

Packages sent ... and received?

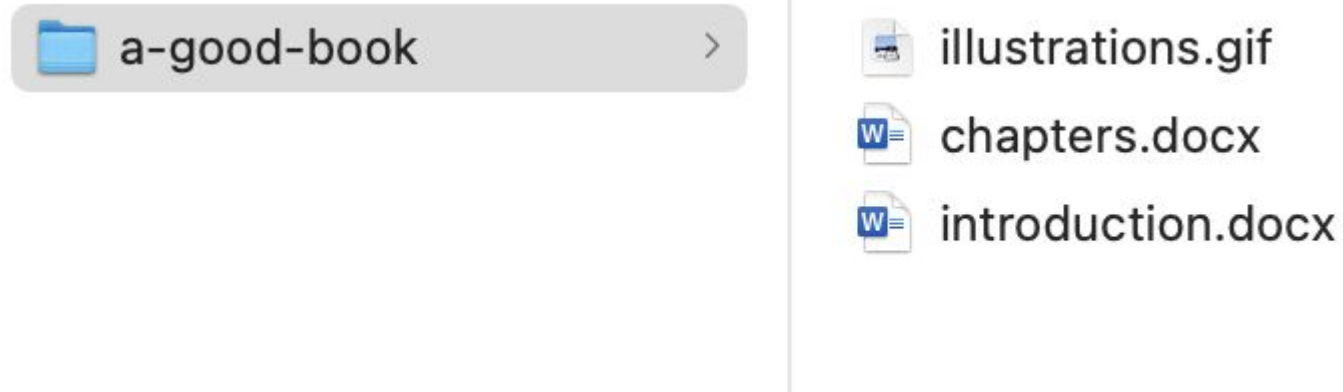
Documenting digital authenticity and creating preservation
metadata with BagIt

Providing Authentic, Verifiable Digital Content

- “Did you get what I sent you?”
- “Do you have the file ...
that I sent you last year?”
 - “Are you sure that it’s the same?”



It's more complicated with “complex” digital objects . . .



Think of the folder as an “information package”



1-index.html



2-scripts.js

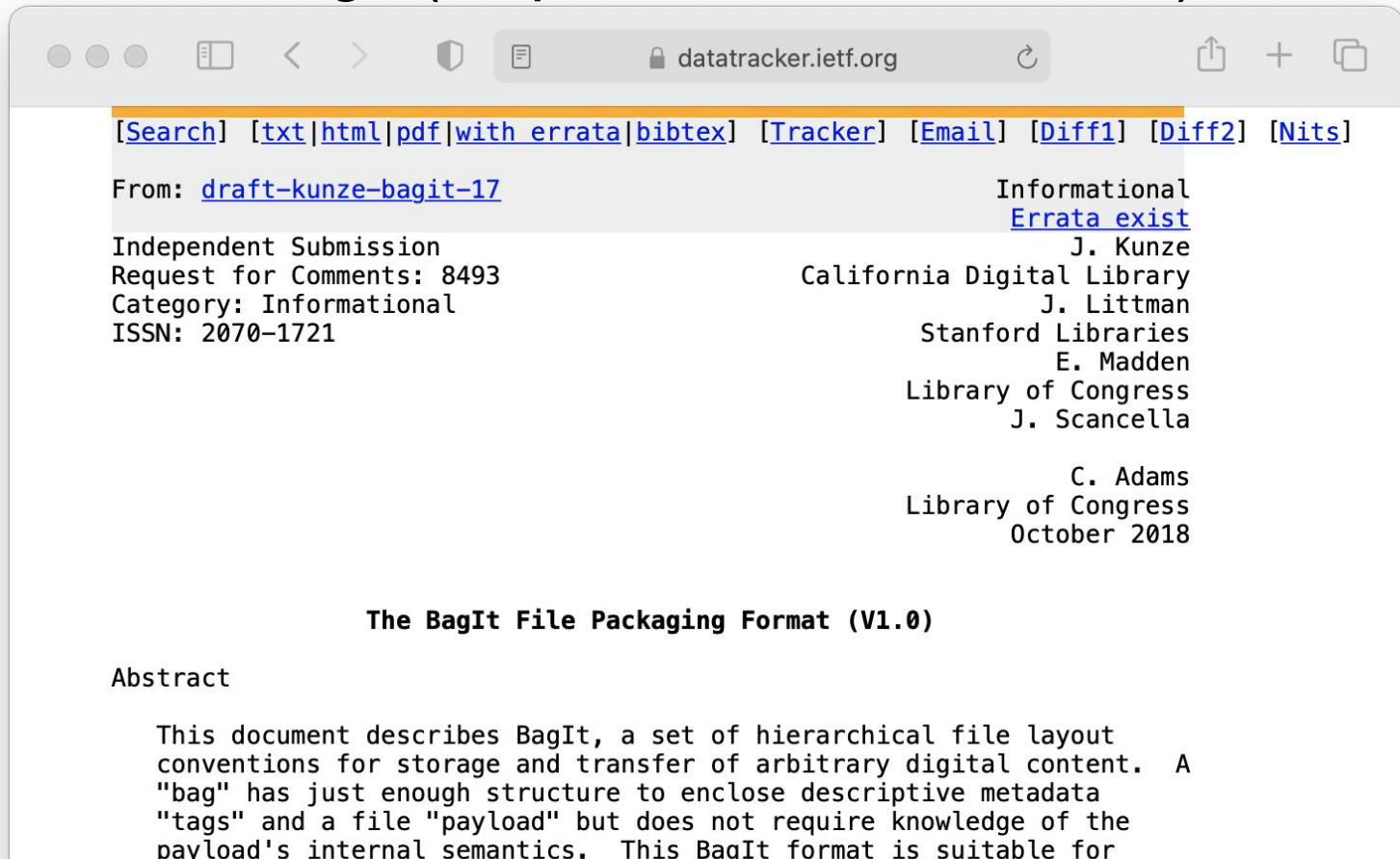


3-images.jpg



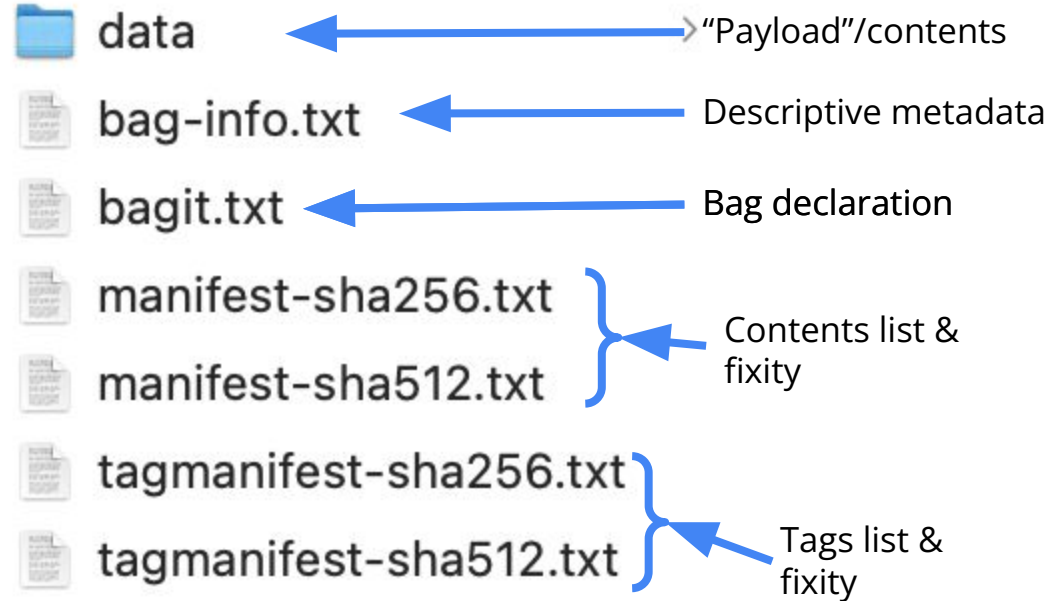
4-style.css

One Answer: BagIt (A specification, and a tool)



What does it look like? How is it structured?

 bag-folder >



Use “Bag Info” for descriptive information

Data is stored as label-name: value

(this is like a Python dictionary... ‘key’: ‘value’)

Sample BagInfo:

Contact-Name: Jesse

External-Description: These are some cool files!

Bagging-Date: 2022-01-08

Payload-Oxum: 26923687.23

Manifests & sample bag

myfirstbag/

manifest-md5.txt

(49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png)

(408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt)

bagit.txt

(BagIt-version: 1.0)

(Tag-File-Character-Encoding: UTF-8)

\--- data/

27613-h/images/q172.png

(... image bytes ...)

27613-h/images/q172.txt

(... OCR text ...)

Demo

Open up notebook 01b-using-bagit...

Use Cases

- Transfer and verify large amounts of heterogeneous web archive data (this was the problem in 2005, when California Digital Library and Library of Congress defined BagIt)
- Store large groups of digitized content, such as digitized microfilms (this is what the National Digital Newspaper Program does to preserve millions of pages of digitized newspapers)
- “Bag” local records, transmit them to a state or federal archive, and the archive can confirm they got the authentic files (this is used by the North Carolina State Archives)