Bagger

Overview

Bagger is a desktop software tool developed at the Library of Congress that packages files into a specific directory structure, called a bag, for transfer and digital preservation. When an online bookstore sends your order, they take your books and pack them in a box, add your shipping information to it and include a packing slip listing what books you've ordered. Bagger acts in a similar way: it takes your files and packs them in a directory structure, lets you add an information form to the shipment and includes a list of all files included in the package.

In the same way that a physical package can be damaged in the delivery process, digital packages can also be damaged. Sometimes an electronic transfer is cut-off mid-stream and files are missing. In other cases, you may have received all of the files but one or more have become corrupted. Bagging with Bagger lets you create an inventory before it leaves one place and then validate that inventory after arrival at the destination. Checking the inventory tells you quickly if something went wrong during transfer. Likewise, if there are problems with a file after a bag has been validated, the problem can be localized to the sender's original file, rather than the transfer process.

What Is a Bag?

In general a bag is a group of digital files of any type (metadata, media, text, etc.). It can be an arbitrary grouping and does not have to be in a specific file/folder structure. A Bagit bag is also a group of digital files of any type. The difference is that it has a file and folder structure specified by the BagIt standard (http://tools.ietf.org/html/draft-kunze-bagit-06). The Bagger software application creates BagIt bags.

1. The top-level directory is the packaging for the shipment. It is the equivalent of the box for shipping. All of the files and packing lists will be inside this directory. The name given to this directory is generally considered to be the name of your bag.

   There is no standard naming practice for BagIt bags. A general guideline, though, is consistency and uniqueness, particularly if you are going to be using the bags for digital preservation in long-term storage. WDL generally names its bags with a partner code and the year and month of delivery (qatar_2011_06).

2. The data directory holds all of the files you are planning to deliver. These files include metadata in any format, media of any type and accompanying Word or text documents. They can be single files or directories and multiple levels of subdirectories. This collection of files is collectively called the payload. During the bagging process, you choose the files to be included in the payload and Bagger places them automatically in the data directory.

3. The bag-info.txt file is a text file that is a lot like the shipping label on a physical package. It has information about who sent it and to whom it's going as well as information about the bag generated in the bagging process. Bagger creates this file and there is a window to edit information during the bagging process. You can also use Bagger to edit this information after the bagging has been completed.

4. The bag-it.txt file is another text file that Bagger creates with information about the bag itself. There is no need to do anything with this file.

5. The manifest-md5.txt file is a text file that is the equivalent of a packing list for shipping. It is an inventory of files in the payload along with their checksum and is automatically generated.

   A checksum is a type of digital fingerprint for a file. The validation process after transfer creates a new list of checksums and compares it to the old one. If the fingerprint has changed, then something has been corrupted and a replacement needs to be requested. The validation process also checks the inventory to see if files were lost or added in the process. The most common type of checksum is md5, which refers to the name of the algorithm that creates it. There are other types but for transfer and digital preservation, md5 is most widely used. The text of the file looks like this:

```
1fce6033e68ff6bc2a45da317ce9dbc3  data/qatar_2011_06.mrc
18b786ad6ebd2b0ae164023a2dc0e6d7  data/852 HLM.MRK
e9b9fd83898b46f77d65dd04063202ce  data/810 HLM.MRK
736790981dcf5ee4b93ffac8631e0809  data/25488/0001.tif
c6a7de1aa7c34ee4737502ad1aceaa7d  data/25488/0002.tif
```

6. The tagmanifest-md5.txt file is a text file automatically generated by Bagger that does not need to be edited by the user.

Bagging

Bagger has two ways to create BagIt Bags:

1. Bag a Copy – this creates a copy of your original materials, which may be in one or more source locations, and puts them into a BagIt bag, leaving the originals untouched.
2. Bag in Place – This bags the original materials themselves.

Which Way Is Better?

Important factors in deciding which way you bag items are the amount of space on your computer and the number of source locations.

- You must have sufficient space on your computer to make a second copy of the materials or your machine can bog down or crash.
- If you have multiple sources of files, you can join them together when bagging a copy but not when bagging in place. You have to manually organize your source files together *before* bagging if you want to bag in place.
- If you have a need to zip your files, you need to bag a copy.
- If you need to use a project profile, you need to bag a copy.
- If you have "loose" files, ie, files not in a directory, then you must bag a copy. Bag in Place can only process at the directory level.
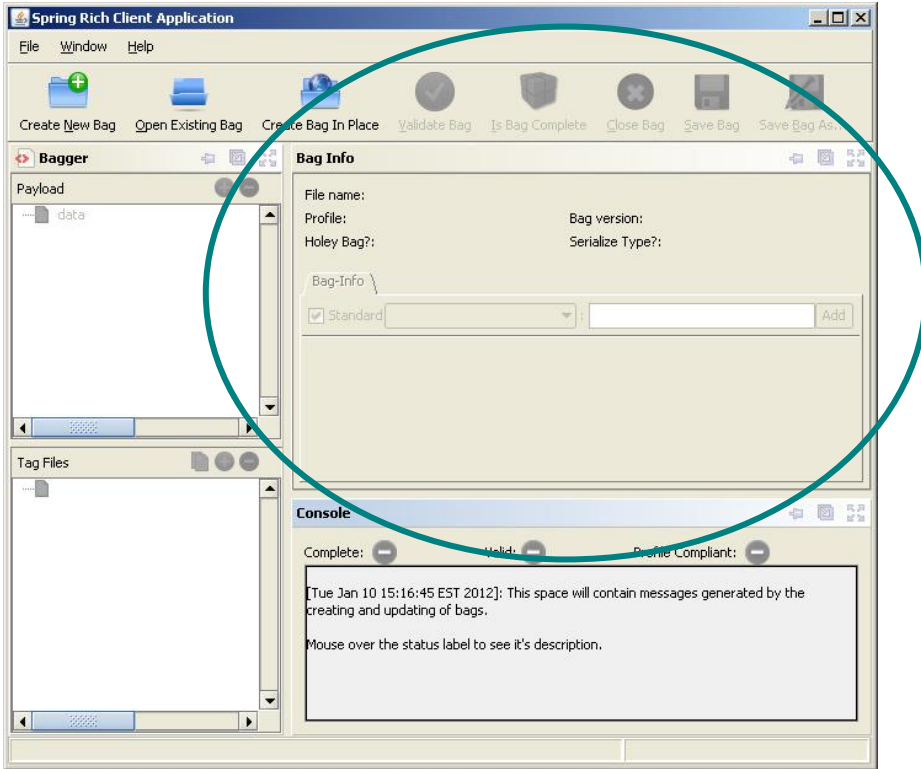
Bag a Copy

- At the main bagger screen, choose "Create New Bag".
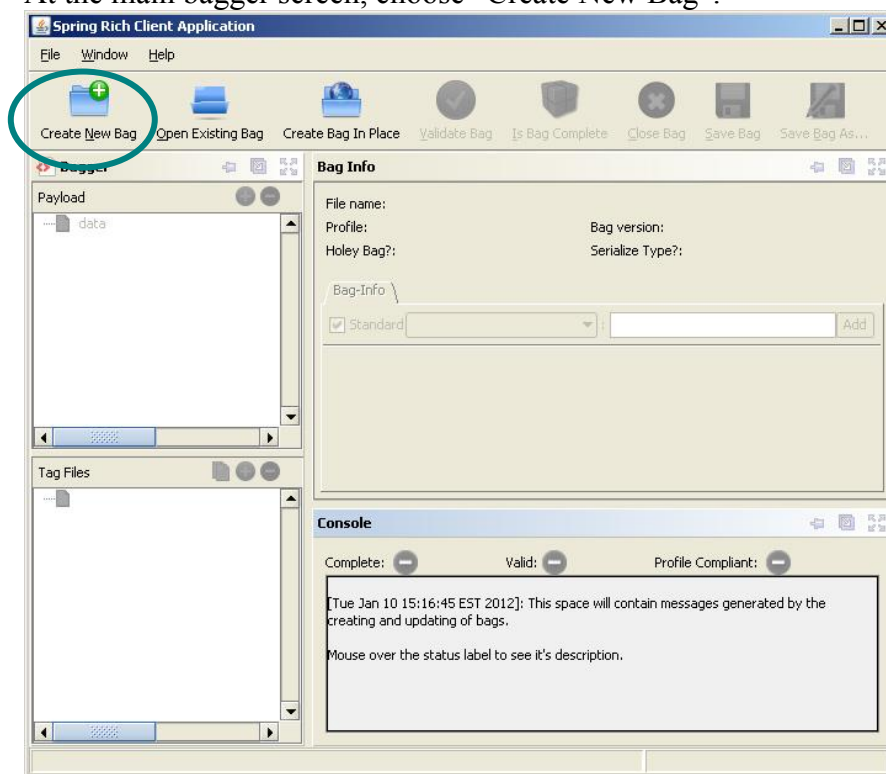
- In the pop-up screen:
    1. Leave the default bag version as is.
   2. Select "wdl" from the drop-down list of profiles.
    3. Click "OK".
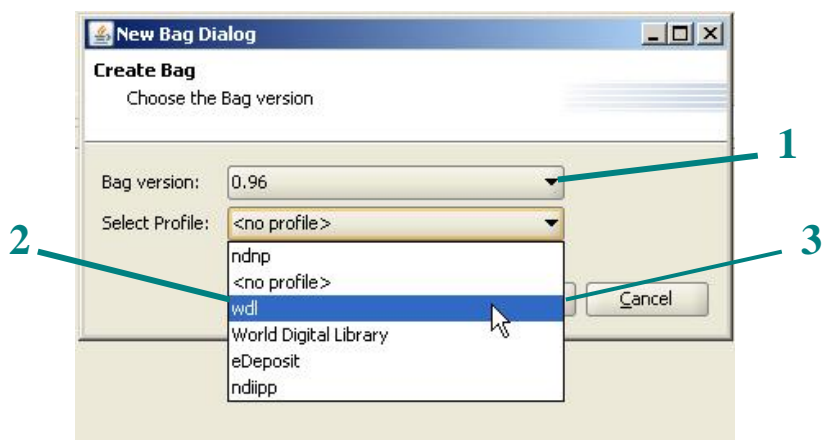
- Fill out the Bag-Info section of the main window. This is your "shipping label". The following fields are required to complete the bag:

1. Sender-Email – the contact person's e-mail in case we have questions.
2. Sender-Contact – the contact person (so we know to whom we should address questions).
3. Sender-Institution – We often get several deliveries at once and it helps us keep straight what belongs to whom.
4. Type-of-Transfer – Choose Hard Drive, CD/DVD or Network from the drop-down menu.
5. Sender-Address – The address of the sending institution. This is not required but is important for hard drive deliveries. We need to know where we should send the drive back.
6. The recipient (WDL) information has been already entered in the profile.

Bagger Guide
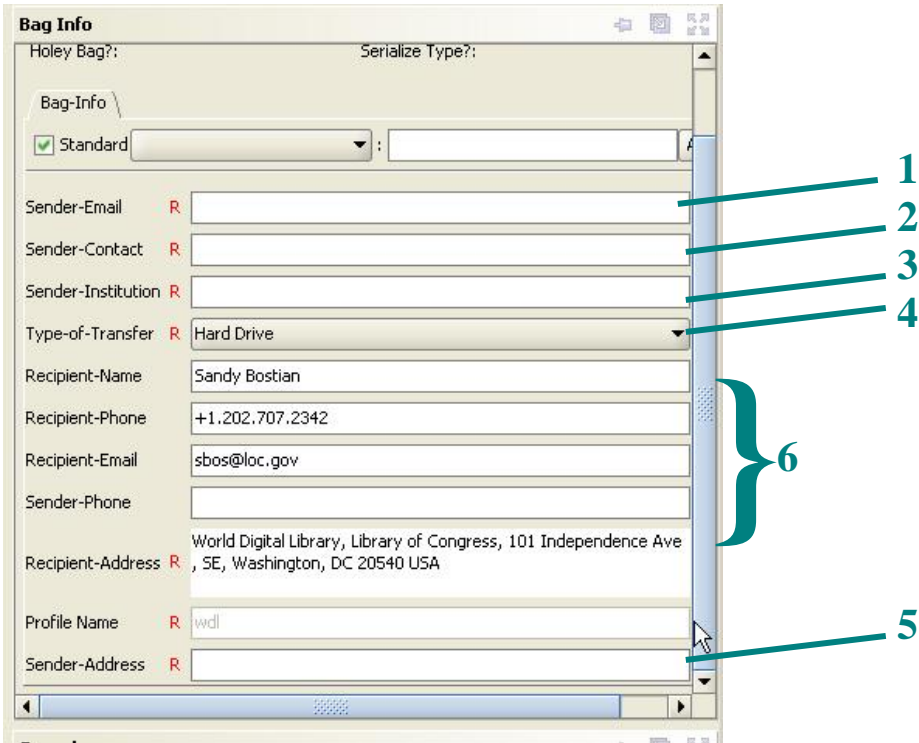
- The payload window shows you what files are in the data directory.
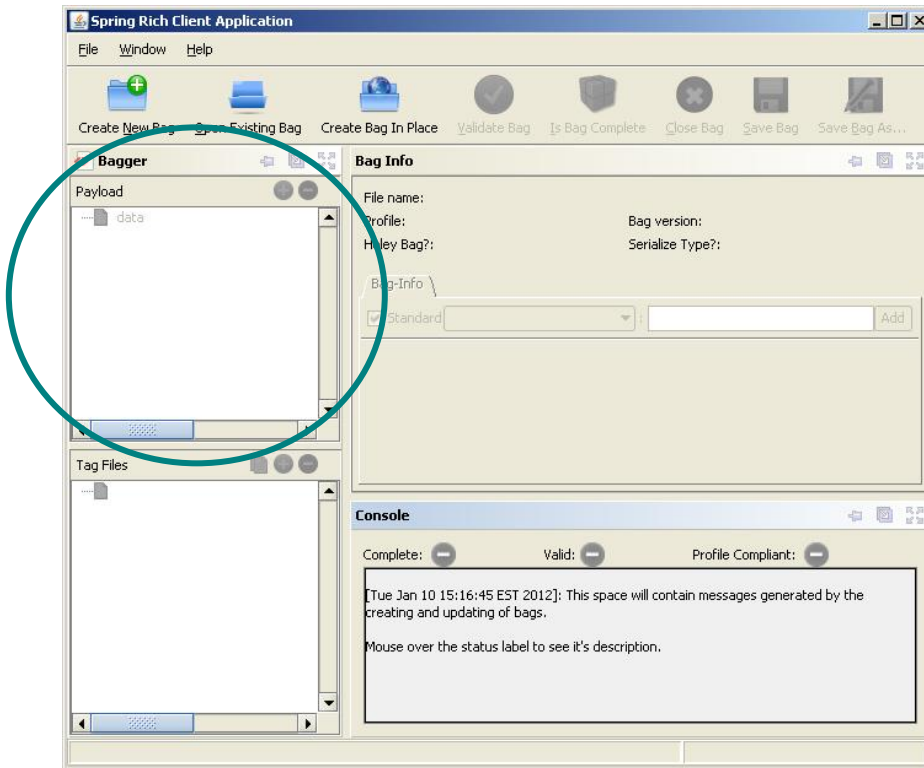
 o Add payload files by clicking the green plus sign (+) in the payload window.
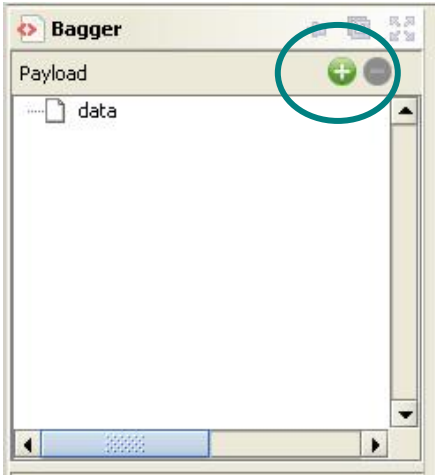
- Navigate to the directories or files you want to add through the pop-up window.
  - Directories can be added even though the label says "File Name:"
  - Multiple directories and files can be added at one time by holding down the shift button when selecting.
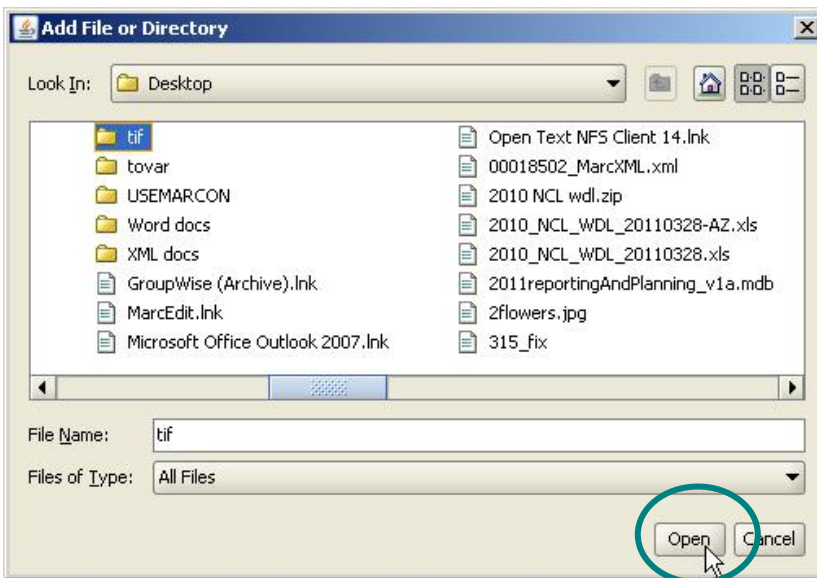

- Clicking open will add the chosen directory or file(s) to the payload window under the data directory.
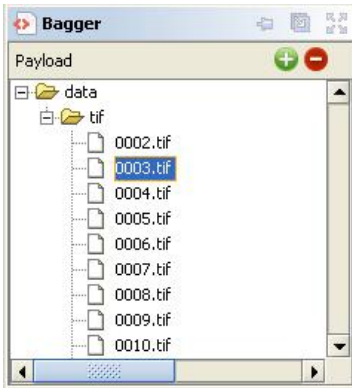
- To delete files from the payload, highlight the directory or file and click the red minus sign (-).
- To add files from another source location, repeat the above steps until the payload is complete.
- Your bag is not complete until you click the "Save Bag As" button at the top.
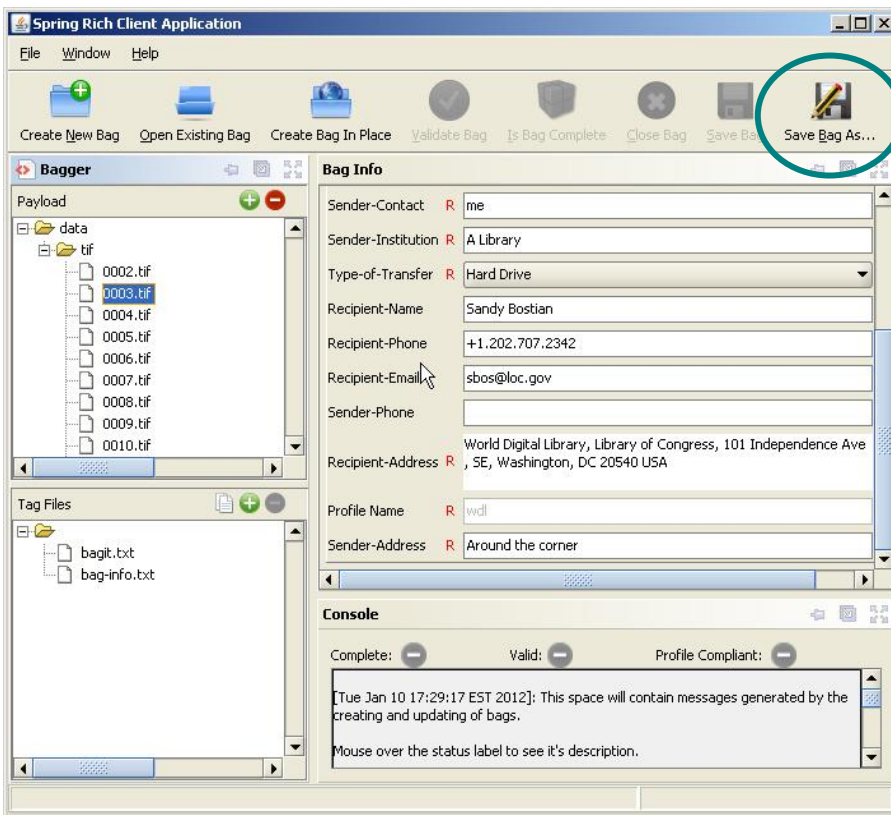
- The "Save Bag" pop-up window allows you to manage settings before final bagging.
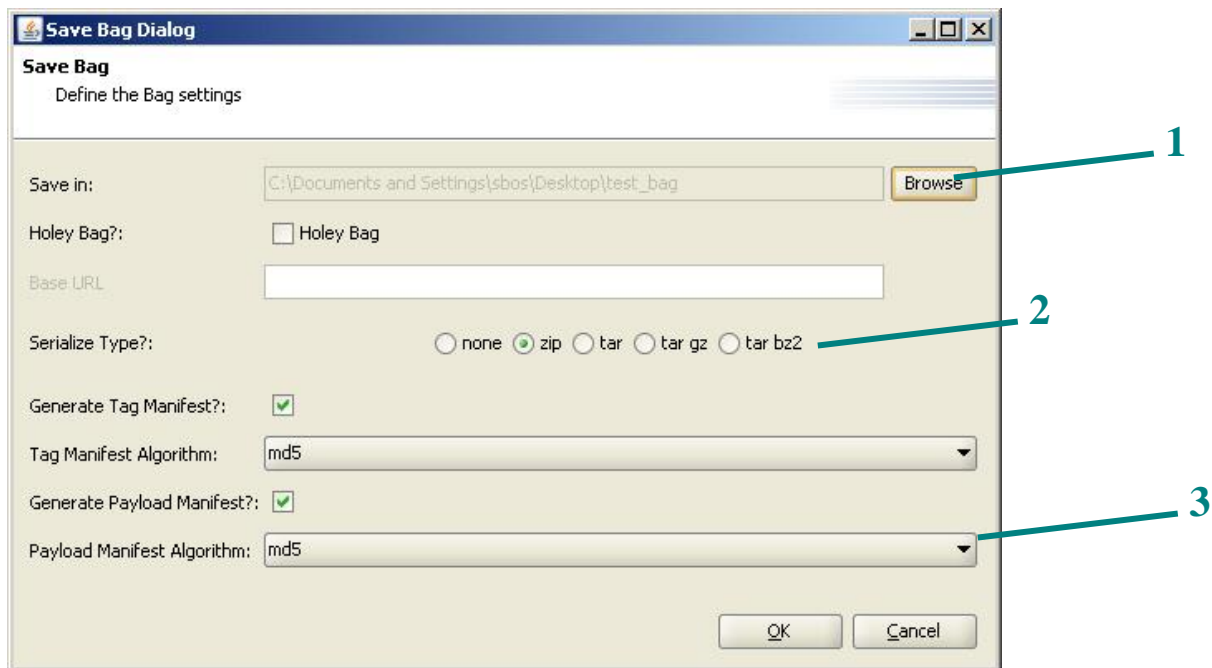
1. Save in – Click "Browse" to choose the location where your bag will be saved and to name your bag. This creates the top-level directory that holds the payload and accompanying information files (the packing list and shipping label) and is the equivalent of the shipping box. This is also where you name your bag.
    o You'll see another pop-up box:
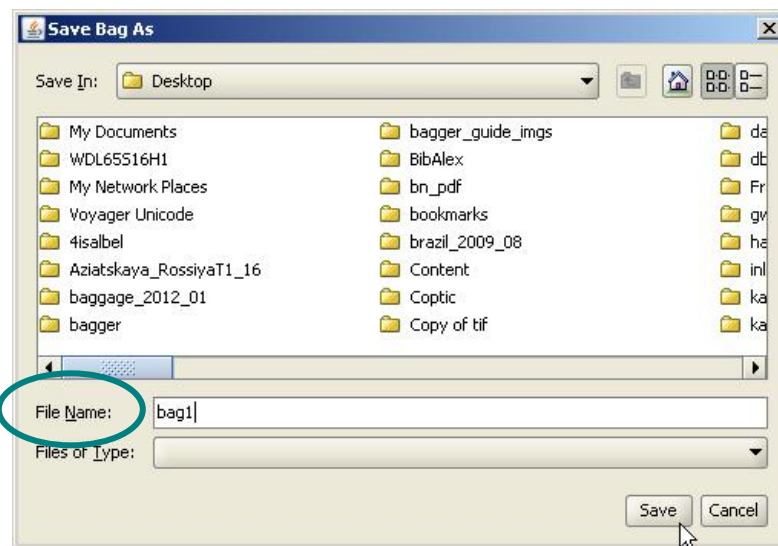
    o Navigate to the location where you want to save the bag.
    o Enter your bag name next to "File Name".
    o Click Save.
2. Serialize Type? – This determines if the bag will be bundled into a single file, what type of file that will be and if there will be compression. Choose the appropriate radio button:

- o None – no bundling, no compression (the same as it is on your desktop)
- o zip – bundling as a zip file (.zip) with zip compression. This type of bundling is useful if materials will be used/stored on Windows machines or servers.
- o tar – bundling as a tar file (.tar) with no compression. This type of bundling is useful if materials will be used/stored on UNIX/LINUX servers.
- o tar gz – bundling as a tar file (.tar.gz) with gzip compression (the more common tar compression format).
- o tar bz2 – bundling as a tar file (.tar.bz2) with bzip2 compression

3. Choose a checksum algorithm – This is where you can choose the type of checksum (or fingerprint) that will be generated. The default is md5 and, unless you have a need to change it, it's best to leave it that way. The most commonly used algorithm is md5.
4. Click OK.
5. There may be a pause while Bagger sets up the process. You'll then see a progress bar and, depending on the size of your bag, it could take a while to perform the bagging operation.
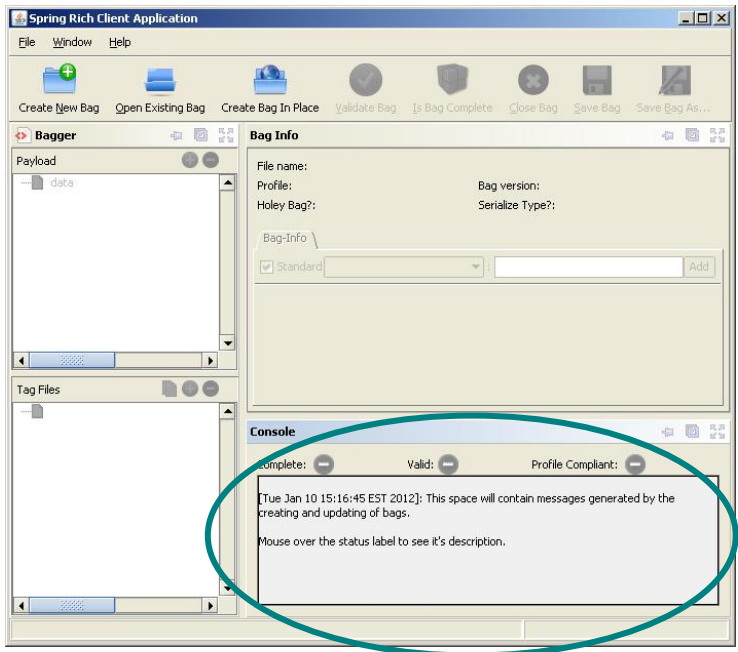

6. If everything goes well, you should see a pop-up telling you your bag was saved.

7. Click OK.

- The Tag Files window allows you to read the packing list and shipping information (bag-info.txt). Generally, you don't need to do anything with this.
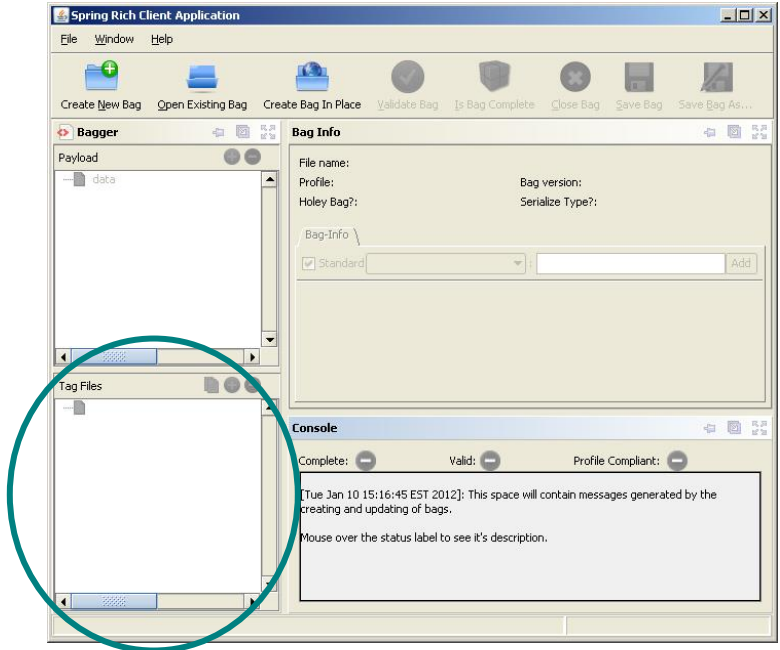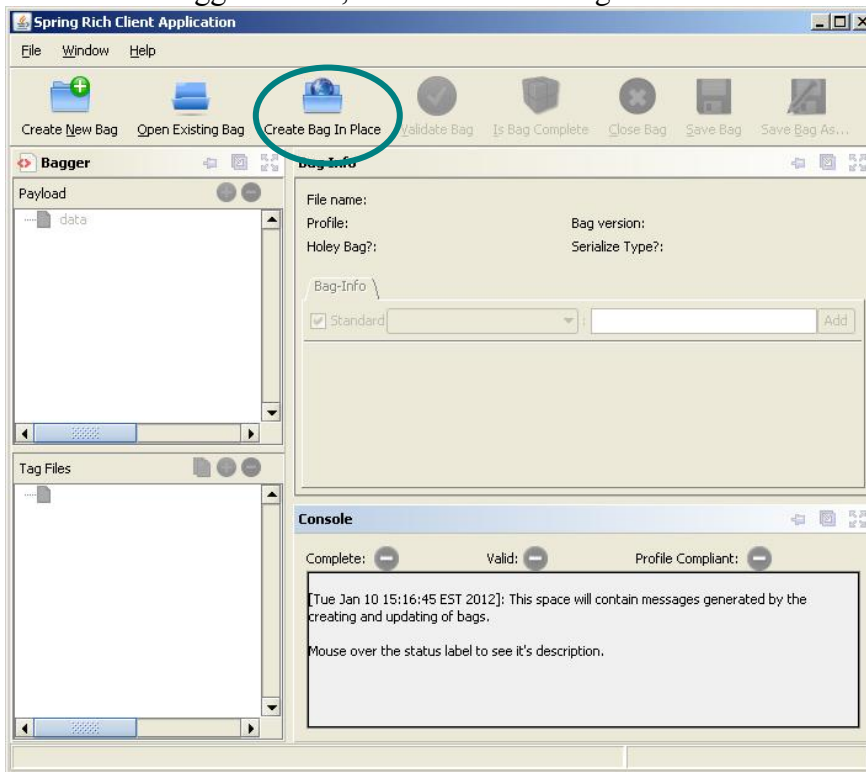
- The Console window gives output messages from the bagging process. Generally, you don't need to do anything with this.
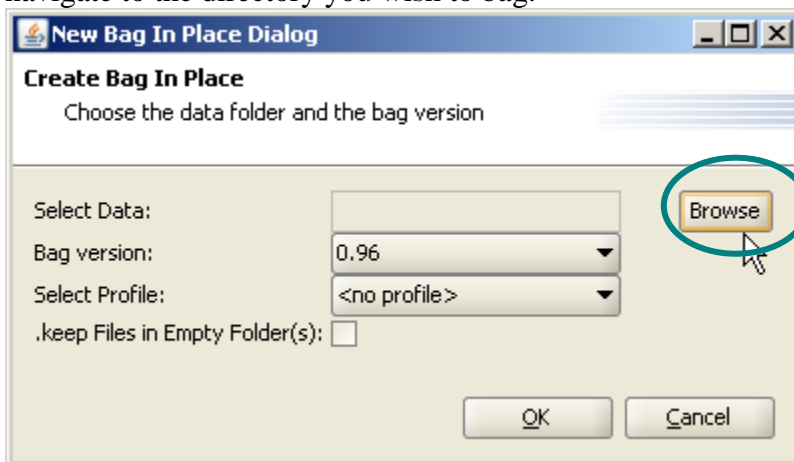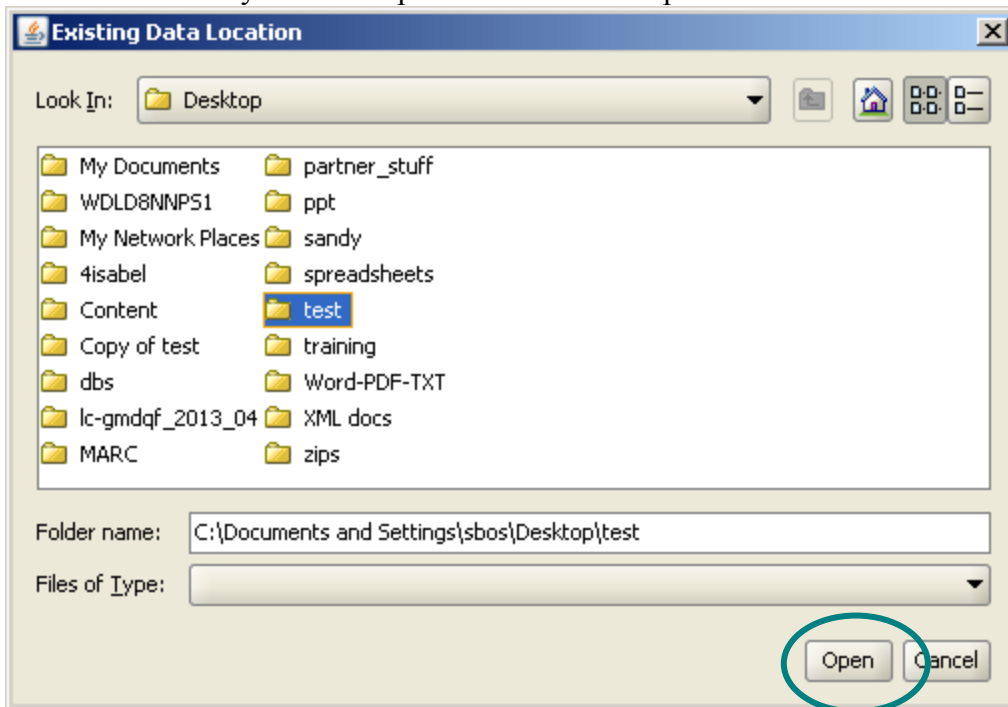

Bagger Guide

Bag In Place

- At the main bagger screen, choose "Create Bag In Place".
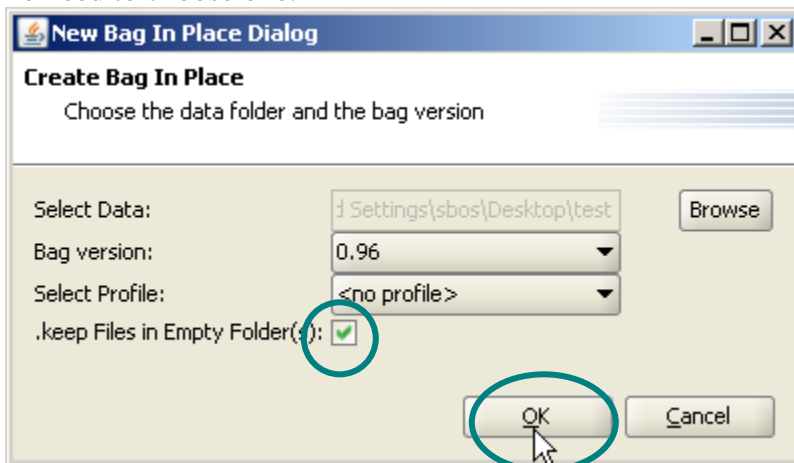

- In the pop-up "Create Bag in Place" dialog box, click "Browse". This will let you navigate to the directory you wish to bag.

- Choose the folder you wish to process and click "Open".

- If you have empty directories that you want to preserve, check ".keep Files in Empty Folder(s):" and then choose "OK". Profiles are not enable for "Bag in Place" so there's no need to choose one.

- If everything goes well, you should see a pop-up telling you your bag was saved. Click OK.

Validating

Validating a bag does three things.
1. It checks whether all the files in the manifest are still there.
2. It checks whether any files have been added to the bag since the manifest was created.
3. It checks the checksums of each file to determine whether the file itself has been modified.

Note: Some Windows machines create hidden thumbnail files for image directories. If the machine you are working on does this, you will need to turn off this functionality in order to run the validation process.
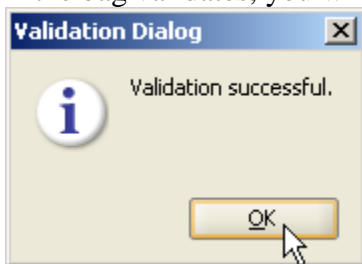
- Start the Bagger tool and choose "Open Existing Bag" from the icons across the top. Navigate to the previously bagged directory that you want to check.
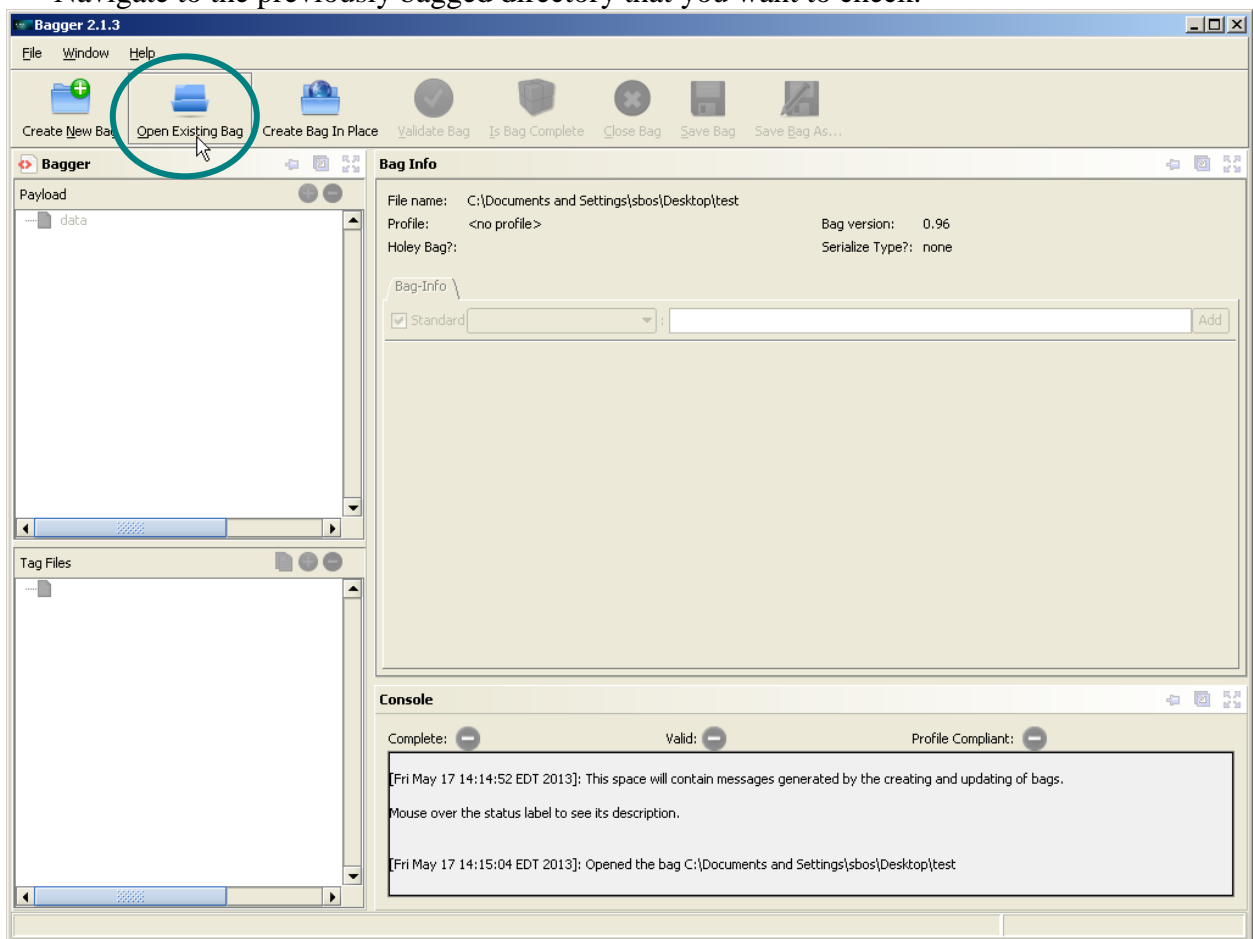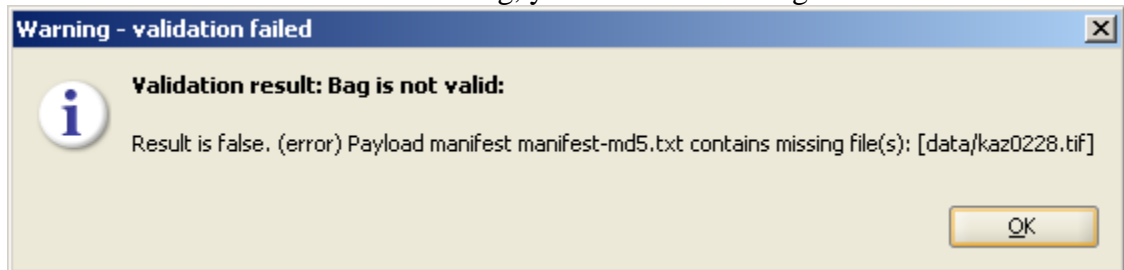

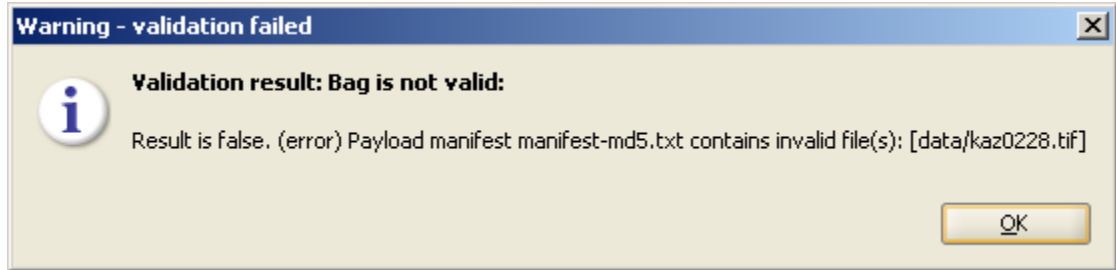- Choose "Validate Bag" from the from the icons across the top.
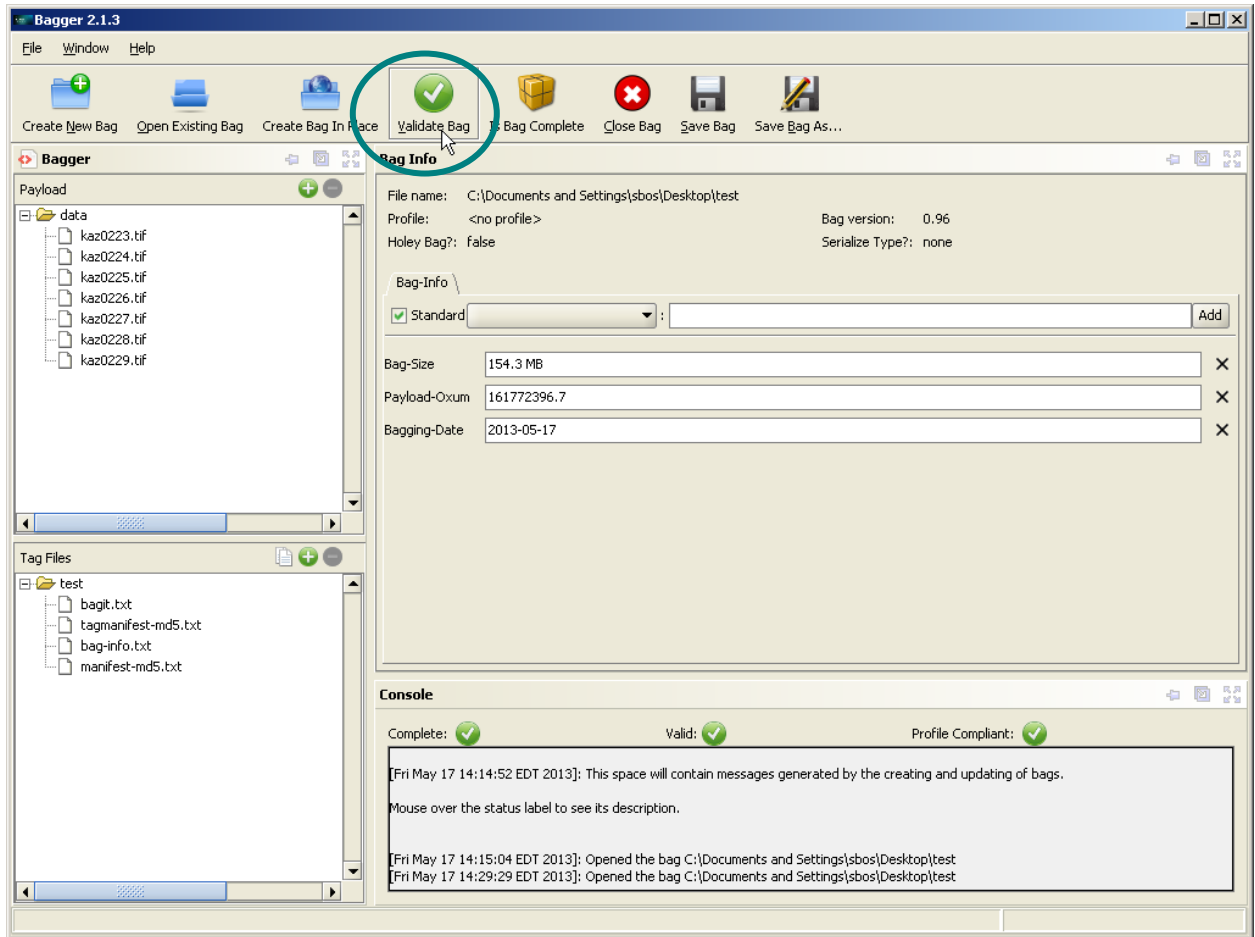
- If the bag validates, you will get a pop-up box telling you so. Choose "OK".
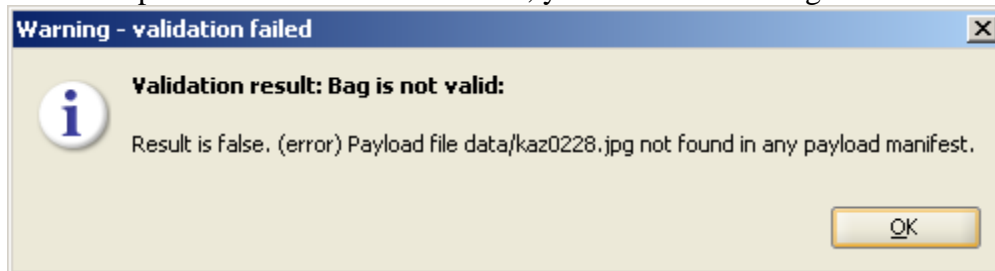

- If the bag does not validate (something has changed), you will get one of three messages:

  1. If a file listed in the manifest is missing, you will see a message like this:

2. If a file has changed since the manifest was generated, you will see a message like this:

> **Warning - validation failed**
>
> **ⓘ** **Validation result: Bag is not valid:**
>
> Result is false. (error) Payload manifest manifest-md5.txt contains invalid file(s): [data/kaz0228.tif]
>
> OK

3. If a file is present but not in the manifest, you will see a message like this:

> **Warning - validation failed**
>
> **ⓘ** **Validation result: Bag is not valid:**
>
> Result is false. (error) Payload file data/kaz0228.jpg not found in any payload manifest.
>
> OK

Appendix A: Installing & Running Bagger

Installing the Application

Bagger is an open source application. The Library of Congress' Repository Development Center maintains a SourceForge site where users can download the latest versions at http://sourceforge.net/projects/loc-xferutils/files/loc-bagger/ . It is the same download for both PC and Mac.

1) You will see two versions there. The executable needed for your installation contains the version number .zip while the source code (for developers only) contains _src.zip. Make sure you download the executable version.
2) After download, unzip the file.
3) You should see a directory named bagger-*version.number* (where *version.number* is something like 2.1.1). That's it.

Installing the WDL Profile

The World Digital Library profile is a customized template for the shipping information to be included in the bag. It is a JSON file. The file is the same for Macs and PCs but the placement of it is different.

PC

1) Locate the WDL profile file in the Bagger installation folder.
2) Place the file in the directory C:\Documents and Settings*<user>*\bagger . This is the default location where profiles are located.
3) Restart Bagger.

Mac

1) Locate the WDL profile file in the Bagger installation folder.
2) Place the file in the directory /Users/*<user>*/bagger . This is the default location where profiles are located.
3) Restart Bagger.

Running Bagger

PC

1) To start Bagger, double-click the bagger.bat file in the application directory.
2) You should see a windows command line screen with text scrolling across. This will remain open as long as Bagger is running.
3) You should also see a blue-green startup screen with Bagger across it. This will only stay open for the start-up process.
4) The main Bagger window should open.

Mac

1) To start Bagger, double-click the bagger-*version_number*.jar file in the application directory.
2) The main Bagger window should open.