Tackling Tangible Media: Projects, Workflows, and Issues

Moryma Aydelott
Library of Congress
Program Specialist, LS/TECH/O
07/2012

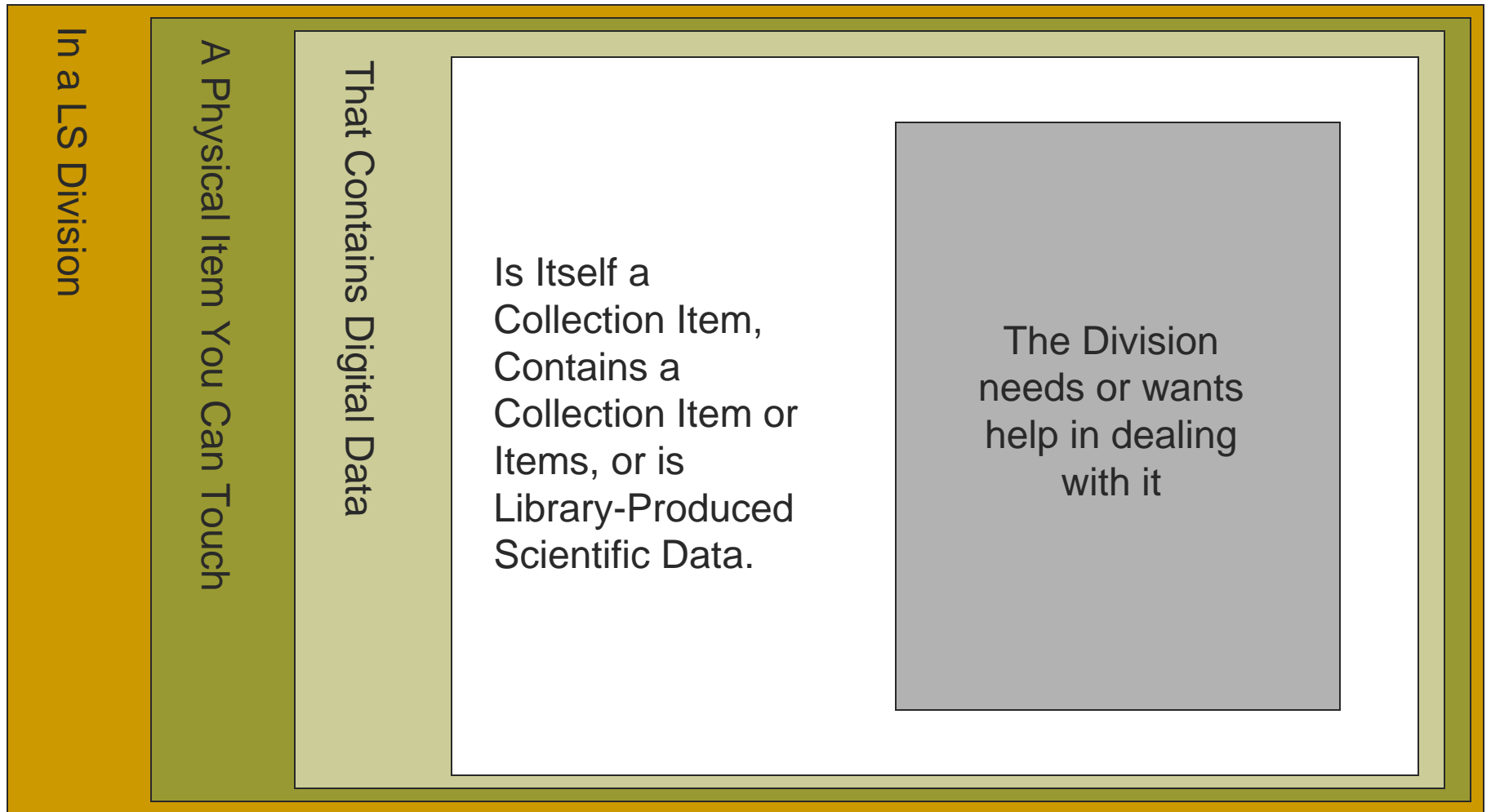*(Left)* Disks, in their Natural Habitat

(Right) A Whole Lot of CDs

The Problem

- From a 2009 Survey – approx 300 TB
  - DLT Tapes, DAT Tapes, Mini DVDs, Zip Discs, Flash Drives, Hard Drives, Floppy Discs (various sizes), DVDs, and CDs
- Being transferred, stored, and processed (or not) in different ways and using different tools by each of the Library of Congress' curatorial divisions.

The Projects

In a LS Division

A Physical Item You Can Touch

That Contains Digital Data

Is Itself a Collection Item, Contains a Collection Item or Items, or is Library-Produced Scientific Data.

The Division needs or wants help in dealing with it

The Goal

- A generic workflow for aggregation, bagging, transfer, inventory, access, verification and long term storage that can be used by multiple LS Divisions, various types of tangible media, and any digital formats.

- A centralized set of resources and help

The Plan

1. Talking things over with the Divisions
2. Triage – Saving and Bagging
3. Making the content available in the way, or at the level it is with the current item
4. Making content more available, as legal and appropriate
5. Migrating, standardizing, reformatting content to make it more accessible and stable over time

Talking with the Division

- About the materials
  - What are they?
  - How are they used now?
  - How would they like them to be used in the future?
- About the staff and the work
  - What do they already know about the process and tools?
  - How much do they want to learn?
  - How much time do staff have?
  - When do they want this done?

Answers to the above shape the generic workflow, at all points in the process.
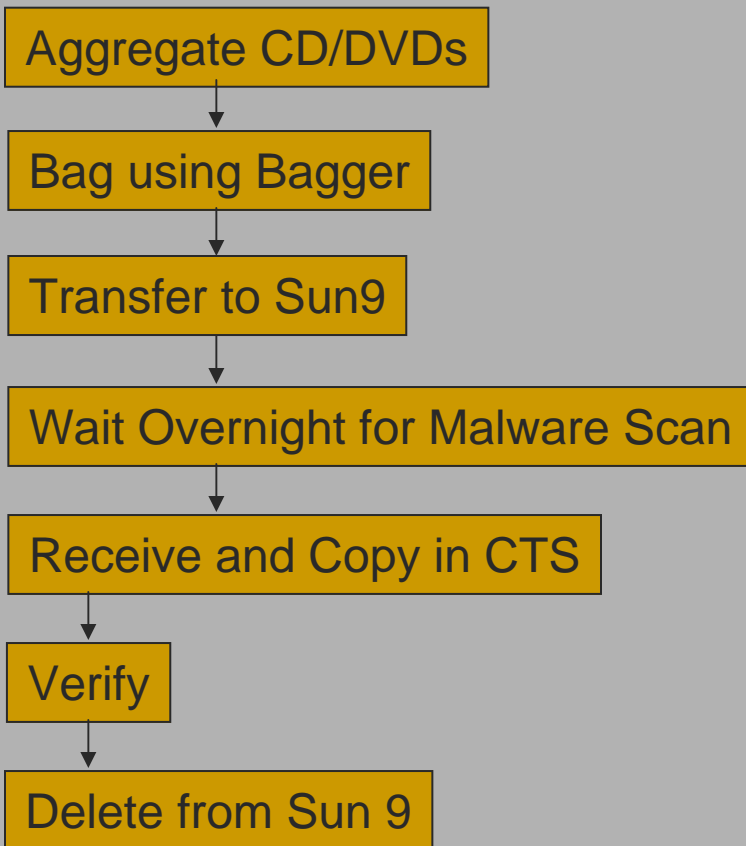
Triage – Saving and Bagging

Order varies on the content but consists of:

- Getting the Content onto a Server so that if something happens to the Items we have a backup copy

- Bagging the Content so that we have a manifest, the data to run checksums later, and some additional metadata.
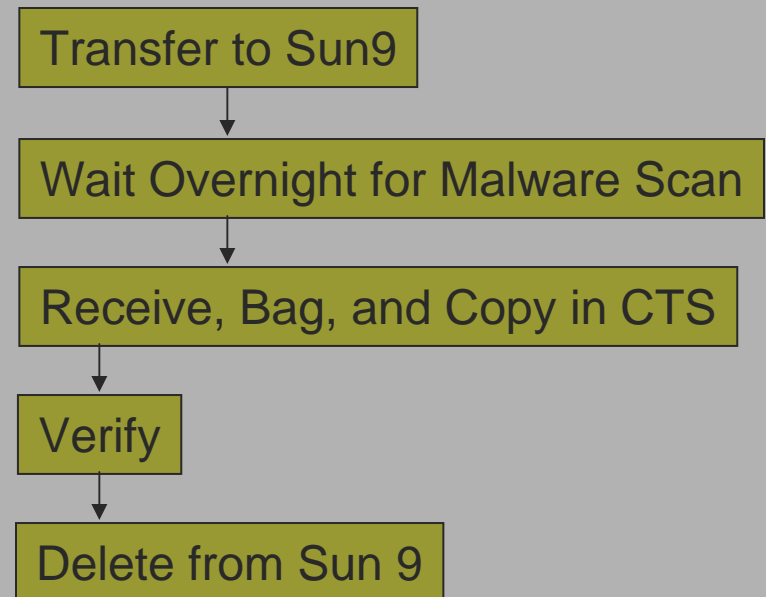
Triage Steps, by Media

(Assuming everyone has needed accounts and there's available space on servers)

For Tangible Media Items on CD/DVD

- Aggregate CD/DVDs
  ↓
- Bag using Bagger
  ↓
- Transfer to Sun9
  ↓
- Wait Overnight for Malware Scan
  ↓
- Receive and Copy in CTS
  ↓
- Verify
  ↓
- Delete from Sun 9

For Tangible Media Items on Hard Drives

- Transfer to Sun9
  ↓
- Wait Overnight for Malware Scan
  ↓
- Receive, Bag, and Copy in CTS
  ↓
- Verify
  ↓
- Delete from Sun 9

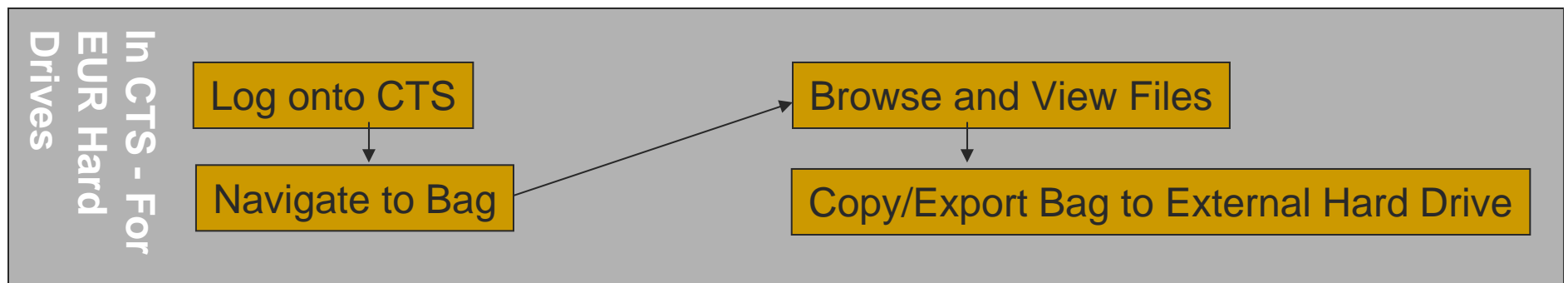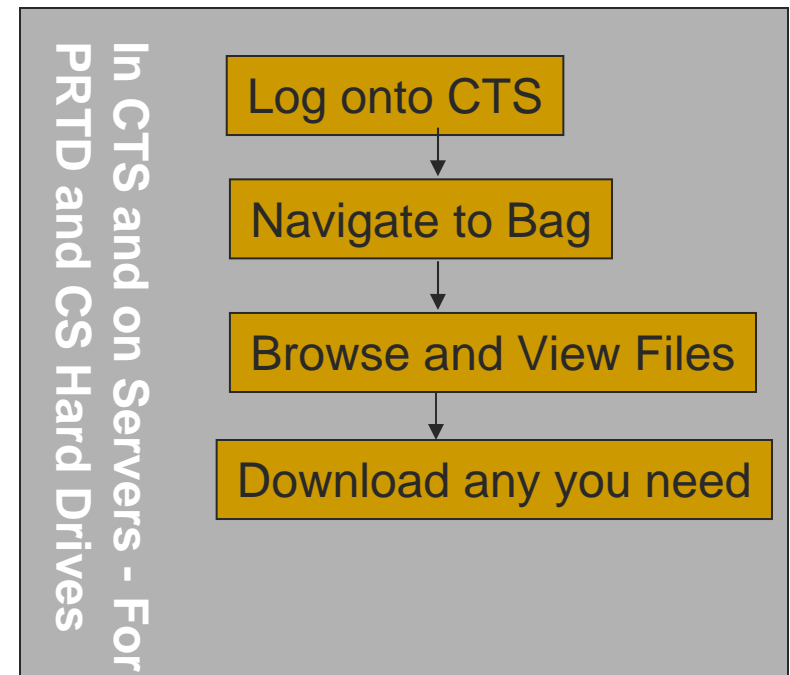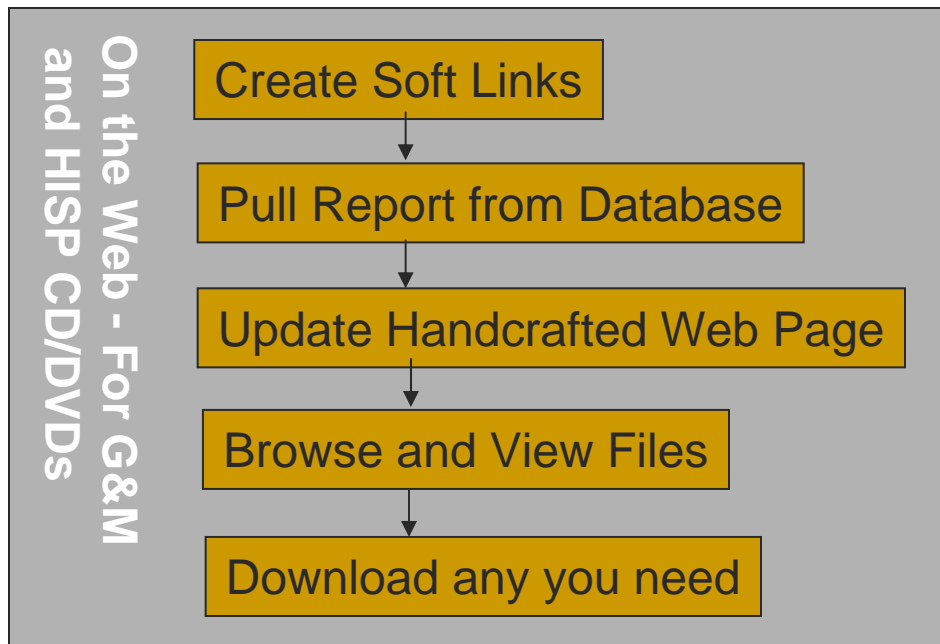*(Left)* The Cart, onsite in G&M

Making Content Available as it is Currently

Providing Access to Files

- Web Links

- Using CTS

- Via Filezilla/FTP

Making Available As Is – Steps, by Media and Permissions

(Assuming everyone has needed accounts and there's available space on servers)

On the Web - For G&M and HISP CD/DVDs

Create Soft Links
↓
Pull Report from Database
↓
Update Handcrafted Web Page
↓
Browse and View Files
↓
Download any you need

In CTS and on Servers - For PRTD and CS Hard Drives

Log onto CTS
↓
Navigate to Bag
↓
Browse and View Files
↓
Download any you need

In CTS - For EUR Hard Drives

Log onto CTS
↓
Navigate to Bag → Browse and View Files
↓
Copy/Export Bag to External Hard Drive

Actually running/viewing files

- Some Files can be viewed directly in the browser
  - Example - http://rs16.loc.gov/natlib/tangible/gmd/00198643774/DOCUMENT/DOQTEXT/DOQLIST.TXT
- Some Files be downloaded and run in applications we have
- In the process of acquiring software to mount ISO images – allowing us to use images of a CD or DVD in the same way we do the original
- As for the rest – all done on a case by case basis

As for the rest….

1. Talking things over with the Divisions
2. Triage – Saving and Bagging
3. Making the content available in the way, or at the level it is with the current item
4. Making content more available, as legal and appropriate
5. Migrating, standardizing, reformatting content to make it more accessible and stable over time

For those remaining two steps there's tons of work being done and to be done.

- Best Practices
- Policies
- Workflow Optimization
- New Tools
- Additional Training

What's Been Done so Far?

- In Tangible Projects in 5 Divisions:
  - Over 700 CD/DVD Titles
    - File and ISO Images of each
    - Individually Bagged
    - About 540 GB for the 1400 Items
  - 14 Hard Drives
    - Individually Bagged
    - About 6000 GB
- Documentation of Workflows and Processes
- Staff training in Digital Preservation work
- Use of Jukebox by other Library projects

Questions?

Want more info? Have a suggestion?
Care to share war stories?

Moryma Aydelott
Email: mayd@loc.gov
Phone: 202-707-1605