BagIt

BagIt is a specification for packaging/storing content for preservation and transfer. For Chronopolis, we use the bagit specification as a base, only updating certain files to be required.

Structure

```
Bag Structure
<Collection-Name>/
        |   bagit.txt
        |   manifest-sha256.txt
        |   bag-info.txt
        |   tagmanifest-sha256.txt
        \--- data/
              |   [payload files]
```

Currently we require sha256 for Chronopolis, so all manifests are appended with "-sha256"

Data Files

The data files (the payload of a bag) in the BagIt spec are all the files found under the data directory. If there are orphaned files (those not found in the manifest, but exist in the data directory), a bag is deemed to be invalid.

Tag Files

Tag files are all those in that exist outside of the data directory. The bagit.txt and bag-info.txt files are standard and we follow the same requirements as in the BagIt specification for them.

manifest-sha256.txt

- Required as part of the bagit spec
- Contains the digest of a file with the relative path from the bags root directory
- Uses the same output as the md5sum/sha256sum command line utilities

tagmanifest-sha256.txt

- We require the existence of a tagmanifest, which contains the digest of each tag file and the manifest.
- This lets us validate not only the tag files but also the manifest we were given.
- When doing bagging ourselves, we create this only if each file in the manifest has been validated

bag-info.txt

- Required for UCSD deposits.
- Information listed should come from the UCSD depositor SLA.

```
Source Organization:
Organization Address:
Contact Name:
Contact Phone:
Contact Email:
```

- bag command will add

```
Payload-Oxum: 163450283.76
Bagging-Date: 2017-10-05
Bag-Size: 155.9 MB
```

Optional Tag Files

Duracloud Bridge:

- manifest-md5.txt
- content-properties.json
- .collection-snapshot.properties

Each optional file is digested and added to the tagmanifest. Validation on optional files is not currently supported.

Unaccepted Tag Files

- fetch.txt files are not accepted into Chronopolis (although legacy examples may exist)