet
  - BagIt-Version: 1.0
  - Tag-File-Character-Encoding: UTF-8

Example image: The data folder and four txt files together are the bag.



Troubleshooting DART

If there's a validation failure (bag does not validate in DART):

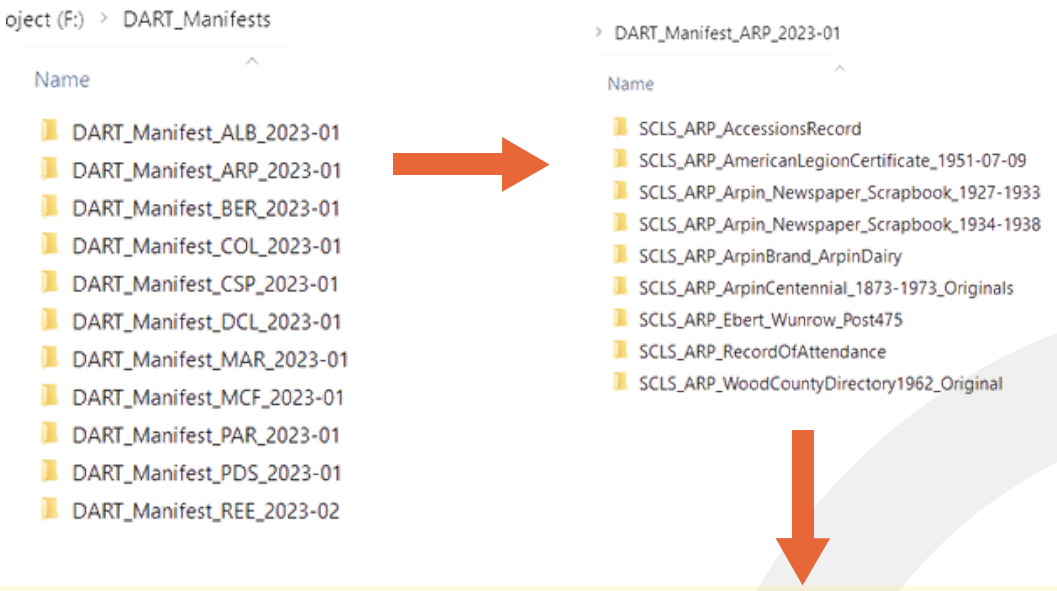
- Check the amount of memory available on your computer. If there is not enough memory, bagging may fail.
- Close DART and retry bagging.

If bagging or validation goes on for a longer than normal period of time:

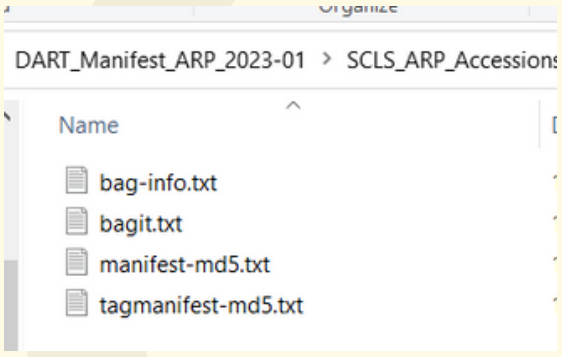
- Check the folder size. Large folder sizes (greater than 100 GB) may cause this to happen. Break the folder into several smaller folders.

LOG THE BAG AND SAVE THE MAINFESTS

- Record in the DART Log tab of your digital archival storage log that this bag has been created in DART, indicating that it is ready to be moved into storage.
- Open the manifest-md5.txt document and make a note of the file types (jpg, tif, etc.) contained in the bag. This data may help identify future migration needs.
- Create a unique folder on your local system dedicated to manifests. There can be one top-level folder (example: DART_Manifests) with a sub-folder for each submitter. Folders within each sub folder should be manually renamed to match the name of the bag.
- Copy the four text files (bag-info.txt, bagit.txt, manifest-md5.txt and tagmanifest-md5.txt) to the unique folder on your local system dedicated to manifests for this project. This will allow you to easily access the information contained in the manifests if you need it in the future, without having to access the entire bag in the storage system. **Do not rename or modify these files.**



Tip: The data folder and the four text files will always have the same names. Do not modify these names; they are important for future validation. Because they do not have unique names, it is important that the folder where you save them locally does have a unique name.



MOVE FILES INTO STORAGE

After you have created and validated your bags, you can begin uploading them into your institution's assigned bucket within the storage system. These data transfers will be made by connecting to the Dell ECS system over a VPN and uploading through an S3 browser (see Appendix B: Accessing ECS with VPN).

SCLS Technology staff will provide access credentials. You will need to install a VPN client such as FortiClient (free) and the S3 Browser (freeware version available, but the $40 pro version is highly recommended because it will significantly increase data transfer speeds).

Moving data into the Dell ECS system can be time consuming, so we recommend that you only start an upload when the computer can be left on indefinitely until the upload is completed. It is likely that all other activity on the computer will be slower during the upload process.

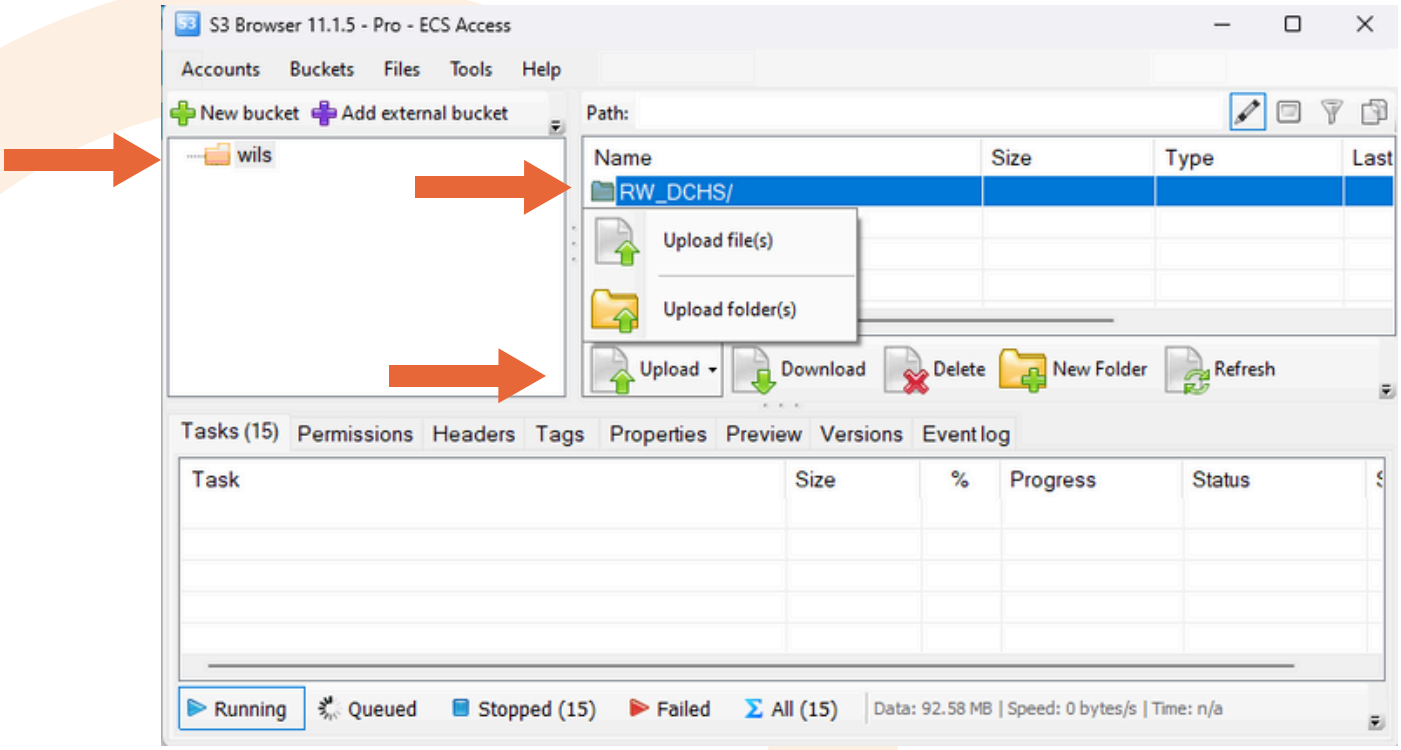
Tips:

- If possible, using a direct Internet connection rather than wi-fi is recommended.
- Uploading more than one bag at a time is an option. However, doing so is not recommended. It can get messy tracking the progress status when multiple bags are being uploaded at one time.

UPLOADING A BAG

- Authenticate through the VPN (see VPN and ECS instructions in Appendix B).
- Open the S3 browser.
- Open Dell ECS platform.
  - Select your bucket in the window on the left (Example: SCLS_Archive or WiLS)
  - **Important: Highlight the save (level/location) where the folder should be added**.
    - Save location can be a sub-folder of an existing folder.
  - Select Upload, Select Folder (the bag that was created by DART, with the data folder and the four txt documents)
    - The upload will start immediately after selecting the folder.
  - Verify Failed = 0

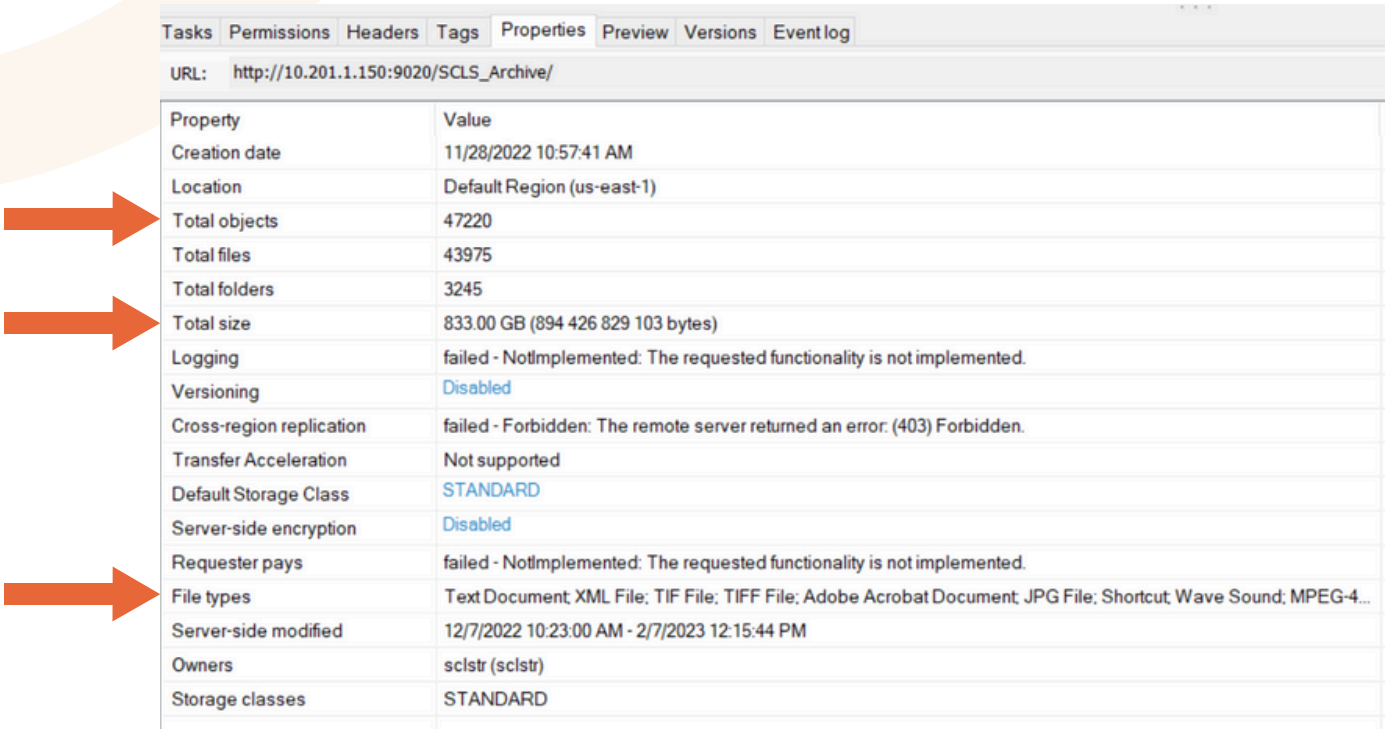
The bag is initially uploaded to the SCLS data center in Madison. It will be replicated to the LEAN data center in Eau Claire within 24 hours. ECS administrators will get alert emails if a replication fails and will share any alerts about any failing replications.



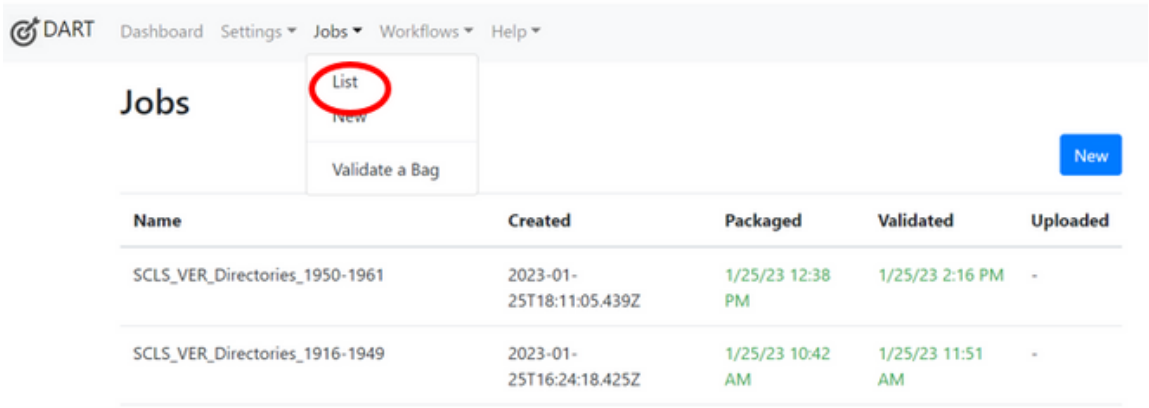
AFTER UPLOAD IS COMPLETED

- Check the **Properties** tab in the S3 browser for details about the upload. The information here can be copy/pasted into your Digital Archival Storage Log. Information from the DART manifest files (see p. 18) can be compared to information in these fields.
- Recommended fields to log include:
  - File Types
  - Total Size
  - Total objects/files/folders/

Note: You will see several fields (Logging, Cross-region replication, and Requester pays) marked as “failed” in the Properties tab. You can ignore these! They are the result of some features that are available in the Amazon S3 environment (which the S3 browser was developed to navigate) but not available in the Dell ECS environment.



- After a bag has been uploaded, the job can be deleted from the DART software.
  - Select **List** from the Jobs tab.
  - Click on a job name.
  - Select **Delete**.



If you have not done so already, save a copy of the text files from the bag folder (see p. 23). At this point, you can also delete the bag and the files that had been saved locally for the bag from your working computer.
Tip: Remember to empty the recycle bin after deleting a bag, since it takes up so much space!

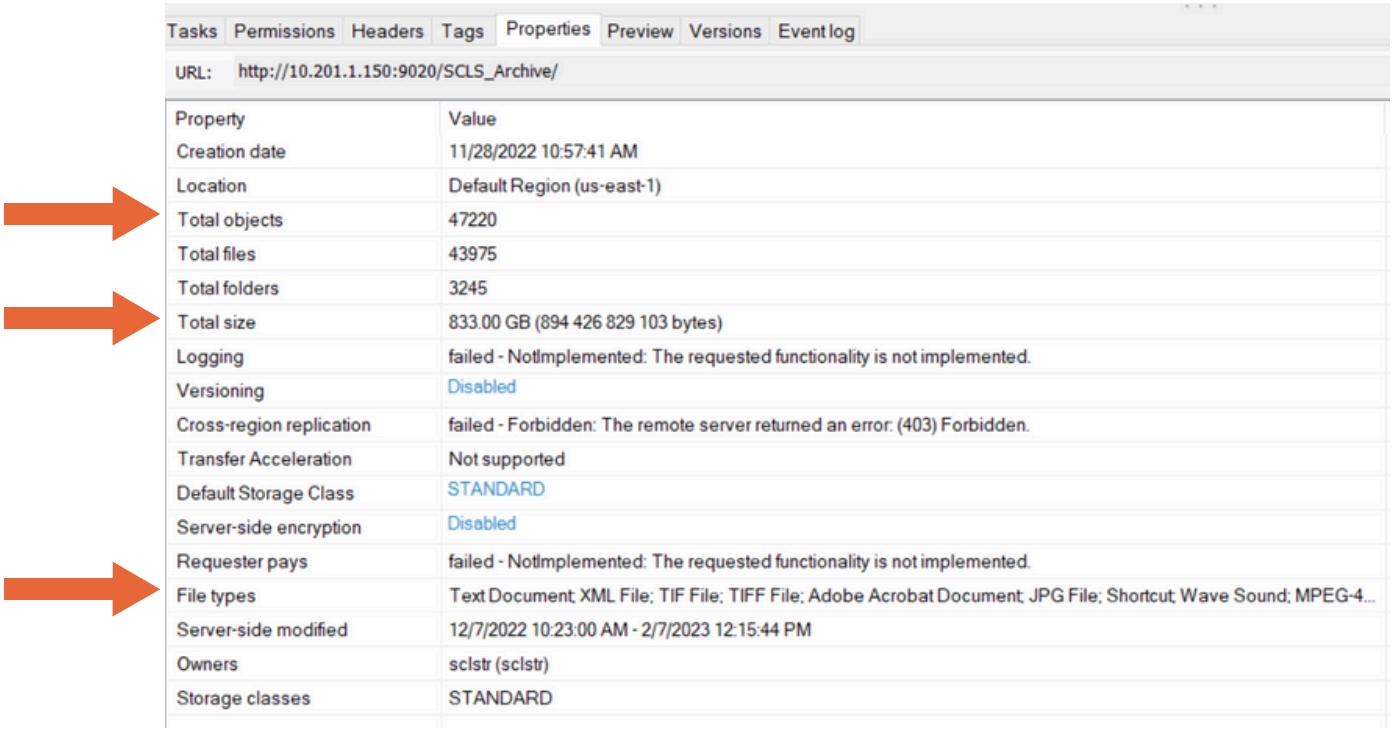
CHECK FILES IN STORAGE

The functions of the Dell ECS system include automatic validation and healing. In addition to this automatic monitoring, it's a good idea to run occasional manual checks on your files to confirm the data has remained unchanged. Each depositor is responsible for running these checks on their own schedule. We recommend spot-checking a selection of your stored files (or, if time and resources allow, a full check of all files) every 12 months.

Checks within the ECS System

A very quick, minimal way you can check your files is to review bag properties within the storage system.

- Authenticate through the VPN, open the S3 browser, then open Dell ECS platform (see Appendix B).
- Check the Properties tab in the S3 browser to view details about the bag. The information here can be compared to the information from the original upload in the digital archival storage log.
    - Recommended fields to compare include:
        - File types
        - Total size
        - Total objects/files/folders



Using DART to Audit Files

A more thorough and accurate way to check your files is by using DART or another checksum validator to audit the integrity of your files. For more information on file fixity and integrity checks, see the Store and Maintain section of the Recollection Wisconsin Digital Readiness Toolkit.

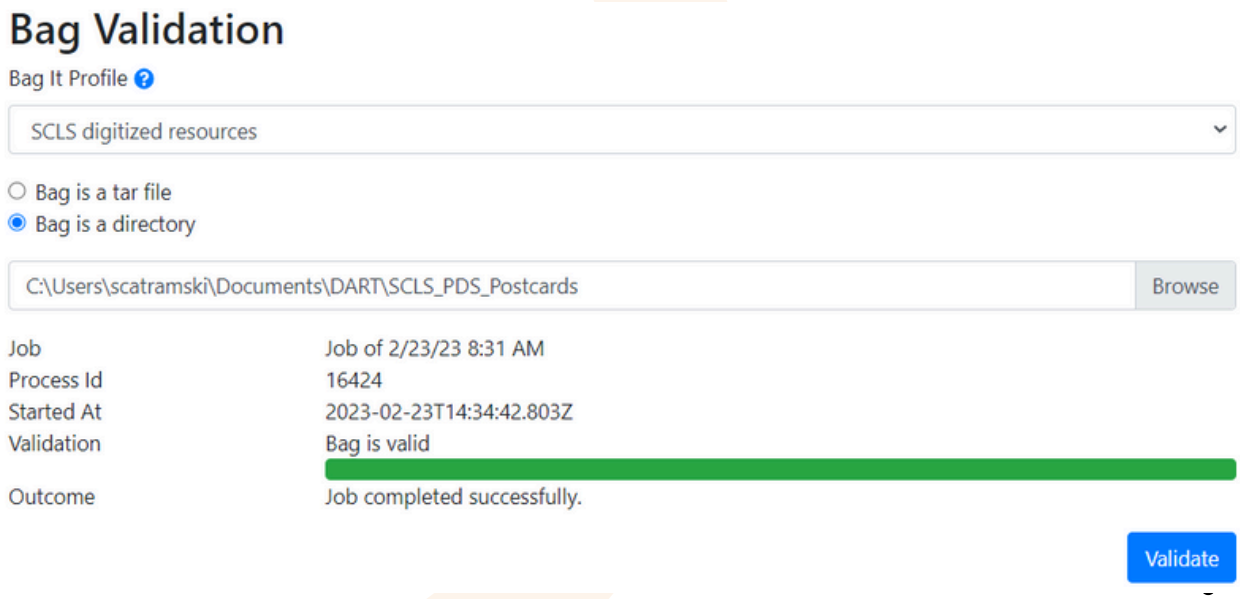
To perform a spot-check using DART, choose several bags in storage to download and re-validate.

- See download instructions in the Retrieve, Remove, or Update Stored Files section.
- Open DART
- Select **Jobs >> Validate a Bag**.
- Ideally, all bags would be periodically downloaded and verified. You can choose to do all bags at the same time, or do several per day or week over a period of time. This practice should be repeated annually.
- BagIt Profile: Select your profile
- Select: **Bag is a directory**
- Browse: Select the folder that was downloaded and **Upload**.
- Select: **Validate**
- DART should tell you that the bag is valid; the Outcome screen should say that the job was successfully completed.

If the validation fails, first re-download the bag from the Dell ECS platform and attempt to re-validate.

If the bag still does not validate properly, continue checking other bags in the Dell ECS platform. You may need to re-upload the original files if there has been a failure.

Log the spot-checked files and results in the **Quality Control** tab of your digital archival storage log.

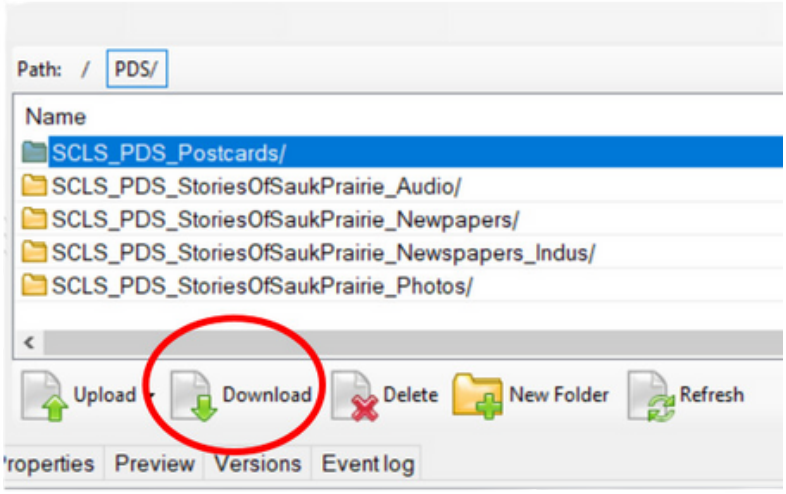


RETRIEVE, REMOVE, OR UPDATE STORED FILES

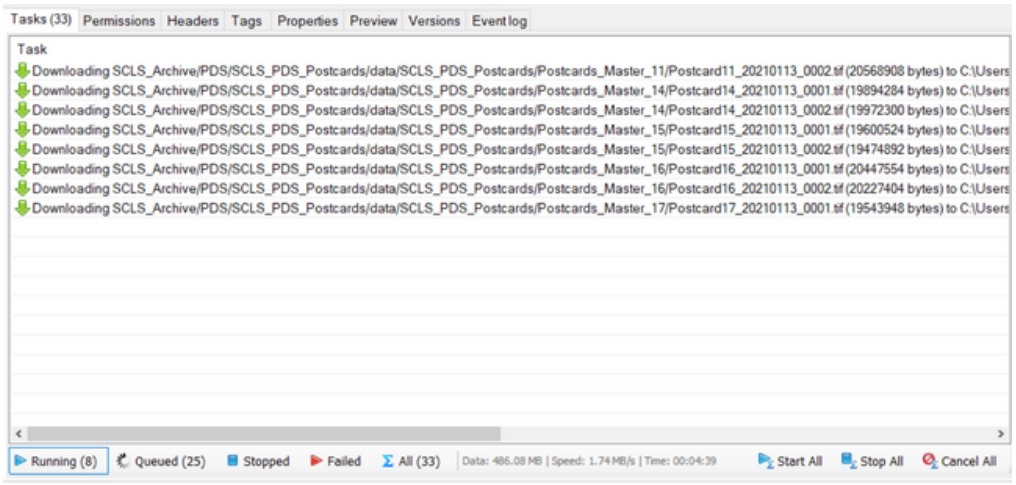
RETRIEVING FILES

In the event of a disaster or other local data loss, retrieving bags from storage may be necessary in order to replace a submitter's local files.

- Authenticate through the VPN (see Appendix B).
- Open the S3 browser
- Select the bag to download.
  - *Tip:* Check the total size of the bag by selecting the **Properties** tab to determine how much memory will be used by the download. Verify there is that much memory available on the save-to location.
  - *Tip*: **Downloading does not remove the bag from the Dell ECS system**. The bag remains in place and a copy is downloaded.
- In the pop-up window: **Select Folder**: Determine save location
- The save location should be on your local workstation. After re-validating using DART and virus-checking (see steps below), the bag's contents can be delivered to the submitter using your preferred file transfer method (i.e. FTP, external hard drive, etc.).
- After clicking **Select Folder**, the download will begin.

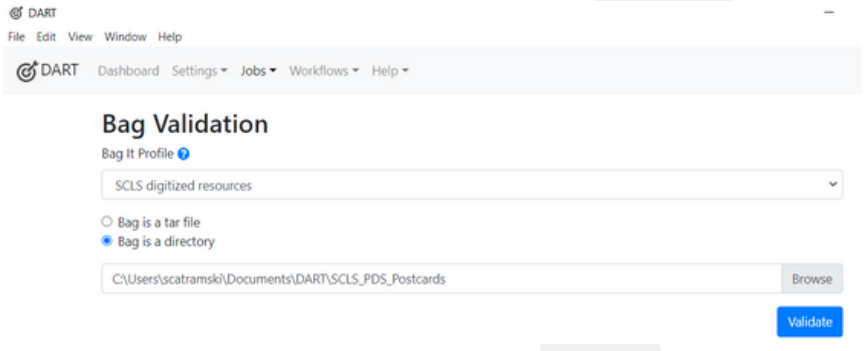


- The download can be monitored by selecting the Tasks button.
- When the download is complete, the **Queued** number will be **0**.
- Check that **Failed**=0.
  - A possible reason for a failure is if there is not enough memory on your save to location.
- The download speed will vary depending on your Internet speed. Even small folder sizes can take an hour or more to download.
- To stop or cancel a download, the buttons are on the lower right-side of the screen.
- If a download is stopped, it can be resumed by selecting **Start All**. It is recommended that a download not be stopped AND the computer shut down, as the download may result in failures after reconnecting to the server.

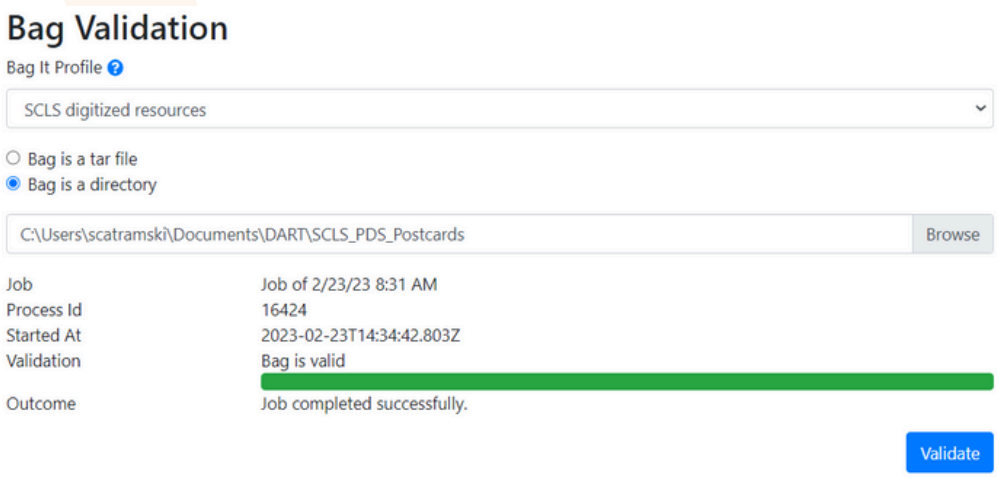


After downloading, the bag should be re-validated using DART.

- Open DART
- Select **Jobs >> Validate a Bag**.
- Bag It Profile: Select your profile
- Select: **Bag is a directory**
- Browse: Select the folder that was downloaded and **Upload**.
- Select: **Validate**



DART should tell you that the bag is valid; the Outcome screen should say that the job was successfully completed.



If the validation fails, first re-download the bag from the Dell ECS platform and attempt to re-validate.

After the bag has been validated, run a virus check on the bag using your tool of choice **before** delivering the requested files to the submitter.

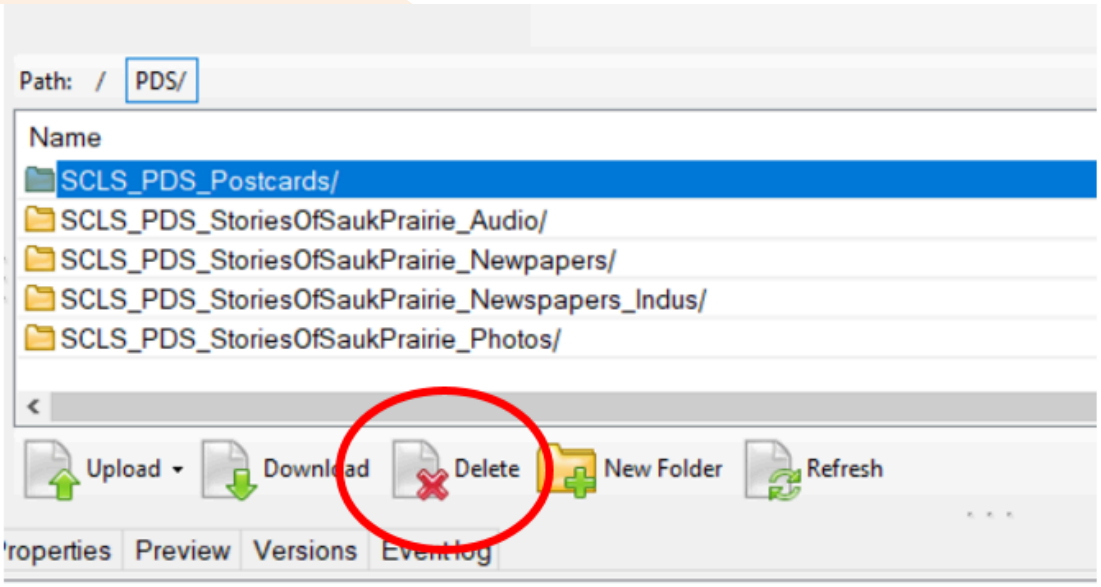
Log the retrieval request and details about fulfilling the request in the digital archival storage log.

REMOVING FILES

Removing individual files from a bag is not possible. All files that were bagged together have to be deleted together. If a single file needs to be deleted, the bagged folder must be deleted. Then the materials can be re-bagged and re-uploaded without the deleted file.

To delete a bag:

- Authenticate through the VPN, open the S3 browser, then open Dell ECS platform.
- Select the folder you want to delete.
- Select Delete.
- A window will open: Confirm File Delete
- Select Yes or No.
- Note: **The folder will be immediately deleted when Yes is selected. The file does not go to a recycle bin and there is not an option for undo or recovery after selecting Yes. No other administrator confirms the deletion of these files, so be very sure you want to delete this bag before selecting Yes.**
- Log deleted bags and reasons for deleting on the digital archival storage log.



UPDATING FILES

To add newly digitized files, updated files, updated metadata, or to include any other changes to stored files, you will:

- Create a new bag (using DART) **with all of the previously uploaded files PLUS the newly created or updated files and/or metadata. ALL of the files from this collection will then be housed in this new bag.**
- Move the new bag into the Dell ECS platform.
- Delete the original bag from the Dell ECS platform.

Appendix A: Configuring BagIt (DART) Profile

You should only need to set up this profile one time at the beginning of your project. You'll download a profile template and customize it for your own organization.

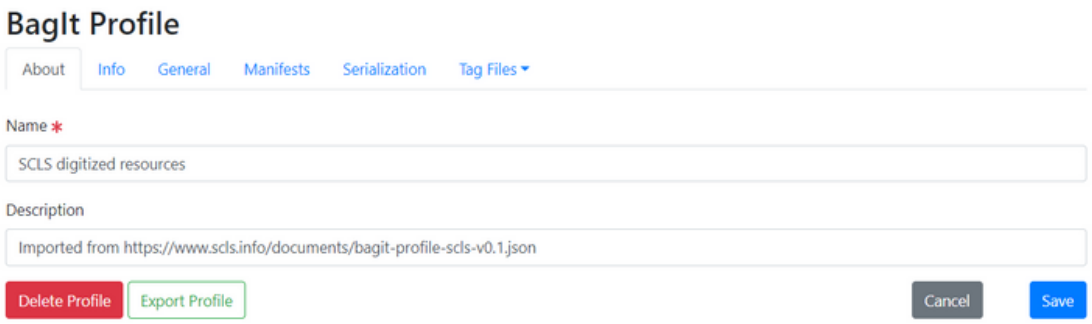
Set save location for bags in this profile:
In DART, go to **Settings >> Application Settings**
Use the default location or create new a new folder, for example:
C:\Users\scatramski\Documents\DART

Import profile template
Go to **Settings >> BagIt Profile**.

- Import profile
- Import JSON code by cutting and pasting the text in this document.
  - This is based on the South Central Library System profile, but you will customize to your own settings in the next step.
- Paste the JSON code
- **Import**

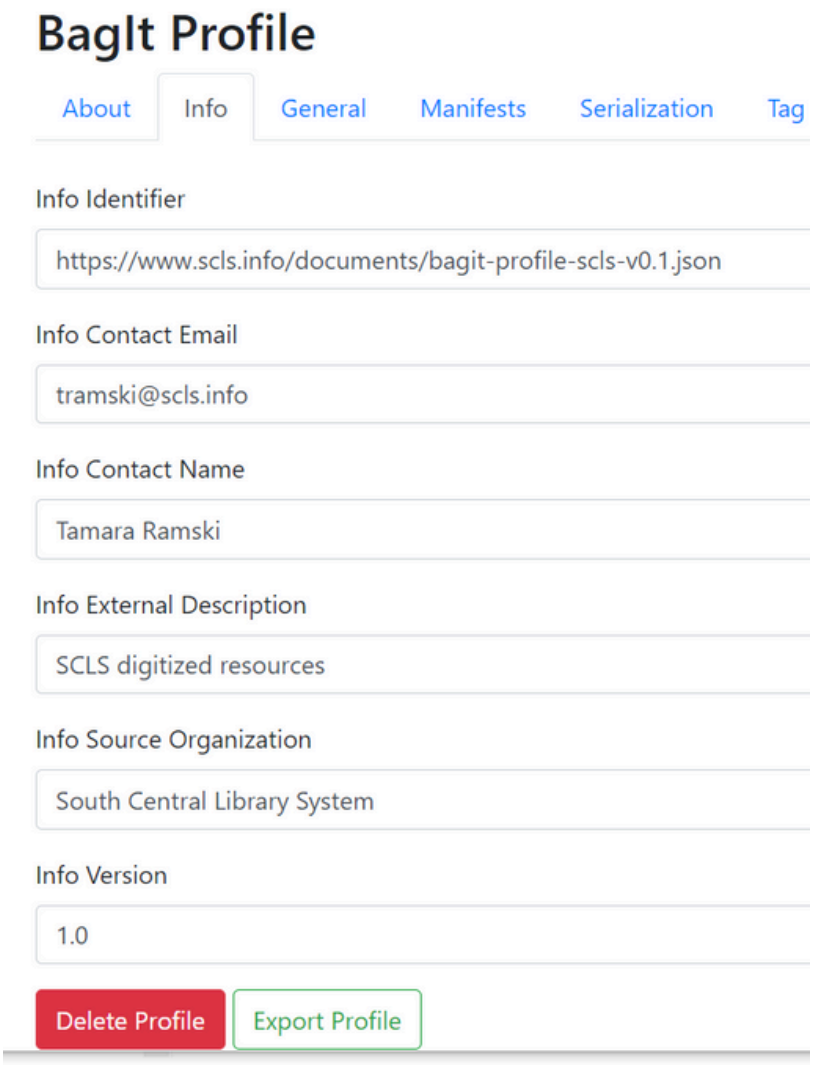
Customize profile
Select the newly imported profile and set up the following configuration:

- **About**
  - **Name** - for your system's profile, i.e. SCLS Digitized Resources
  - **Description** - same as the profile name, a description of what the profile will be used for
  - **Save**



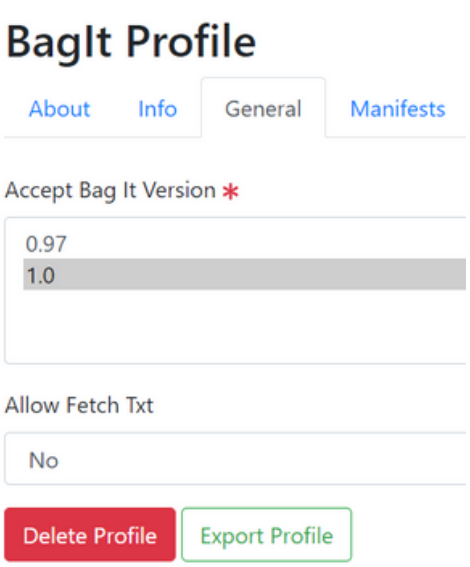
- Info (see screenshot below for SCLS example data)
  - Info Identifier - the profile identifier - that is, the permanent URI where the profile will live online.
  - Info Contact Email
  - Info Contact Name
  - Info External Description = Brief description of the function of the profile.
  - Info Source Organization = Full name of institution.
  - Info Version

To add more detail to your profile specifications, see this guide.



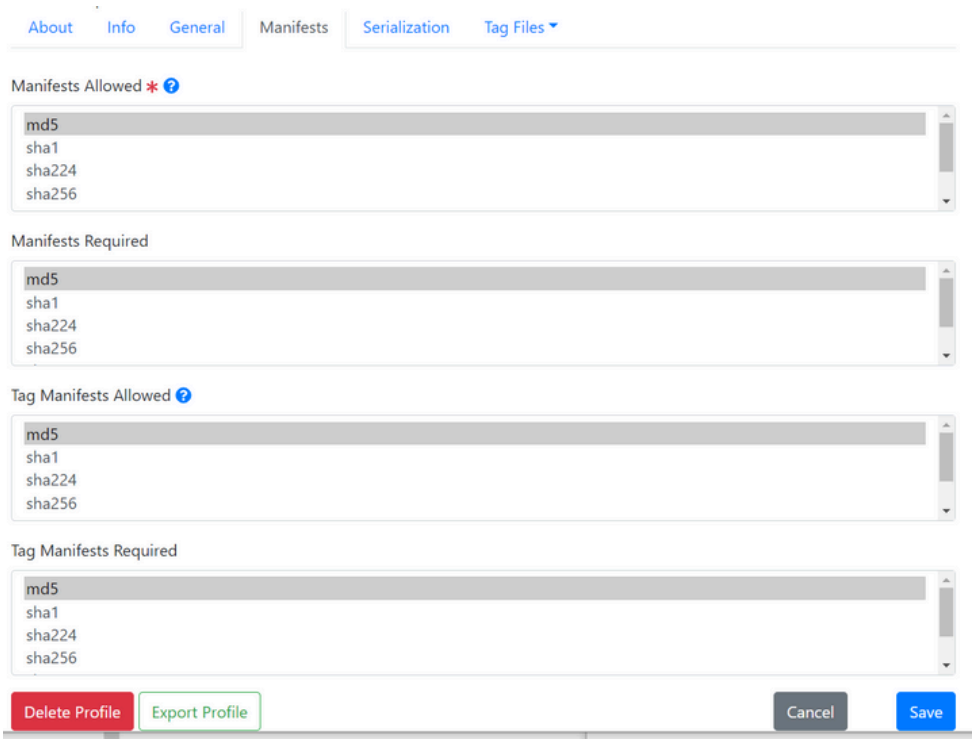
General

- **Accept Bag It Version** = 1.0
- **Allow Fetch Txt** = No



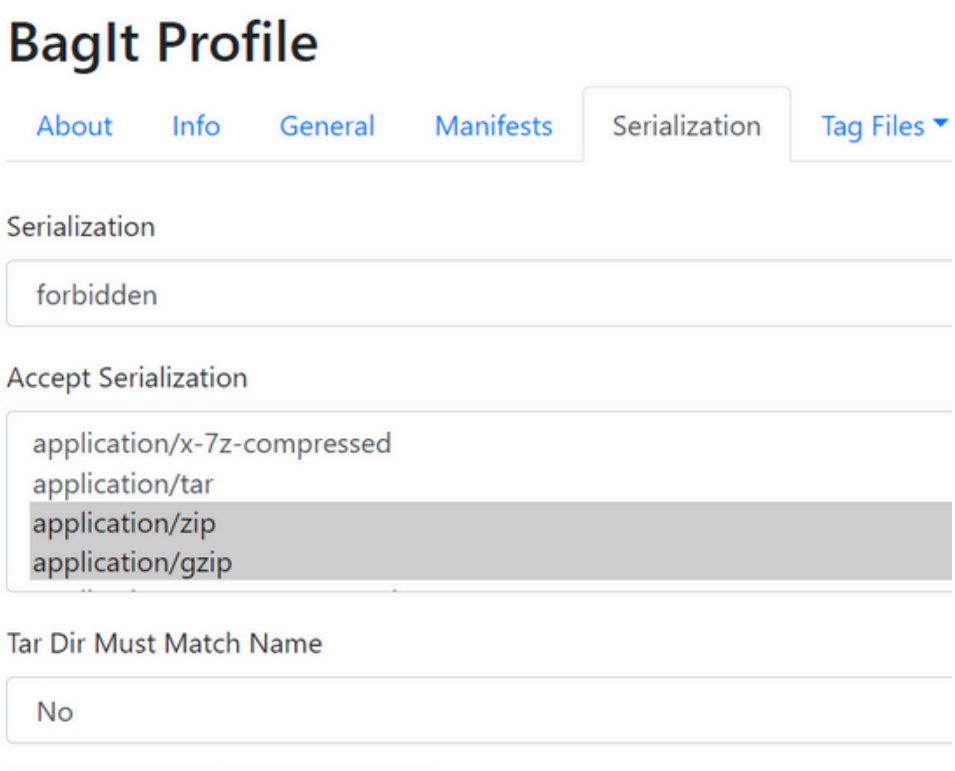
Manifests

- Select **md5** (and only md5) for all
- Select **Save**

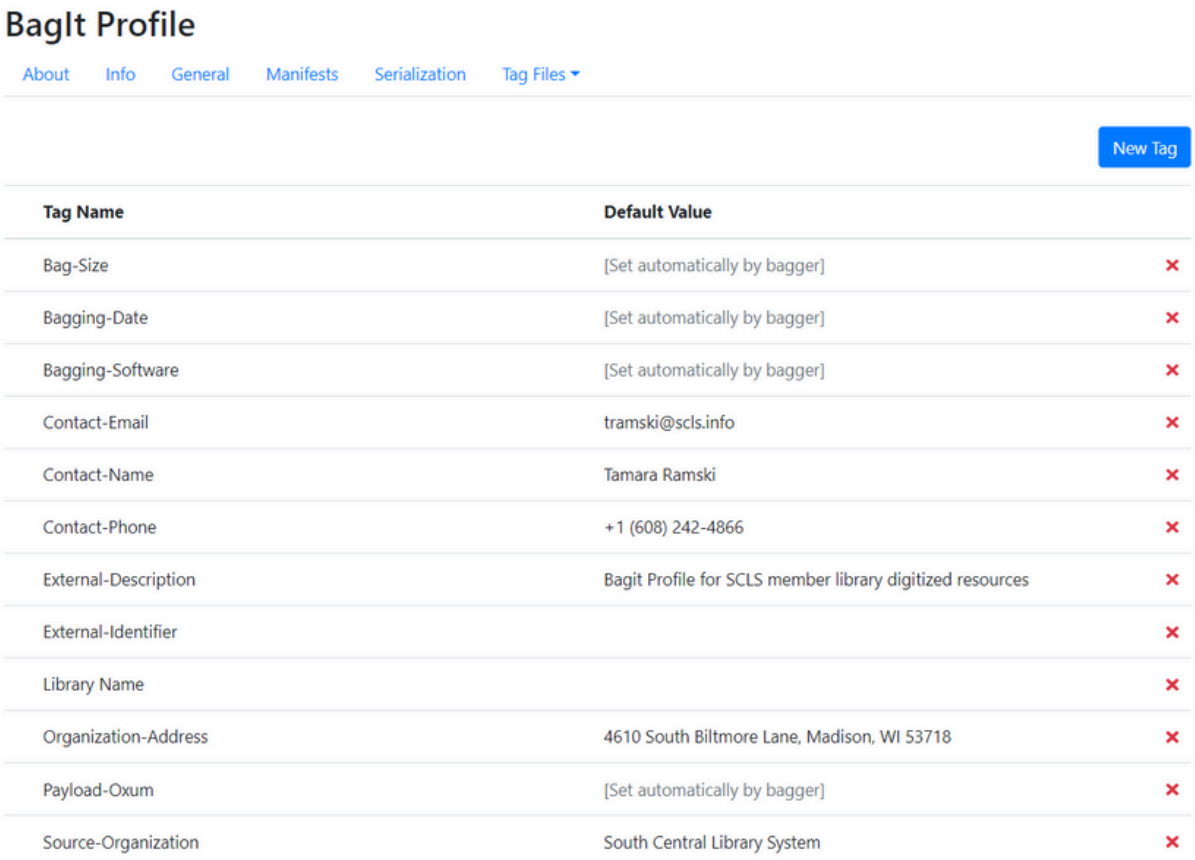


Serialization

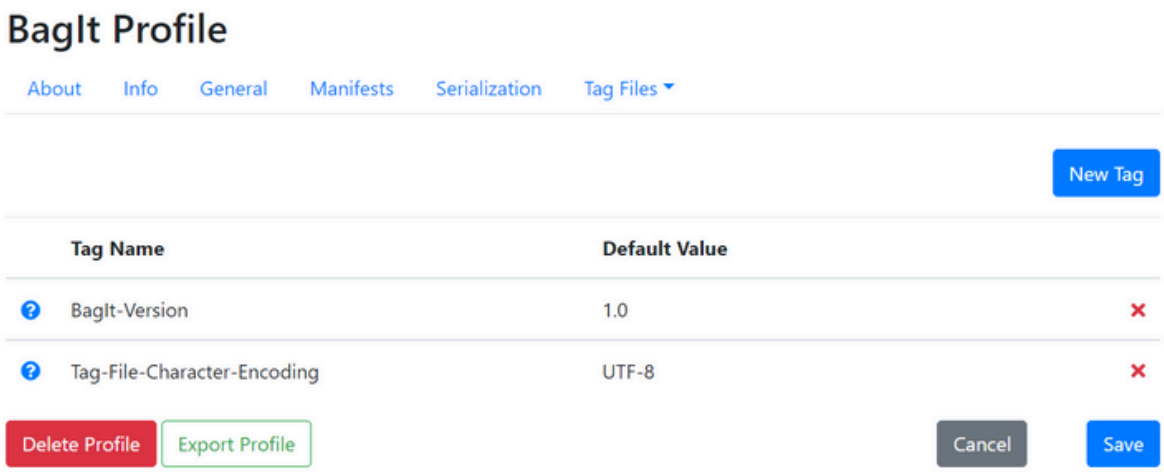
- **Serialization** = forbidden
- **Accept Serialization** = select (only) application/zip and application/gzip
- **Tar Dir Must Match Name** = No



- Tag Files
  - Bag-info.txt (Leave as default, or click on a Tag Name to edit it, or add New Tag) See below for SCLS settings
  - Add Library Name identifier tag for each separate library
    - Select bag-info.txt
    - Select New Tag
    - Tag Name = Library Name
    - Required = No
    - Values = Leave blank
    - Default Value = Leave blank
    - Save
  - The following tags can be deleted:
    - Internal-Sender-Description
    - Internal-Sender-Identifier
    - Bag-Count
    - Bag-Group-Identifier



- Bagit.txt (Leave as default or click on a Tag Name to edit it)
  Example: See screenshot on previous page



Other DART settings in the **Settings** menu: Leave unchanged.

Once your profile is updated to your specifications, export the JSON code and publish it online in a permanent location.

- **DART >> Settings >> BagIt Profiles**
- **About** tab
- Click "**Export Profile**"
- Copy the resulting code.
- Publish publicly online to a web server with a URL that will be unchanged so that bag validators can retrieve it and validate the profile.

Appendix B: Accessing ECS with VPN

These instructions outline the steps needed to download, log in to, and use both the recommended VPN and the Dell ECS interface to upload and manage files in the digital archival storage platform. Note: the software only works on Windows-based PCs.

Credentials

Anyone who accesses the ECS will receive their own VPN and ECS credentials. Please contact WiLS (kristen@wils.org) or or South Central Library System (andrew@scls.info) to request credentials, reset passwords and to remove access when an employee leaves your system.

Downloads

You will need to download two software installers.

FortiClient VPN

1. Go to https://www.fortinet.com/support/product-downloads
2. Download FortiClient VPN for Windows.

S3 Browser

1. Go to https://s3browser.com/
2. Download the most recent version of the S3 Browser Freeware.

NOTE: We highly recommend purchasing the Pro Version license of the S3 Browser. The free version allows a maximum of two concurrent downloads and uploads at a time. The Pro version allows you to set this number higher. You will notice a difference in transfer speed. As of late 2024, the Pro Version license was $40.

Install FortiClient VPN

1. Double-click the installer you downloaded and run through the install wizard.
2. After the install completes, open FortiClient VPN.
3. Accept the Licensing Agreement.
4. Configure VPN connection
    a. Click the **home button** at the top-right of the screen.
    b. Click **Configure VPN**.
        i. Connection Name: Anything works here, but let's go with **ECS Access**.
        ii. Remote Gateway: **205.213.104.2**
        iii. Check the box next to Customize port and enter **10443**.
        iv. Client Certificate: None
        v. Authentication: It's up to you if you want your Username to be saved**.**

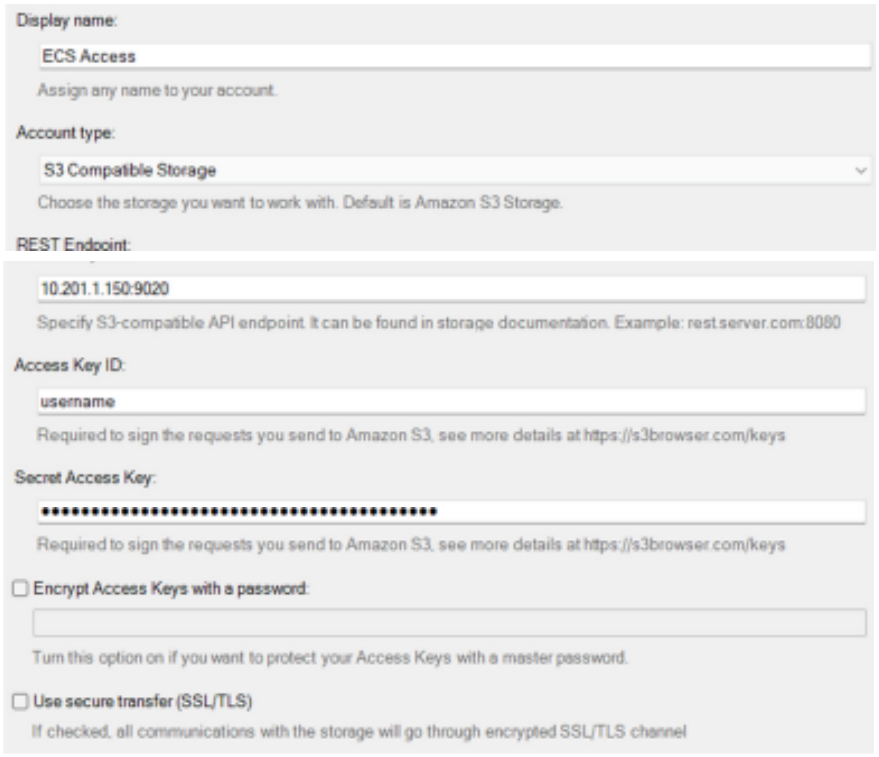
        i. Click **Save**.
5. Turn off Certificate Warnings.
    a. Click the **lock** at the top-right of the window.
    b. Enter administrator credentials or click **Yes** on the UAC prompt.
    c. Click the **settings gear** at the top-right of the window.
    d. **Uncheck the box** for **Do not Warn Invalid Server Certificate**.

Authenticate the VPN

1. **Right-click the FortiClient** shortcut in the notification area of your taskbar.
2. Select **Connect to "ECS Access."**
3. Enter your credentials and click **Connect.**

Configure S3 Browser to Access the Dell ECS

1. Open S3 Browser.
2. **Add New Account** will appear. Configure as shown below. **Access Key ID is your ECS Username.**



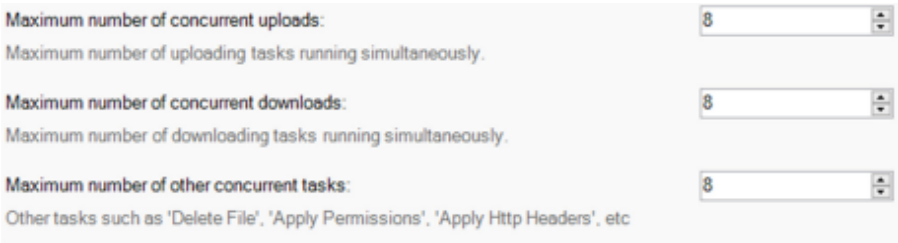
3. Click **Add new account**.
4. S3 Browser should automatically connect to the ECS now.

Install an S3 Browser Pro License

1. Click **Help**, then **Activate Pro Version**.
2. Enter your License Key.
3. Click **Activate**.

Increase the Number of Concurrent Tasks (Pro License Only)

1. Click Tools, then Options.
2. Select the Queueing tab.
3. Set the concurrent uploads/downloads and tasks to a number between 8-10 if you would like. This should improve performance.



Enable Verbose Logging

Verbose logs improve the ability to diagnose issues with the service. The logs are stored locally on your PC in .txt files.

1. Click **Tools**, then **Options**.
2. Select the **Logging and error handling tab**.
3. Set **Log Verbosity level** to **Verbose**.
4. Click **Save Changes**.

If you have questions, concerns, or need troubleshooting, please contact

Appendix C: Participation Agreement

Download a fillable version of this form here.

This agreement is between collaborating Wisconsin public library systems, acting as individual systems, together with an established library system partnership, the Libraries and Enterprise Applications Nexus of Wisconsin ("LEAN WI"), the South Central Library System ("SCLS"), and the contributing library system ("Depositor").

A Memorandum of Understanding and an Addendum to the Memorandum of Understanding for this project exist separately and outline the terms and responsibilities of participating parties. Additional roles and responsibilities are detailed in the Service Model. Your signature below affirms that you have read and understand the Service Model, understand each party's roles and responsibilities, and agree to abide by them.

This Participation Agreement is for the purpose of documenting the Depositor's contacts and preferences regarding the digital archival storage materials.

Depositor contact information
Who is responsible for preparing and contributing the digital files to the servers?
Name: _____
Phone number: _____ Email address: _____
Job title: _____

Who is a secondary contact responsible for preparing and contributing digital files?
Name: _____
Phone number: _____ Email address: _____
Job title: _____

Contact information will be verified annually, and updated if necessary.

Content Monitoring
As described in the Service Model, 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.

Beyond the automatic monitoring described in the Service Model, 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.

Please describe your organization's plan for monitoring content while it is at rest in the storage system:

Does your organization intend to add new content or update existing content in the storage system? If so, how often?

Content Withdrawal
Depositors may withdraw from the MOU by providing a 12 month notice of intent to leave. When a request is submitted to withdraw from the program, all existing content on the servers will be bagged and transferred to the Depositor.

If the program ceases operation, a 12 month notice will be given to depositors and existing content will be bagged and transferred to the Depositor.

_____
(Authorized signature of Depositor)

Name:
Institution:
Date:

_____
(Authorized signature for Statewide Digital Archival Storage Project)

Name:
Institution:
Date

Appendix D: Succession Policy

Scope

This succession policy covers how content managed by the digital archives backup service will be disposed of in the event of the dissolution of the WPLC Digital Archives Backup Collaboration Workgroup, or in the event that a Depositor wishes to withdraw from the service. The policy applies to only and all digital content owned and/or collected by Depositor and placed in the service for long-term management by the WPLC Digital Archives Backup Collaboration (known hereafter as “the project”), a project of the Wisconsin Public Library Consortium.

This policy does not cover who assumes ownership of content in the event of dissolution/change of the depositing entity.

This policy does not cover changes in technical infrastructure, as the WPLC Technology Collaboration Steering Committee reserves the right to change the infrastructure over time to meet current requirements, expectations, and budgetary constraints.

Objectives

This Policy establishes a plan for the orderly dissolution of the project and for the disposition of deposited digital content, should those steps prove necessary. It also empowers the WPLC Technology Collaboration Steering Committee, under the guidance of the WPLC Board, to approve and oversee a transition of the service to another entity for the operations and management of the project.

Policy
In the event that the WPLC Technology Collaboration Steering Committee terminates the service entirely or transitions the service to another provider, or a Depositor wishes to withdraw from the service, the following options are available to the Depositor:

1. **A Depositor may elect to have the digital content they previously deposited in the project preservation storage environment returned to them.**
    - Depositor must notify the WPLC Technology Collaboration Steering Committee in writing that they wish to withdraw from the service, with a minimum of 60 days notice.
    - Project staff will determine any required actions, such as the return, transfer, or conveyance of assets.
    - The WPLC Technology Collaboration Steering Committee will be available to consult, advise, and help with a transition to the Depositor's chosen storage solution.

2. **A Depositor may elect to have project staff delete the digital content they had previously deposited in the preservation storage environment.**
    - Depositor must notify the WPLC Technology Collaboration Steering Committee in writing that they wish to withdraw from the service, with a minimum of 60 days notice.
    - WPLC Technology Collaboration Steering Committee will inform the Depositor when the digital content will be removed and what content will be removed.

In any of the above scenarios:
- The Steering Committee will notify the Depositors and provide Depositors with a proposed transition plan **fourteen months before the end of its responsibility for the service**.
- Depositors will have ten months before the Steering Committee's last day of responsibility that their deposits must be removed from the project's preservation storage environments by that last day.
- If the Depositor does not remove its deposits by the deadline, project staff may delete those deposits.

Policy last reviewed May 2024

Appendix E: Service Model

This document outlines the original vision for the Digital Archival Storage System approved by the WPLC Digital Archives Backup Workgroup in 2022 and 2023. Much of this information is included elsewhere in this handbook, but the service model is shared in its entirety here for reference purposes.

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. 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).

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.

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.”

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.

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).

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.

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:

1. One bucket for each depositor (bucket owner = primary contact at depositor institution).
    a. One folder within the bucket for each submitter (individual library or cultural heritage institution).
        i. 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.

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." (see documentation here)

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.

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.” (see documentation here)
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.