

# Managing and Preserving Digital Files

---

A workshop developed by  
Moriah Caruso and Ann Lally  
University of Washington Libraries  
For We Are History Keepers  
February 27, 2021

# Welcome!

---

Who are we?

- > **Moriah Caruso**
- > **Digital Preservation Librarian**
- > [neilsm@uw.edu](mailto:neilsm@uw.edu)
  
- > **Ann Lally**
- > **Digital Collections Curator**
- > [alally@uw.edu](mailto:alally@uw.edu)

# How to preserve and manage digital files

---

2021-02-27: We are history keepers workshop

- > **10:15-10:30** Getting to know each other
- > **10:30-10:50** UW Case Study
- > **10:50-11** Break
- > **11-11:35** Digital pitfalls and best practices
- > **11:35-11:50** Q&A
- > **11:50-12** Wrap-Up

# **Welcome!**

---

Who are you???

- > Feel free to introduce yourselves in the chat and get ready to answer some polls!**

# Case Study - James McDermott Papers

---



- > U.S House of Representatives 1989 - 2017 for Washington's 7th District
- > Mixed materials collection - documents, photographs and digital
- > First digital materials appear ~1995
- > Original transfer was 80 GB



# Technical Processing

---

- > Preponderance of PDFs downloaded from the internet that were useless
- > Photographs named DCS1893042, DSC129847 DSC20957 organized by event
- > Long file names coming from default names (like entire first sentence of a document) Nested deeply, long path.



# Long file paths

---

James McDermott/2005/ Correspondence/Dear Colleagues/March 2015/to Approps Committee Transportation HUD Subcommittee Leaders re funding for McKinney-Vento Homeless Assistance Grants & HEARTH Act.pdf

> 198 characters with spaces



# **Adding local naming practice**

---

**E:/Shared Drive/20181206\_James McDermott  
Transfer/Initial Copy/James McDermott/2005/  
Correspondence/Dear Colleagues/March 2015/to  
Approps Committee Transportation HUD  
Subcommittee Leaders re funding for  
McKinney-Vento Homeless Assistance Grants &  
HEARTH Act.pdf**

**> 261 characters with spaces**





## **Review by McDermott staff**

---

- > Two staff from Jim McDermott's Seattle office were paid to organize the collection including the digital files**
- > Removed everything that was not an original document including those downloaded from the internet (research reports, white papers, etc)**
- > Reorganized the digital files to be the structure as the paper**



# Finding aid

## Table of Contents

[Overview of the Collection](#)

[Biographical Note](#)

[Content Description](#)

▶ [Use of the Collection](#)

▶ [Administrative Information](#)

▼ [Detailed Description of the Collection](#)

[Accession No. 4074-001: James McDermott papers, 1972-1987](#)

[Accession No. 4074-002: James McDermott papers, 1940-2016](#)

[Names and Subjects](#)

### Accession No. 4074-002: James McDermott papers, 1940-2016

[↑Return to Top](#)

approximately 72.25 cubic feet (88 boxes )

SCOPE AND CONTENT: This collection contains Congressman Jim McDermott's congressional papers and other historical materials related to his service in the U.S. House of Representatives from 1989-2016, such as writings, photographs, awards, etc. Includes media and photographs

ARRANGEMENT: Arranged in 16 series.

- Series 1, Biographical Materials
- Series 2, United States Congressional Campaign
- Series 3, Committee Files
- Series 4, Communications
- Series 5, Newsclippings
- Series 6, Speeches
- Series 7, Legal and Ethical Cases
- Series 8, Travel
- Series 9, Meeting Logs
- Series 10, Audiovisual Materials
- Series 11, Photographs
- Series 12, Issues
- Series 13, Schedules
- Series 14, Awards and Honors
- Series 15, Ephemera
- Series 16, Artifacts

RESTRICTIONS ON ACCESS: No restrictions on access.

RESTRICTIONS ON USE: Creator's copyrights transferred to the University of Washington Libraries Special Collections.

ACQUISITION INFO: Donation from Jim McDermott's Office



# Finding aid 2

## Table of Contents

[Overview of the Collection](#)

[Biographical Note](#)

[Content Description](#)

▶ [Use of the Collection](#)

▶ [Administrative Information](#)

▼ [Detailed Description of the Collection](#)

[Accession No. 4074-001: James McDermott papers, 1972-1987](#)

[Accession No. 4074-002: James McDermott papers, 1940-2016](#)

[Names and Subjects](#)

Container(s)		Description	Dates
Series 1: Biographical Materials			
Box/Folder	Accession		
21/1	4074-002	New Democratic Members, 101st Congress	1988
65/2	4074-002	McDermott Congressional Memberships	1989-1990
22/15	4074-002	McDermott Biography and areas of interest	1989-1993
22/14	4074-002	Staff Biographies and issues areas	1989-2002
22/16	4074-002	Redistricting and complications	2002
65/1	4074-002	McDermott Committee/Sub-Committee Memberships	1989-2014
65/3	4074-002	McDermott Legislative Highlights	1989-2015
Electronic_file			
Reading Room Virtual Vault	4074-002	McDermott Medical Biography	2013
Reading Room Virtual Vault	4074-002	McDermott Multi-Use Biography	2013



# Final steps

---

- > Virtual Vault + locally stored backup copy
- > Send to Preservation
  - Using Bagger, package everything for transport and create a list of checksums
- > Preservation receives the package, validates it, and puts it to bed (details to follow)
- > We receive a Universally Unique Identification Number (UUID) and METS file from Preservation that is added to our internal accession database so we can recall the files if need be



# **General practices**

---

**This collection represents the most processing that has been done with digital files that are not exclusively photographs.**

**Personally Identifiable Information - PII**

**Online access**



# Break Time

---

Please enter any lingering questions in chat!

# Digital Preservation

---

*“Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change.*

*The goal of digital preservation is the accurate rendering of authenticated content over time.”*

<http://www.ala.org/alcts/resources/preserv/defdigpres0408>



# Why is digital different???

---

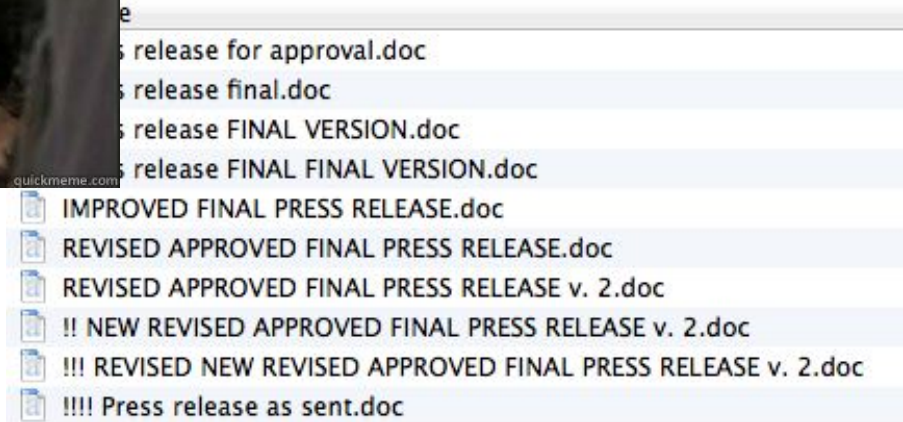
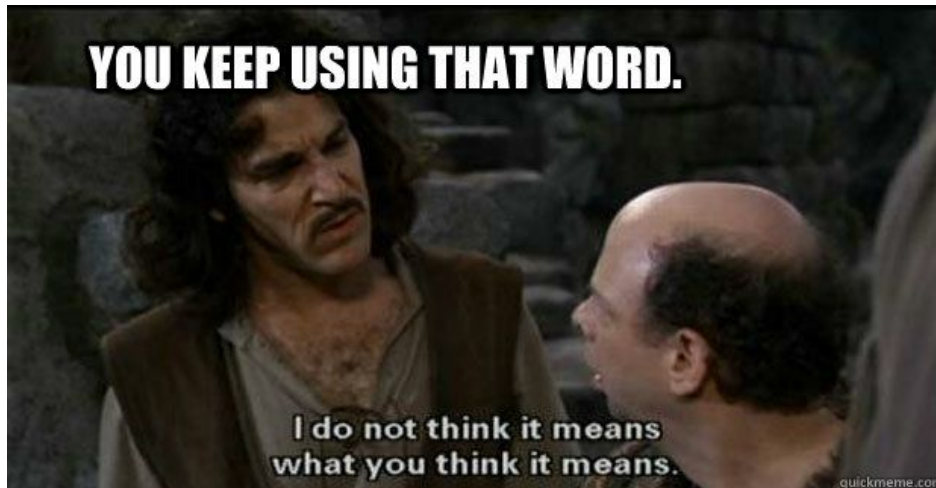
> Ease of production





# Why is digital different???

## > Ease of change/alteration



# Why is digital different???

## > Software and hardware obsolescence



*"L'evoluzione della specie" by Aldo Cavini e Benedetti  
on Flickr*

# Why is digital different???

## > \$\$\$ Licensing vs ownership (and ©)



**Terms and Conditions**

Agreement between User and \_\_\_\_\_  
Welcome to \_\_\_\_\_. The \_\_\_\_\_ website (the "Site") is comprised of various web pages operated by \_\_\_\_\_. \_\_\_\_\_ is offered to you conditioned on your acceptance without modification of the terms, conditions, and notices contained herein (the "Terms"). Your use of \_\_\_\_\_ constitutes your agreement to such Terms. Please read these terms carefully, and keep a copy of them for your reference.

\_\_\_\_\_ is a Blog Site.

\_\_\_\_\_ Privacy Policy. Please read our Privacy Policy, which also governs the Site and informs users of our data collection practices.

**Electronic Communications**  
Visiting \_\_\_\_\_ or sending emails to \_\_\_\_\_ constitutes electronic communications. You consent to receive electronic communications and you agree that all agreements, notices, disclosures and other communications that we provide to you electronically, via email and on the Site, satisfy any legal requirement that such communications be in writing.

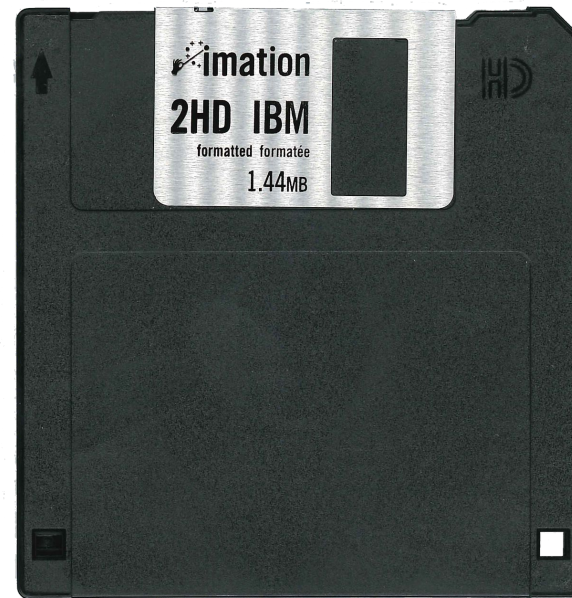
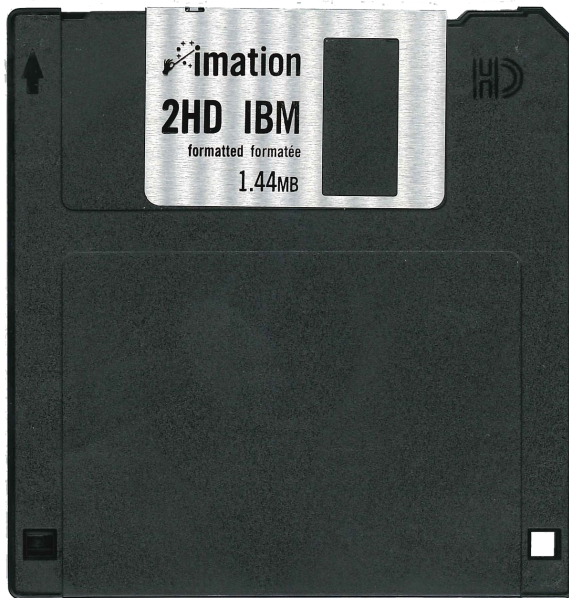
**Your Account**



# Why is digital different???

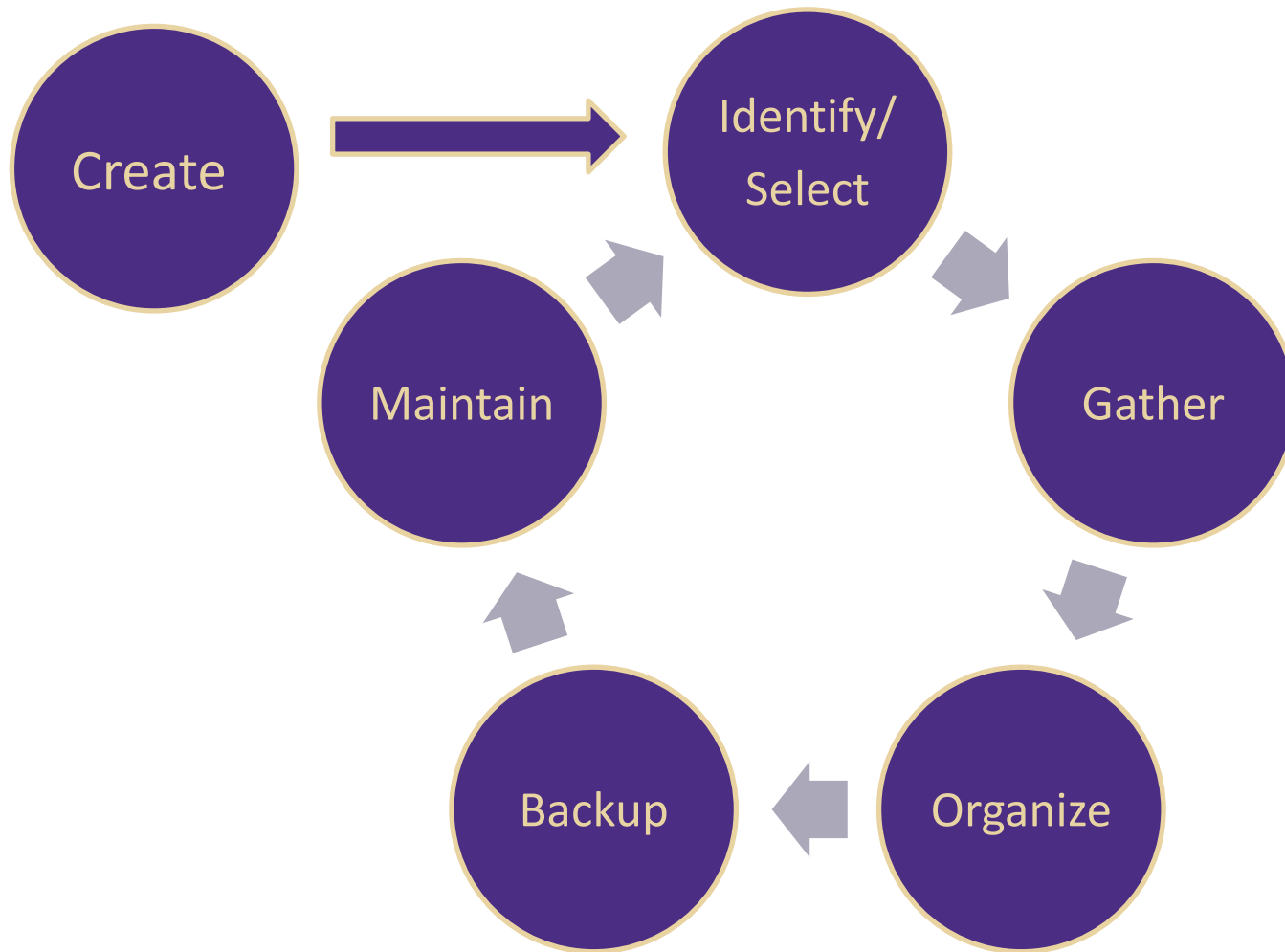
---

## > Inherent vice



# Digital Archiving Lifecycle

---



# Create

---

Creating your own files gives a unique opportunity to make them last!

## Choose widely adopted file formats

- **Text-based: PDF**
- **Still images: TIFF, JPEG**
- **Moving images: MOV, MP4**
- **Sounds: WAVE, MP3**
- **Tabular data: CSV**

# Identify and Select

---

A few questions to get started...

- > Which digital files represent irreplaceable memories, important life events, or valuable creative work?
- > Which digital objects tell meaningful stories about your life?
- > What digital objects do you expect or need to have access to in 5 years?

# **BUT CAN'T I JUST SAVE EVERYTHING???**

---

Short answer is no. Even the professionals can't.

## **> Twitter archive at the Library of Congress**

- 2006-2017 all public tweets
- 2017+ : selective policy
- <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/>

## **> Internet Archive Wayback Machine**

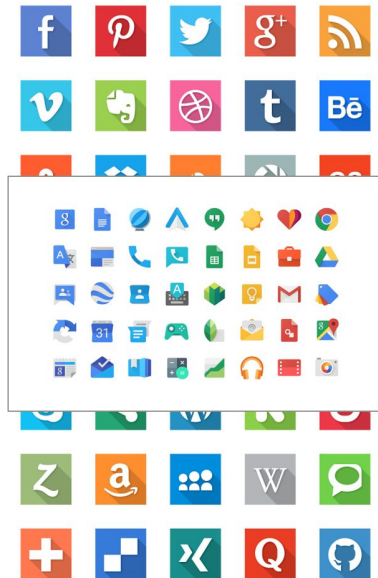
- 525 billion web pages and counting
- The average life of a web page is about 100 days
- New Yorker Article "Can the Internet be archived?"  
<https://www.newyorker.com/magazine/2015/01/26/cobweb>



# Gather

---

Think about where you may have stored or used important files.



Computers and phones

Other electronic devices

music players

digital cameras

digital audio/video recorders

External hard drives and disks

Email

Cloud services and storage (Google, Box, etc)

Personal websites/blogs

Social media and file hosting services

(Facebook, Twitter, etc)

# Exporting content from social media

---

Google:

<https://takeout.google.com/settings/takeout>

Twitter:

<https://help.twitter.com/en/managing-your-account/how-to-download-your-twitter-archive>

Facebook:

<https://www.facebook.com/help/1701730696756992>

# Organize

---

Now that you have some files together, what's next?

- > Goal: Organize your files into folders so they'll be easy to search or browse.
- > Use folder and file naming conventions
  - Keep names short (under 25 characters)
  - Avoid special characters and spaces ()^ ~ \ / : \* ? " < > | ! # % & \$ , .
  - Start with a date for easy sorting (2021-02-27)
- > **Write down your decisions so you remember next year!**

# Backup

---

## The 3-2-1 rule



# Cloud Storage

---

When you store your data in “the cloud” it is actually stored on computers owned by the service provider.

**Before deciding on a vendor, ask:**

- > **Do I have a reliable internet connection?**
- > **How much will it cost to store my data?**
- > **What technical skills do I need to get started?**
- > **How can I get my content out again?**

# Maintain

---

Digital preservation is never finished, only in progress.

- > **Check your storage annually**
- > **Refresh storage every 5-7 years**
- > **Spot check for obsolescence annually**
- > **Use a password manager**
- > **Review your list of important files**



# **Thank you!**

---

Questions?

**Moriah Caruso**

**Digital Preservation Librarian**

**[neilsm@uw.edu](mailto:neilsm@uw.edu)**

**Ann Lally**

**Digital Collections Curator**

**[alally@uw.edu](mailto:alally@uw.edu)**