# Digital POWRR

## Preserving Digital Objects with Restricted Resources

Meg Miner, University Archivist,
Illinois Wesleyan University

National Archives Conference for Fraternities and Sororities,
June 13, 2014

# Who we are….and how we got here….

- Defining Moments → Found Some Friends

- Applied for an Implementation Grant → Received a "Figure It Out" Grant

## The Digital POWRR Team: Proud to be works-in-progress

**ILLINOIS STATE UNIVERSITY**
*Illinois' first public university*

**ILLINOIS WESLEYAN UNIVERSITY**

**NIU**
**Northern Illinois University**

**CHICAGO STATE UNIVERSITY**

**WESTERN ILLINOIS UNIVERSITY**

Sponsored By: **INSTITUTE of Museum and Library SERVICES**

# Assumptions in the Digital Age

Storage is cheap and everywhere!
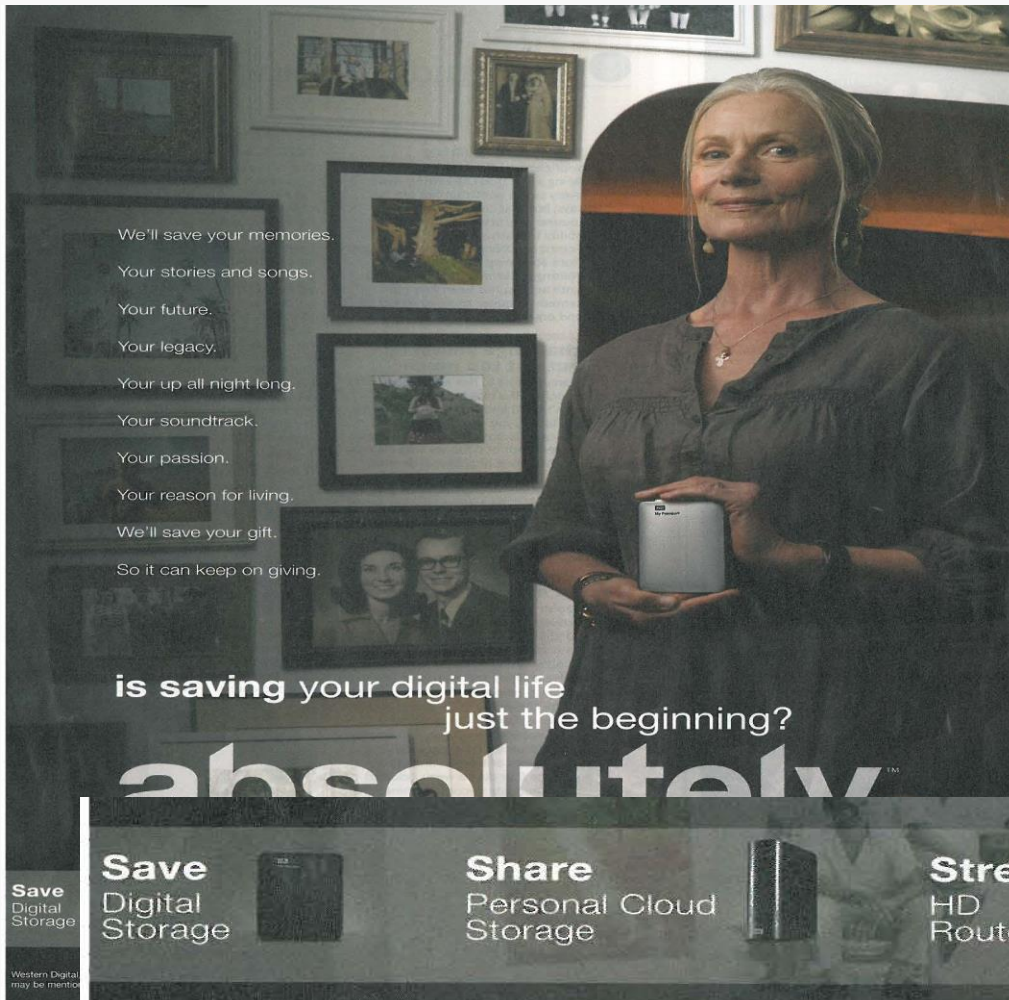*Is the content secure/accessible? How? By whom?*

If it is digital it will always be available!
*Will it be useable? If software changes or files decay how will we know what is gone before it is too late?*

# Storage

- Storage options are cheap and everywhere!

Storage
≠
Preservation

# Storage (cont'd)

- Storage options are cheap and everywhere!
  - But human intervention is not  (and we are so busy that it is easy to put off)
    - Files need naming, tagging, organizing
      AKA *metadata*
  - Do your backed up versions all match?
    - Do you spend time synching devices? How much?

# Storage (cont'd)

Don't forget: all drives fail eventually



"Do you have a back–up plan?"
by John K
http://www.flickr.com/photos/johnkay/5200871042/

# Storage (cont'd)

Media needs to be replaced every five years (aka migration)

"Bit Rot" by The Joy of the Mundane

# Storage (cont'd)

- Storage options are cheap and everywhere!
  - But human intervention is not (and we are so busy that it is easy to put off)
  - Do your backed up versions all match?
  - Remember: you get what you pay for with physical storage and in the Cloud!

# Storage (cont'd)

Backups on cheap and/or fragile media

"Data deterioration" by Univ.of VA Library, Preservation Services



http://www.flickr.com/photos/uvalibpreservation/8033085992/
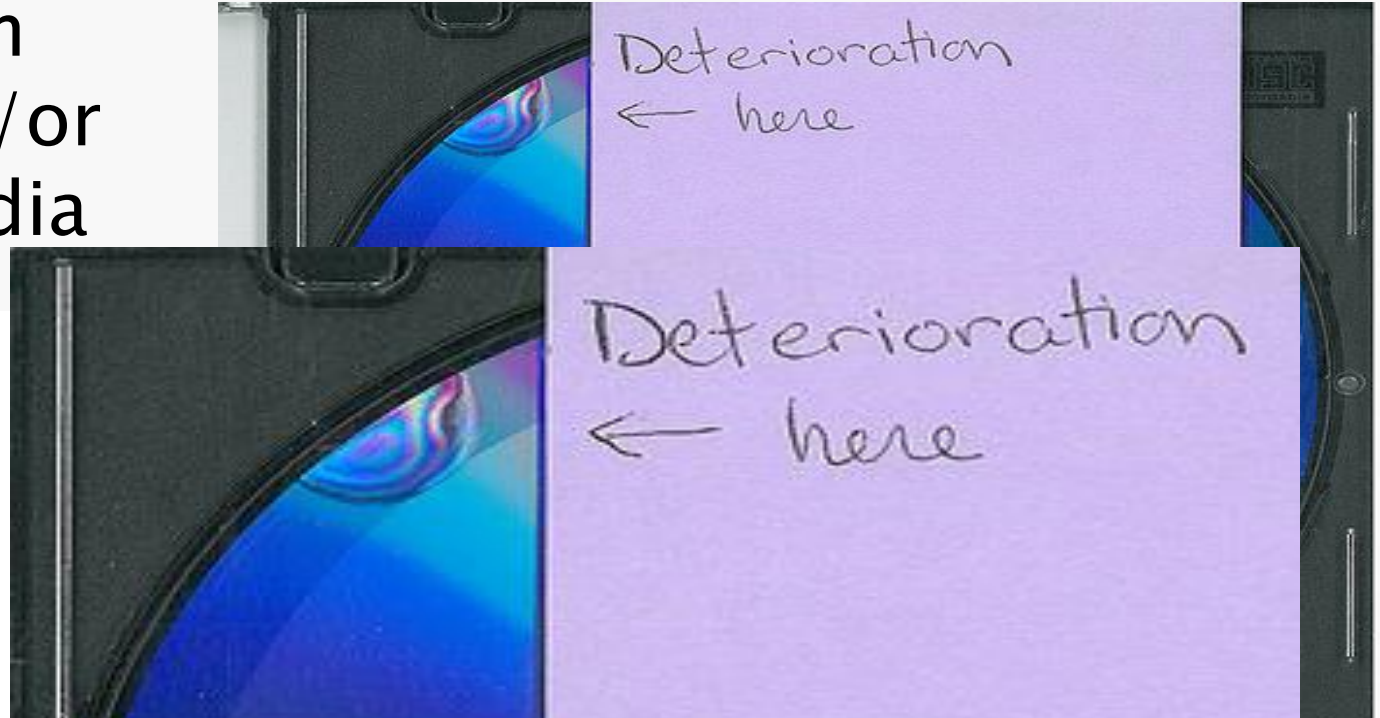
# Storage (cont'd)

Cloud providers
*may not* = stable, consistent access

- What security measures do they promise?
- Do they back up anywhere? How will you know?
- Do they offer an exit strategy if they go out of business?
- If the person who set up the account dies, who has access?

# Served ≠ preserved

# Clarification: Preservation vs. Access

## Long term access (Preservation)

- **Purpose**: ensure long-term access

- **Focus:** current & **future** users

- Relies on **proven (reliable)** technologies to preserve digital objects across generations of technology

- **Accumulate**s metadata over the life cycle to trace preserved content

- Preservation systems **create** new versions of digital objects for access to deliver as needs change over time

## Short term access

- **Purpose**: provide content to users now

- **Focus:** current

- Relies on **cutting edge** technologies to provide best and fastest access at a point in time

- **Selects** metadata needed to use and understand content

- Access systems **deliver** objects with user-oriented services

# Two steps towards preservation

Making back up copies of digital objects
vs.

Making back up copies of digital objects **and storing them in different places**

- The minimum recommendation: 2 copies

- Geographic distance = more protection
  - A Web site and networked servers that are backed up to another city.
  - Specialized storage sites (details to come!)

# Full digital preservation

**Decide** which digital objects need to be saved, **describe** them,
backing up (**duplicate**) in geographically **distant** places, *and* **monitor** the digital files for deterioration (the *bit rot* factor—more to come!).

*4Ds and 1M*
**Decision making + Description + Duplication + Distance + Monitoring = Digital Preservation**

Files become corrupted or degrade (aka, bit rot)

# Digital inherent vices

- Bit Rot
- Proprietary formats
- Software
- Hardware dependencies
- Practices of people who create these objects!

# Solution in Practice
## AKA Good Enough DP for real people!!

**Our take on what you need to consider when thinking about your digital stuff…..**
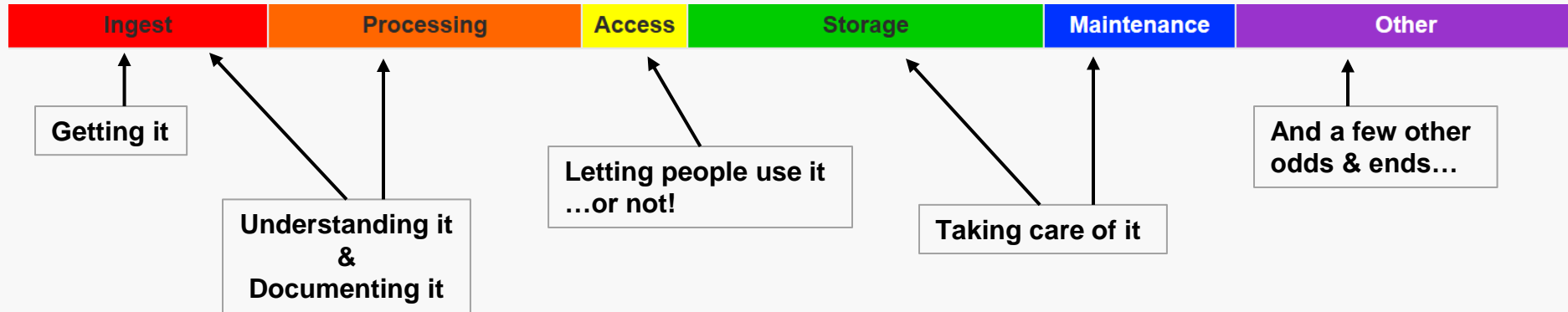
| Ingest | Processing | Access | Storage | Maintenance | Other |
|--------|-----------|--------|---------|-------------|-------|

**Getting it**

**Understanding it
&
Documenting it**

**Letting people use it
…or not!**

**Taking care of it**

**And a few other odds & ends…**

| Ingest | | | | | Processing | | | | | | Access | | Storage | | | | | Maintenance | | | Other | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Copy | Fixity Check | Virus Scan | File Dedupe | Auto Unique ID | Auto Metadata Creation | Auto Metadata Harvest | Manual Metadata | Rights Management | Package Metadata | Auto SIP Creation | Public Interface | Auto DIP Creation | Auto AIP Creation | Reliable, Long-Term Bit Preservation | Redundancy | Geographically Dispersed Data Storage Model | Exit Strategy | Migration | Monitoring | Auto Recovery | Open Source | Clear Documentation | Cost |

Our take on some things that need to happen or be considered along the way to this *"Digital Preservation"* thing….

The live site is at http://digitalpowrr.niu.edu/tool-grid

# Activity Time!
## 5 Minutes

1) What content do you collect that increasingly only comes in digital form?

    -- Example: Pictures, news items, etc.

2) How are you receiving it?

    -- Example: Email, flashdrive, website?

3) What problems are you encountering that are new to you?

# Format examples

Word documents or other text-based material: we may decide we only need to print the documents and store as we used to.

A/V content: we need more information. Was the object created in a format you can monitor for obsolescence?

# Saving originals is good practice but…

Consider "normalizing" content, *too*!

- Copy HTML to Word or ODF, PDF or PDF/A
- Transfer audio/visual content from CD/DVD and save files locally with an open multimedia player like VLC http://videolan.org

Check out tips for media and social media at http://digitalpowrr.niu.edu/digital-preservation-101/personal-preservation/

# Format bottom-line

- **Make a decision**: what you think you can care for informs what you should accept.

- **Start educating**: content creators/donors need to know about the formats you accept and your users need to know what you are able to provide.

*Add to Policies: Coll.Dev. and Preservation!*

# Transfer questions

Will you collect at individual file increments?

　Or create some kind of schedule for transferring specific content in bulk at periodic intervals?

If in bulk, will you know how/if files relate to each other if there's no donor face-to-face?

*Create a transfer template for donors that matches your accessions info needs!*

# Transfer template

What information do you collect for analog content? Get it for e-content! Plus…

Ask questions about activities and people present for A/V. Multimedia content without metadata has limited long-term value!

And just like analog, you will need to ask the donor to declare any issues with privacy, security or copyright restrictions.

# Pre-Ingest Inventory Spreadsheet Categories

–These suggestions follow the recommended DPOE step "Identify" as locally defined by curator/archivist. Example at:

http://www.carli.illinois.edu/sites/files/digital_collections/documentation/digipres_identify.pdf

This is YOUR inventory… YOU get to decide if it needs additional fields, if some can be deleted, etc. You are the boss of this!

❑ Category (digitization project; born digital; university archives)
❑ Title and Description
❑ Date(s) (date range of what's IN there or date of creation if born digital)
❑ Location (CD, Jump drive, server location?)
❑ Extent (quantity: 48 journal issues; 106 images; 2 TB of video)
❑ Format (file formats: PDF, .Jpeg, Animated GIF, Wordstar2.0 file)

| Category | Title and Description | Date | Location | Extent | Format |
|----------|----------------------|------|----------|--------|--------|
|          |                      |      |          |        |        |

**FILL OUT WHAT YOU CAN AS YOU WOULD WITH ANY NORMAL ACCESSION**

LIBRARY OF CONGRESS

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Category** | **Title & Description** | **Date** | **Location** | **Extent** | **Format** |
| 2 | (locally defined: project name? content creation method?) | (donor applied and/or yours…what's your practice?) | (YYYYMMDD or other locally defined format for accession date) | (storage place of choice--networked server recommended) | (quantity of folders, files by type or total size) | what extensions are involved: .jpg, .tif, .xls?) |
| 3 | Special Collections, mixed: digitized and born digital | A Curator's Cat Collection. Donated by Jane (nee Pennypincher) and John Moneybags, Class of 2006. Feline Health research. No restrictions on access; some material may have copyright restrictions by law. | 20140410 | C:\Users\User\Desktop\NewAccessions\Masters | 42MB in four folders: Classic Kitties, Kitty Research, Kitty Videos, Stacey's Kitties | 23 octet/stream, 10 pdf, 2 ppt, 24 .jpg |
| 4 | | | | | | |
| 5 | | | | | | |

# National Digital Stewardship Alliance

- Definition of "basic" fixity: expected file size and file count on information capture.

  Example: Does the detailed view in a file window show a size of 0 bytes? Does your donation form (or the donor's emailed message!) list quantity and does that match the attachment?

- More complexity of file analysis possible with checksum tools and systems (not our focus today).

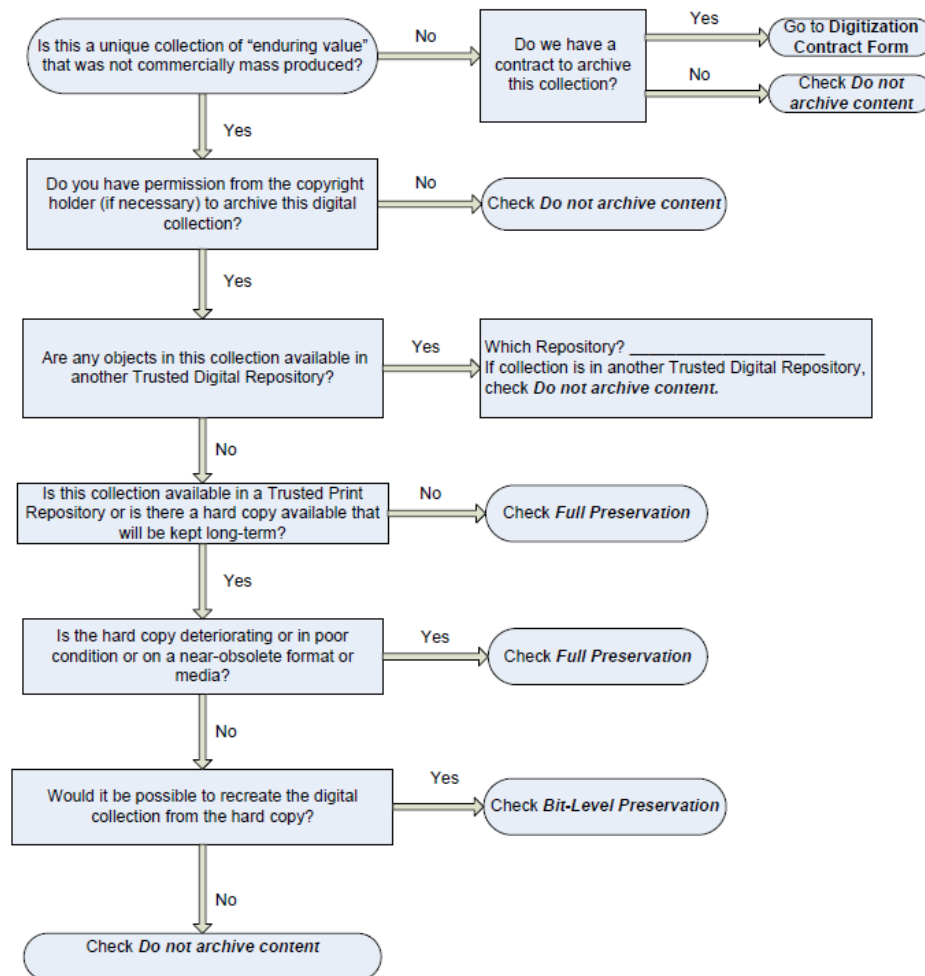**NDSA**

**Table 1: Version 1 of the Levels of Digital Preservation**

| | Level 1 (Protect your data) | Level 2 (Know your data) | Level 3 (Monitor your data) | Level 4 (Repair your data) |
|---|---|---|---|---|
| Storage and Geographic Location | - Two complete copies that are not collocated<br>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system | - At least three complete copies<br>- At least one copy in a different geographic location<br>- Document your storage system(s) and storage media and what you need to use them | - At least one copy in a geographic location with a different disaster threat<br>- Obsolescence monitoring process for your storage system(s) and media | - At least three copies in geographic locations with different disaster threats<br>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems |
| File Fixity and Data Integrity | - Check file fixity on ingest if it has been provided with the content<br>- Create fixity info if it wasn't provided with the content | - Check fixity on all ingests<br>- Use write-blockers when working with original media<br>- Virus-check high risk content | - Check fixity of content at fixed intervals<br>- Maintain logs of fixity info; supply audit on demand<br>- Ability to detect corrupt data<br>- Virus-check all content | - Check fixity of all content in response to specific events or activities<br>- Ability to replace/repair corrupted data<br>- Ensure no one person has write access to all copies |
| Information Security | - Identify who has read, write, move and delete authorization to individual files<br>- Restrict who has those authorizations to individual files | - Document access restrictions for content | - Maintain logs of who performed what actions on files, including deletions and preservation actions | - Perform audit of logs |
| Metadata | - Inventory of content and its storage location<br>- Ensure backup and non-collocation of inventory | - Store administrative metadata<br>- Store transformative metadata and log events | - Store standard technical and descriptive metadata | - Store standard preservation metadata |
| File Formats | - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs | - Inventory of file formats in use | - Monitor file format obsolescence issues | - Perform format migrations, emulation and similar activities as needed |

# Collection decision points

Fact: Not every e-record needs the maximum amount of protection.

How will you decide what can be stored in-house, what goes to the cloud and what might need automatic monitoring?

# DIGITAL PRESERVATION DECISION FLOWCHART



Is this a unique collection of "enduring value" that was not commercially mass produced?

No → Do we have a contract to archive this collection?
- Yes → Go to Digitization Contract Form
- No → Check *Do not archive content*

Yes ↓

Do you have permission from the copyright holder (if necessary) to archive this digital collection?
- No → Check *Do not archive content*

Yes ↓

Are any objects in this collection available in another Trusted Digital Repository?
- Yes → Which Repository? _____ If collection is in another Trusted Digital Repository, check *Do not archive content.*

No ↓

Is this collection available in a Trusted Print Repository or is there a hard copy available that will be kept long-term?
- No → Check *Full Preservation*

Yes ↓

Is the hard copy deteriorating or in poor condition or on a near-obsolete format or media?
- Yes → Check *Full Preservation*

No ↓

Would it be possible to recreate the digital collection from the hard copy?
- Yes → Check *Bit-Level Preservation*

No ↓

Check *Do not archive content*

____ DO NOT PRESERVE

____ FULL PRESERVATION

____ BIT-LEVEL PRESERVATION

**Courtesy of:**
Tawnya Keller, *Digital Preservation Archivist*
University of Utah

# Collection decision points

Questions that are best thought through
sooner rather than later:
1) Does the e-content go with already
acquired material, or
2) Are you being asked to take it just
because it's digital and doesn't take up
"space"?

*Stick to your Collection Development Plan!*

# Storage decision points

Not every storage device is created equal.

> Remember: two copies in distributed storage is the minimum. Can you do this locally? Collaborate with a similar institution? Or will you be able to use a product or vendor?

How will you decide the type of storage to use?

> Hint: This goes back to the collection decision point about what level of preservation you decide your content needs.

# Storage decision points

Is public access necessary?
  And how do you define public—organizational or Google-able?

Does the access system have to be the same as the storage system?
  Will you have ability to provide reference/access on a one-on-one basis? Or the cafeteria-style of letting people retrieve their own content?

*Guess which costs more in money and/or time!*

# Storage decision types

## Distributed storage through collaboration

- With IT, you might choose networked transfer of sites or even specific record types with another organization.

- If no IT, go old school and swap hard drives with someone you trust on a regular basis!

- If no one to swap with use RAID (Redundant Array of Independent Discs) drives as a start.

# Storage decision types

Distributed storage up in the Clouds

- Free but open to the world: *Internet Archive*.
  The lower cost, web-crawling product *Archive-It* does allow for file restrictions.

- Commercial Cloud service providers.
  Numerous but not all equal…Check the fine print!

# Storage decision types

Distributed storage with geographic plus!

- Moderate cost: *DuraCloud*

Allows open or closed file access. Includes fixity checks (AKA checksums, "health checks")

- Full service for a variable cost: *Preservica*

All of the above plus automatic file normalizing and an option to provide a public interface.

# Remember this?

Most tools and services only perform *some* of the functions in a digital curation lifecycle.

**There are front-end/processing tools like…..**

*Duke Data Accessioner*
*Archivematica*

**And there are back-end storage/preservation services like…..**

*MetaArchive*
*DuraCloud*
*Amazon Glacier*
*Internet Archive*

| Ingest | Processing | Access | Storage | Maintenance |
|--------|-----------|--------|---------|-------------|

**But there are a few that will pretty much do it all like….**

*Preservica*
*Dspace Direct (uses DuraCloud)*

**Please Keep In Mind…**
This is NOT exhaustive. Software changes quickly!
Based on availability at time of testing and our perceived needs.

# Don't Panic - Your Pre-Ingest Workflow
aka Wrangling your digital stuff before you can get it into a shiny system

NOTE: This is only ONE way to do this… Everyone's workflow is a little different!

**Starting from scratch:**

- Begin an Inventory Spreadsheet

- Move everything to a stable carrier (like a network drive)

- Make an Access Copy from your Master Copy

- Note these locations in your Inventory Spreadsheet

- OPTIONAL: Keep original media

✓ *Most of these will cost you more time than money*

✓ *Document what you do pre-ingest. For future you.*

✓ *Remember: Good enough is just fine. For now.*

# CONGRATULATIONS!

**We call this "Digital Preservation in Your Office"**

**There are things that need to happen *outside* of your office as well….**

# A tale of two centuries:

# Class of 1914

ALBERT, WILLIAM M. Φ A Δ.
    Vandalia, Illinois.
    Law.
    Modesto (Cal.) High School.
    Peacock Military College, San Antonio, Tex.
    Roswell (N. M.) High School.
    James Millikin Academy, 08.
    Secretary-Treasurer, Senior Class '13-'14.

ALDERSON, OREN.
    Virden, Illinois.
    Biology, B. S.
    Illinois Wesleyan Academy.
    Argus Staff, 1911.
    Debate Board, 1912.

ANNA, GEORGE HERMAN.
    Kinmundy, Illinois.
    Law.
    Kinmundy High School.

ARNOLD, FANNIE.
    Kokomo, Indiana.
    Teacher's Certificate, Piano & Harmony
    Kokomo High School.
    Cooksville High School
    Domestic Science, Purdue University.

# A tale of two centuries:

# Class of 2014



Commencement 2014



*Their legacy is locked up in scattered social media sites—personal and organizational—and who has access?*

*Not me!*

# Outside Your Office

Digital Preservation is not sustainable by just using a tool or selecting a service. Sustainability takes funding and people.

You cannot do this alone. **You will need to talk to other people**… because you are not the only boss of this.

Successful Digital Preservation programs take a team of people at multiple administrative levels.

**Three-Legged Stool of Digital Preservation**
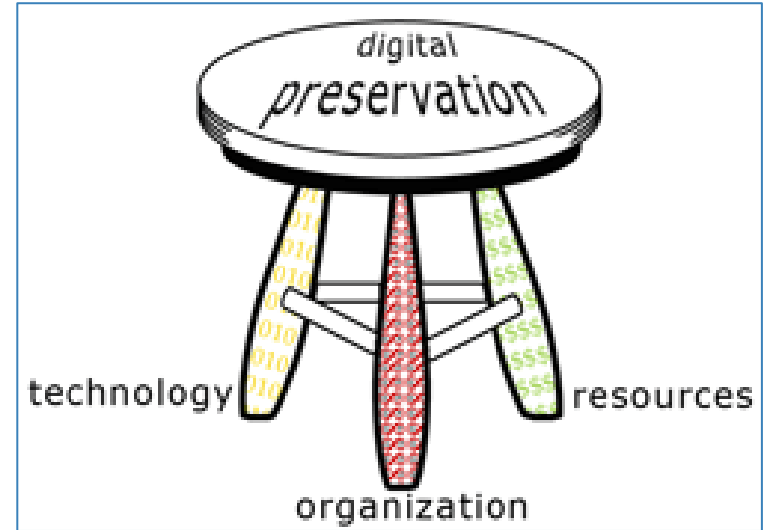


Anne R. Kenney
Nancy McGovern
*Digital Preservation Management Workshop*
http://www.dpworkshop.org/

# Assemble Your Team!



Image: Flickr Commons

# Outside Your Office
## *Group Activity: 3-3-3 Action Plan*

Create a list of all roles in an organization that should play a part in some aspect of digital preservation.

Board? IT? Content creators?

# 3-3-3 Action Plan: Build Your Team

Now let's move from roles to people….

- On your 3-3-3 Action Plan handout, list 3 individuals at your institution in these roles that you already have a working relationship with.

- Which of these folks are you willing to contact in the next 2 weeks?
  …in the following month?
  …in the following 3 months?

- After bringing these colleagues on board, what are 3 concrete, small steps that you can take together to move your burgeoning DP program forward?

---

**~Conversations/Meetings     ~Inventory what you already have**
**~Enhance the metadata of the records you already have**
**~Look at how current policies address digital materials (ex. collection development)**
**~Tool investigation: Dig a little deeper on tools that piqued your interest today**
**~Look at other institutions' DP policies with an eye to crafting your own**
**~Engage in some outreach/education activities…host a Brown Bag!**
**~Read the POWRR white paper**

# Advocacy Before Policy

- Advocacy is valuable because you're educating people about why digital preservation is also THEIR problem.
    - Our one-pagers may help you frame why digital preservation is important to different jobs/function.
    - The risks of doing nothing are a lot greater than they may think.

- Good policies incorporate multiple viewpoints.

- Other people in your organizations will bring up issues – and possible solutions – you may have missed.

- You will discover many things that you don't directly control that still directly affect your work. This will lead you to more people to add to your team.

# Wrapping Up

## Our Final Thoughts & Your Questions

I survived the POWRR workshop! Now what?

https://digitalPOWRR.niu.edu/survived-powrr-wkshp/

We're here to help. Seriously.

YOU CAN DO THIS. Really. But not alone. So bring some friends.
*"If you want to go fast…go alone. If you want to go far…go together."* — African Proverb

Advocacy, Policy and Workflow function best when they are aligned.

Remember: Baby steps still move you forward!

# POWRR Project Team Members
## Contact us…we are here to help!

*Northern Illinois University*

**Lynne M. Thomas**        Curator, RBSC
lmthomas@niu.edu        815.753.0255

**Drew VandeCreek**        Director Digital Scholarship
drew@niu.edu        815.753.7179

**Jaime Schumacher**        Digital POWRR Director
jschumacher@niu.edu        815.753.0576

**Stacey Erdman**        Digital Collections Curator
serdman@niu.edu        815.753.1004

**Danielle Spalenka**        Regional Hist Cntr Curator
dspalenka@niu.edu        815.753.9394

**Matthew Short**        Metadata Librarian
mshort@niu.edu        815.753.9868

**Nathan Books**        Technical Associate
nbooks@niu.edu        815.753.9653

*Chicago State University*

**Aaisha Haykal**        University Archivist
ahaykal@csu.edu        773.995.3843

**Martin Kong**        Systems Librarian
mkong@csu.edu        773.995.3908

*Illinois State University*

**Patrice-Andre Prud'homme**        Digital Collections Head
ppprudh@ilstu.edu        309.438.5385

*Illinois Wesleyan University*

**\*Meg Miner**        University Archivist
mminer@iwu.edu        309.556.1538

*Western Illinois University*

**Jeff Hancks**        Director, Archives and
        Special Collections
jl-hancks@wiu.edu        309.298.2717

**\* Presented at *National Archives Conference for Fraternities and Sororities***