

# Nuts and Bits: Digital Processing 1.0

New England Archivists  
Fall 2010 Conference  
November 6, 2010



Jennifer Phillips  
Digital Collections Archivist

Veronica Martzahl  
Records Archivist

**Tufts**  
UNIVERSITY  
Digital Collections and Archives

## Session Overview

Surveying your holdings

Document and store

Metadata

Workflows and tips



November 6, 2010

## Session Overview

Surveying your holdings

Document and store

Metadata

Workflows and tips



November 6, 2010

---

## Surveying your holdings

Determine the  
size of the problem

On disk

On file servers

On obsolete data storage

Hidden in boxes

---

## Surveying your holdings

Determine the  
nature of the problem

Images

Documents

Spreadsheets

Audio

Video

Email

Web

Databases

## Surveying your holdings

Familiar file types you  
might encounter

Images: .jpg, .tif, .gif, .pdf

Documents: .doc, .txt, .pdf

Spreadsheets: .xls, .csv

Audio: .mp3, .wav, .aif

Video: .mov, .mpeg, .avi

Email: .eml

Web: .html, .css

Databases: .mdb, .sql

---

# Surveying your holdings

Formats you know nothing about

---

<http://www.digitalpreservation.gov/formats/>

---

## Surveying your holdings

Determine who cares  
about the problem

Who are your stakeholders?

Can their expectations be  
leveraged?

---

## Surveying your holdings

Determine who can  
help with the problem

Intra-institutional partners

Inter-institutional partners

Professional associations

---

## Surveying your holdings

Attack the problem

Pick a manageable sub-set  
of your materials

Begin description!

---

## Surveying your holdings

Communication and  
managing expectations

Legacy materials

Accessions

Turnaround times

## Session Overview

Surveying your holdings

Document and store

Metadata

Workflows and tips



November 6, 2010

---

## Document and store

Find a way to track  
what you know

Policies and procedures

Manuals

Spreadsheets

Collection management  
systems

Accession documentation

---

## Document and store

Medium

CD-ROM/DVD

Hard drive

Networked servers

Cloud storage

## Document and store



Multiple copies



Control, cost

Refresh data



Regular back-ups



---

# Document and store

## Case Scenario

Evolution of stabilization procedure  
at Tufts DCA

# Session Overview

Surveying your holdings

Document and store

Metadata

Workflows and tips

The screenshot displays the Tufts Digital Library interface. At the top, there are navigation links for 'Search', 'Browse', and 'Resource Inspector'. The main content area is divided into several sections:

- Thumbnail View:** A small image of a student working with survey equipment, with a link to 'Advanced Image Viewer'.
- Metadata:** A list of fields including Title, Date, Format, Places, and Topics.
- Complete Image Information:** A detailed view of the image's properties, including Title, Date, Publisher, Citable URL, Description, Format, Resolution, Bit Depth, and Color Space.
- Image:** A large view of the survey equipment being used by a student.

**Metadata:**  
Title: Students working with survey equipment, ca. 1930  
Date: 1930  
Format: image/tiff  
Places: [Medford Campus](#)  
Topics: [Academic](#)

**Complete Image Information:**  
Title: Students working with survey equipment, ca. 1930  
Date: 1930  
Publisher: Digital Collections and Archives, Tufts University  
Citable URL: <http://hdl.handle.net/10427/2666>  
Description: 6x8  
Format: image/tiff  
Resolution: 600  
Bit Depth: 8  
Color Space: grayscale

---

# Metadata

## Categories

Descriptive

Administrative

Technical

# Metadata

Question of emphasis

Descriptive: photos, publications

Administrative: institutional records

Technical: databases, course  
management packages

# Metadata

## Question of creation

Who is supplying metadata?

Where is it coming from?

Can metadata be garnered from external sources?

Can you trust embedded metadata?

---

## Metadata

Question of quantity

Risk of too little or too much

How much is enough?

---

## Metadata

Minimum elements:  
descriptive

DC title

DC creator

DC date

DC type

DC format

DC extent

DC identifier

DC subject

## Metadata

Minimum elements:  
descriptive

Title: Catnip

Creator: School of Veterinary  
Medicine

Date: 2006-04-01

Type: text

Format: application/pdf

Extent: 85876 bytes

Identifier: UP015.001.002.00014

## Metadata

Minimum elements:  
administrative

Original file name

Original file path

Accession date and person

Stabilization date and person

Software used to handle file

Rights

## Metadata

Original file name	catnip_spring06.pdf
Original file path	//vet_publications/2006
Accession date and person	2008-05-27, Veronica Martzahl
Stabilization date and person	2009-11-06, Veronica Martzahl
Software used to handle file	OfficeScan 10.5, Adobe Acrobat 9.0
Rights	Tufts University

---

## Metadata

Minimum elements:  
technical

Checksum

Checksum algorithm

Storage location

Storage medium

Dependencies

Capture

## Metadata

Checksum	3c0e49f9d8e33c3acff7c5b83b65029
Checksum algorithm	Md5
Storage location	C:\publictions\UP015\preservation
Storage medium	Hard disk
Dependencies	None known
Capture	N/A

---

# Metadata

## Specialized schema

Darwin Core

FGDC CSDGM

VRA Core

ETD-MS

PREMIS

---

# Metadata

## Case Scenario

Dark archive aggregate files

Individual university publications

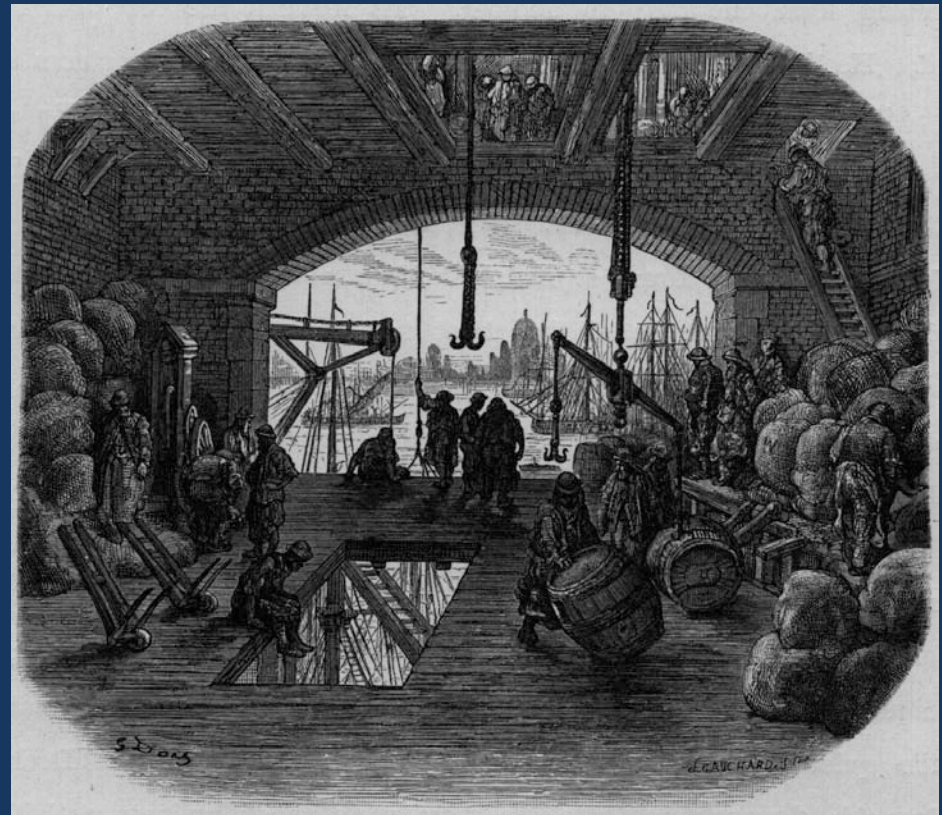
## Session Overview

Surveying your holdings

Document and store

Metadata

Workflows and tips



November 6, 2010

---

## Workflows and tips

### Developing workflows

Core archival principles

Proactive vs. reactive

Flexibility and revision

Unexpected consequences

## Workflows and tips

Tried to do everything

It didn't work . . . too many variables

- Files types
- Media
- Individual vs. aggregate cataloging

---

## Workflows and tips

Broke it down to what  
we knew

### Digital publications

- One file type
- No Media
- Individual cataloging

---

## Workflows and tips

Then we added a layer

Digital publications with print

- Just a cataloging problem
- Work vs. manifestation

## Workflows and tips

More complexity

Desktop publishing files

- Straight-forward preservation formats
- Transfer vs. preservation copies
- Individual vs. aggregate cataloging

## Workflows and tips

After that . . .

Just a matter of different file types

- Access copy considerations
- Storage space
- Preservation standards still being developed

## Workflows and tips

### Tips

Don't convert lossy formats into bulky files just to have a "preservation copy"

- NO .jpg to .tif
- NO .pdf to .pdf/a

## Workflows and tips

### Tips

Go back to what you know

- Appraisal and selection
- Aggregate description

---

## Workflows and tips

### Tips

Find partners and resources

- Share software and hardware

[http://digital-preservation-lab.wikia.com/wiki/  
Archival\\_Digital\\_Preservation\\_Lab\\_Wiki](http://digital-preservation-lab.wikia.com/wiki/Archival_Digital_Preservation_Lab_Wiki)

---

# Workflows and tips

Tips

START!

---

## Workflows and tips

Thank You!

Veronica Martzahl, Records Archivist  
[veronica.martzahl@tufts.edu](mailto:veronica.martzahl@tufts.edu)

Jennifer Phillips, Digital Collections  
Archivist  
[jennifer.phillips@tufts.edu](mailto:jennifer.phillips@tufts.edu)