

## Montana State Library

# Transferring New Spatial Data to the Dark Archive

December 27, 2012 (author: Diane Papineau)

Document Status: Final Draft, ready for internal use

The purpose of the Montana State Library (MSL) spatial data Dark Archive is to store a copy of data that MSL has accepted into its collection. The Dark Archive is a volume in the GIS\_Archive network share alongside three other data management folders, which are made available primarily for processing older MSL clearinghouse data into the Dark Archive (Figure 1). In the future, an offsite replica of the Dark Archive may serve as its security copy and restore point.

**Note:** *The Dark Archive is a preservation repository. The Dark Archive volume may only be accessed for data integrity management, published data restoration, and reappraisal purposes. No Clearinghouse patron tools may link directly to data stored in the Dark Archive.*

All data placed in the Dark Archive will remain in the Dark Archive in perpetuity with few exceptions. Data that is securely archived by another organization (i.e. NAIP) may be removed from the MSL Dark Archive later through a re-appraisal process. Incorrect (obsoleted) data and superseded data placed in the Dark Archive will remain there as a record of data accepted into the collection and served to patrons.

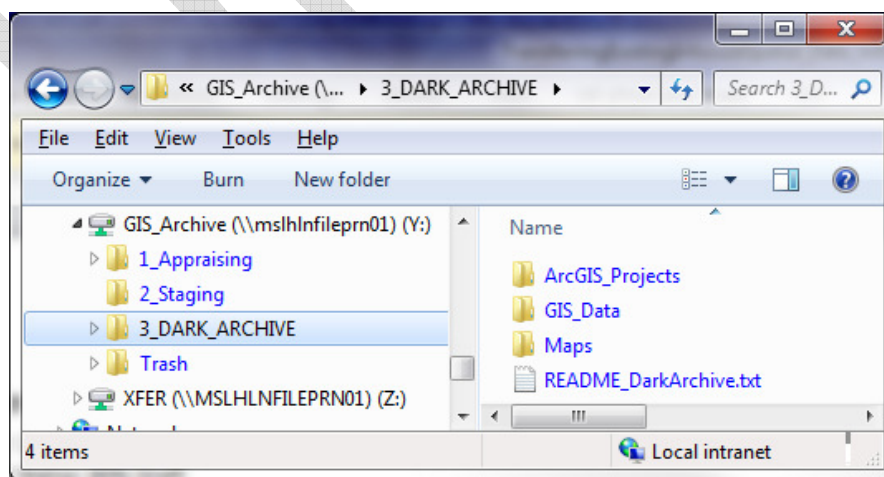


FIGURE 1: GIS\_Archive Folders Including Appraising, Staging, and DARK\_ARCHIVE

*To process new data into the Dark Archive:*

1. Appraise the data's integrity and usefulness. The data must be ready for patron use and meet the following conditions. If necessary, modify the dataset and/or its metadata to meet these conditions before proceeding to Step 2. The dataset must:
  - Meet the requirements of the MSL Collection Development Policy.
  - Be free of viruses/malware. Manually run a virus check on the data.
  - Successfully load in ArcMap.
  - Have a complete metadata record in XML form that meets the state metadata standard.
  - Have a documented coordinate system/spatial reference.
  - Have understandable attribute field names that are present in the metadata and defined.
  - Use the file naming convention.
2. Appraise the data for data management concerns:
  - Is the data related to another dataset we collect (i.e. is it part of a series or a collection)? If no, proceed. If yes, make note of this for metadata work associated with data in series and data collections.
  - Will this data supersede another dataset in a series? If no, proceed. If yes, make note of this for data management work associated with updating Published Data Packages.  
**Note:** Superseded data is data that was expected to change (data in a series), which is different from obsoleted data that had inaccuracies and is being replaced.
  - Will this data make a currently-offered dataset obsolete? If no, proceed. If yes, make note of this for data management work associated with updating Published Data Packages.
  - Will the data be archived by another agency? If no, proceed. If yes, make note of this for metadata work associated with determining and recording a reappraisal date.

- Can the data be freely distributed? If yes, proceed. If no, determine what the limitations are and make note for metadata work.
  - Determine how the data will be served and make note of this.
3. Prepare all the necessary metadata. Store a copy of the published metadata record with the dataset at its current location.
  4. If the data is not part of a series or collection, proceed. If it is part of a series or collection, update the parent metadata record or create a new parent record if necessary and store it for use in the data management and data discovery systems.
  5. Using Windows Explorer and Bagger, package the data for archiving:
    - a. In Windows Explorer, create a compressed zip folder on the drive where the data currently resides. Name that zip folder using the file naming convention.
    - b. Place the dataset contents in the zip folder. The zip folder should include the dataset and the published metadata record.
    - c. Open Bagger, click Create New Bag. In the dialog that displays, set parameters as version 0.96 with no profile.
    - d. Click the Payload green plus icon (Figure 2), navigate to and select the Zip file for bagging, then click Open.

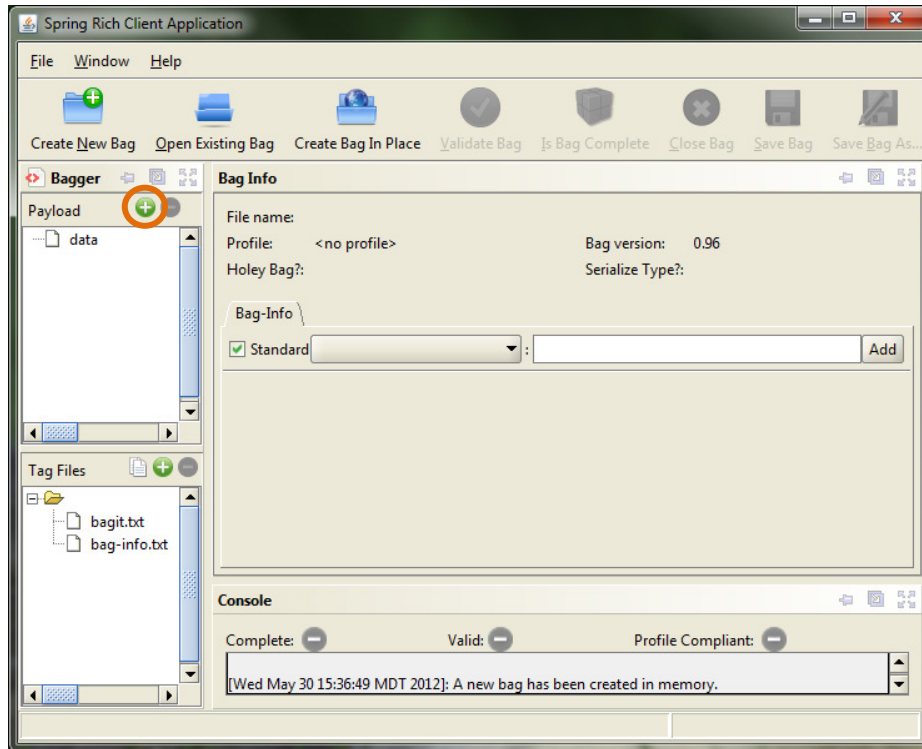
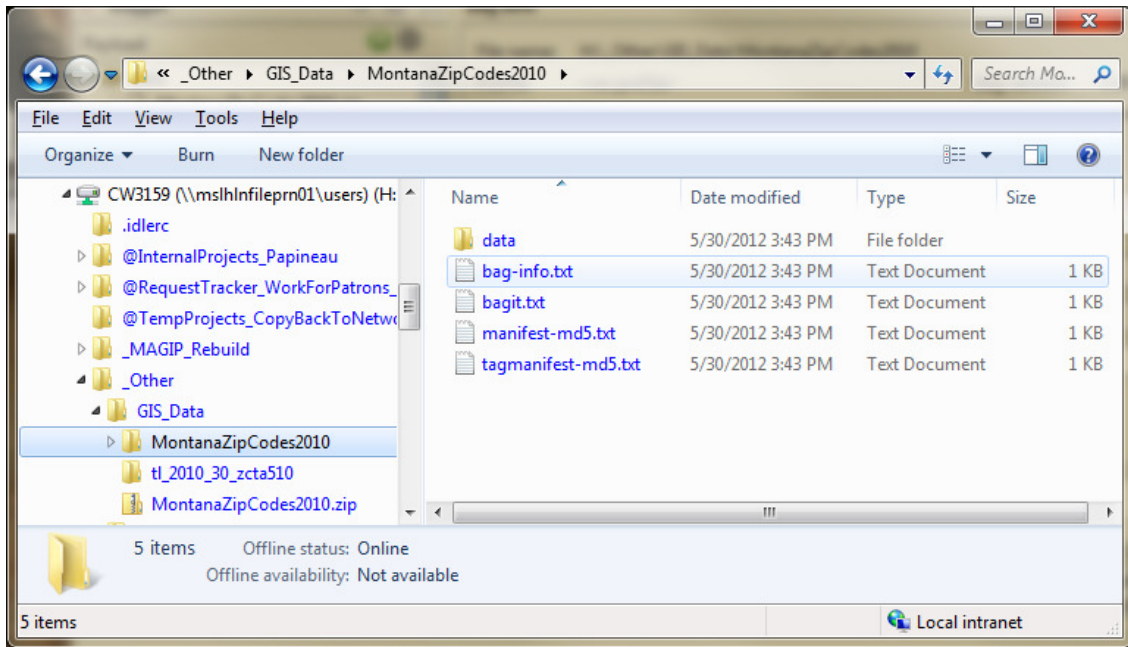


Figure 2: Bagger Interface

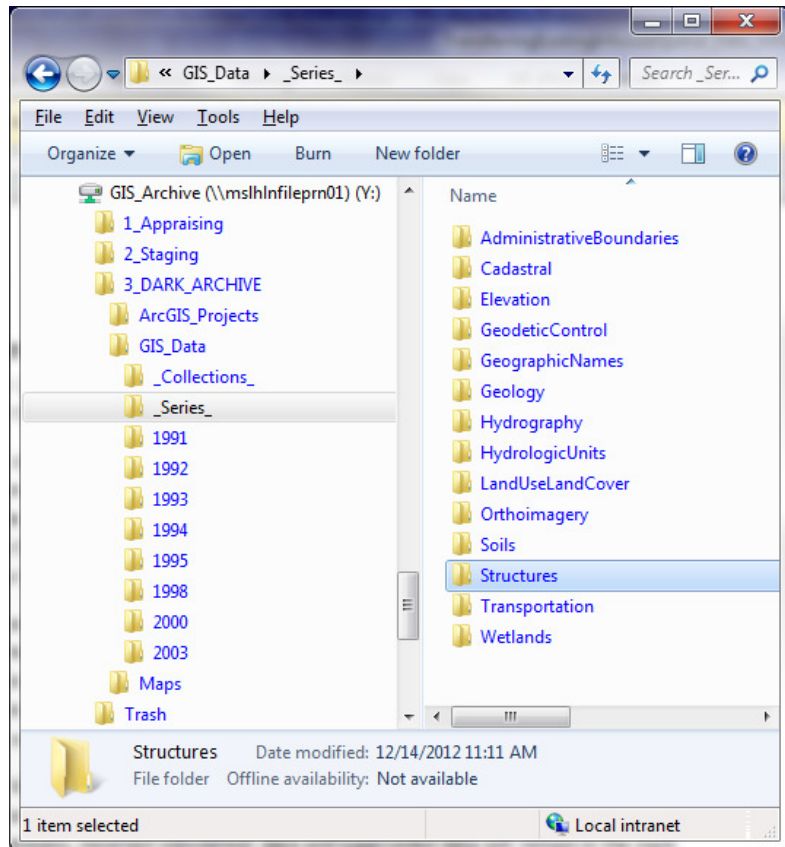
- e. Save the bag: File>Save Bag As.
- f. In the Save Bag Dialog, use the Save in: browse button to choose the location where the bag is currently stored and name the bag the same as the zip file.
- g. Accept defaults in the Save Bag Dialog, but ensure that the dataset will not be zipped by Bagger and choose OK. Dismiss the Bag Saved Successfully dialog.

**Note:** An archive “bag” looks like a Windows folder. This new archive bag uses the naming convention and it contains the dataset and published metadata record (zipped up) in a data folder, plus other Bagger-generated files such as an MD5 checksum (manifest-md5.txt). See Figure 3.



**Figure 3: Bagger Generated Files with Data Folder for Transferring**

6. Navigate to the new archive bag and copy it.
7. Navigate to the appropriate folder in the Dark Archive and paste the archive bag. If the folder needed for this data is not yet present, create it.



**Figure 4: MSL Dark Archive**

8. Open the transferred bag and load it into Bagger. From Bagger:
  - a. Click Open Existing Bag.
  - b. Navigate to and select the newly-placed archive bag in the Dark Archive.
  - c. Click Open. The new bag of data loads in Bagger.
9. Confirm that the data transferred to the folder successfully
  - Click the Validate button, then dismiss the Validation Successful dialog.
  - Click the Is Bag Complete button, then dismiss the Bag is complete dialog.
10. Close Bagger.
11. Enter information in the data management system or spreadsheet including the following: Title, Time Period, Larger Work Citation, Date Archived, Checksum, Last Archive Review Date, File Location, Data Format, Data Format Version, Next review date, MSDI (Y/N), MSDI theme.