

# DigiVengers: Digital Preservation in Progress

Heidi Southworth, Digital Initiatives Librarian

Jamie Dalbey, Gov Docs/Preservation Technician

Anne Stenzel, Archives Technician

Library Technology Conference 2016

March 16, 2016


Big ideas. Real-world thinking.

# Disclaimers



PLUNK!

- We are not all-knowing experts in everything.
- We are still learning about digital preservation.
- Possible side effects from today may include:
  - A need to learn more about digital preservation.
  - A desire to create your own digital preservation plan.



**POW!**

# Our Digitization Experiences

- Minnesota Digital Library (2000)
- Local Digitization:
  - CONTENTdm in 2006 (photographs)
  - Digitizing student magazine and newspapers in 2010
  - Audio oral histories in 2012
  - Video oral histories in 2015
- ARCH = New platform for Digital collections (2014) – Islandora by PALS
- Cornerstone = Institutional Repository (2014) – bepress Digital Commons®

# Preservation Experiences



- **Preservation of Library Materials**
- **Library Assessment**
  - Comprehensive overview of ALL items in the Library collections along with environmental conditions, best practices, tools for building a preservation program, etc.
  - NEDCC is a great resource
- **Disaster Team**
  - Preparing for disasters that affect the building and contents (physical and virtual)

# Our Current Digital Projects

- **ARCH** – University Archives Digital Collections
  - 9000+ photographs
  - Video interviews – Homecoming Stories, Vikings Training Camp
  - Student Magazines and Newspapers
  - Yearbook
- **Cornerstone** – Institutional Repository
  - Electronic Theses and Dissertations
  - 3 Journals
  - Faculty publications
  - Student works



# Digitization projects are

COOL!

FUN!

But what about...

STORAGE?

COPIES?

FILE  
CORRUPTION?

# **Our Story:** **From Average Library Folk to the DigiVengers**





# **Our Mission...** **if we choose to accept it**

The DigiVengers will implement a digital preservation plan for our expanding digital materials, including solutions for effective storage, usability, maintenance and updates to our digital collections for the continued use and accessibility of students, faculty, staff, and the global online community.



# The DigiVengers Team

- Heidi the Metadata Enforcer
- Jamie the Disaster Master
- Anne the Dominating Digitizer



# Who are Our Allies?

- Library Systems Unit
- Campus IT Department
- PALS
- bepress
- Continuing Education
  - University of Minnesota Digital Preservation = <https://www.lib.umn.edu/dp>



# Who are Our Villains?

- Faux Formats
- Storage Scoundrels
- Dastardly Data Corruption
- Frightening Forgotten Files
- Funding Foes



# Where to Begin?

- Start with the “easy” questions
  - **What do you have?**
  - **How it is organized?**
- Gather some helpful tools (allies)



**SPLAT!**

# Digital Collections Inventory

- What is in our digital collections?

- Begin with a list
- Create a log
- Create unique identifiers to keep track
- Collection development policy =

<http://lib.mnsu.edu/archives/diglib/intro.html>



**BANG!**

# Describing Digital Collections



WHACK!

- **What metadata schemas are used?**
  - Minnesota Digital Library and ARCH = <http://mndigital.org/standards-best-practices/minnesota-reflections-metadata-entry-guidelines>
  - MODS based for ARCH (can export in Dublin Core)
  - Home grown schema for Cornerstone (AKA. messy)
  - Dublin Core based for Cornerstone

# Metadata Tools



Metadata!

- Exiftool (DigiVenger approved™) =  
<http://www.sno.phy.queensu.ca/~phil/exiftool/>
- NARA File Analyzer and Metadata Harvester (DigiVenger approved™) =  
<https://github.com/usnationalarchives/File-Analyzer>
- MediaInfo (audio & video files) =  
<https://sourceforge.net/projects/mediainfo>
- MDQC (batch metadata checking) =  
<https://www.avpreserve.com/tools/mdqc/>

# Organizing Digital Items




**BLAM!**

- **How are digital items organized/arranged?**
  - Consistent file naming structure (our examples)
    - ✓ msu00072\_SMHC\_103\_07\_022\_0006.tif
    - ✓ c\_biol\_fac\_pubs\_1000\_Land\_Allison\_Oncotarget\_2015\_10\_31\_Article.pdf
  - Folder structure matches digital collection structure
  - Importance of logs (<https://youtu.be/RTrAVpK9blw>)
  - ReNamer Lite (DigiVenger approved™) = <http://www.den4b.com/?x=downloads&product=renamer>
  - Bulk Rename Utility = <http://www.bulkrenameutility.co.uk/Download.php>




# More Questions

- Where can we store our digital items?
- How many copies do we keep?
- How are they backed up?
  
- And how do we protect our digital items against...



STORAGE  
SCOUNDREL



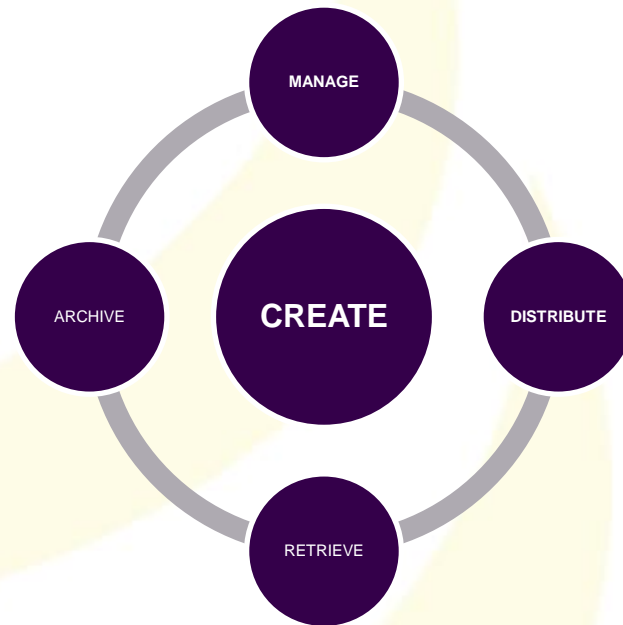
DASTARDLY  
DATA  
CORRUPTION

**DAM!**

# Why should we give a DAM?!

- **What is a DAM?**

- Digital Asset Manager
- Preservation included



- **Our DAMs**

- ARCH and Islandora (true DAM)
- Cornerstone and Digital Commons® (false DAM)

- **Do you know how your DAM is backed up?**

# Duplicate Copies



**SPLONK!**

- **How many copies do we keep?**
  - 3 Copies = Item Born Digital
  - 2 Copies = Physical Item Digitized
- **How are the copies arranged/organized?**
  - BAGGER/BAGIT (DigiVenger approved™)  
(<https://sourceforge.net/projects/loc-xferutils/files/loc-bagger/2.1.3>)

# Duplicate Copies 2.0

- **Where do we keep them?**
  - 4 Drives (networked, backed up by ITS)
    - ✓ Masters are kept here (Born Digital)
  - GOLD (networked storage array)
    - ✓ Masters are kept here (Physical Item Digitized)
  - Hard drives and CDs – backup copy
  - ARCH – backup copy
  - Cornerstone – backup copy



***COPY!***

# Checksums



- **How often do we run checksums?**
  - Every 3 months – check everything
  - Stored with Copies
- **What checksum format do we use?**
  - SHA-1 (recommended by PALS)
- [www.mnhs.org/preserve/records/legislative/records/carol/docs\\_pdfs/ChecksumWhitePaper.pdf](http://www.mnhs.org/preserve/records/legislative/records/carol/docs_pdfs/ChecksumWhitePaper.pdf)

# Checksum Tools



SHA-1

- Exact File = <http://www.exactfile.com>
- DROID = <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid>
- HashMyFiles = [http://www.nirsoft.net/utils/hash\\_my\\_files.html](http://www.nirsoft.net/utils/hash_my_files.html)
- Fixity (DigiVenger approved™) = <https://www.avpreserve.com/tools/fixity>

# Long-Term Storage

- Needs more investigation (Cost, Binding Contracts and Licenses, Security concerns)
- Microsoft Azure (Campus IT Department)
- LOCKSS = <http://www.lockss.org>
  - Digital Commons Private LOCKSS Network
  - Minnesota Private LOCKSS Network



# What's Next?



- Evolving/Living Document
  - Review every semester because it is changing so fast!
  - Draft found at:  
[http://cornerstone.lib.mnsu.edu/lib\\_services\\_fac\\_pubs/53](http://cornerstone.lib.mnsu.edu/lib_services_fac_pubs/53)



# ERM?



- Electronic Records Management
  - Minnesota Historical Society Electronic Records Management Guidelines = <http://www.mnhs.org/preserve/records/electronicrecords/erguidelinestoc.php>
  - Working on our own ERM plans
- ERAS Development (Electronic Records Acquisition Station) DigiVenger™
  - University of Minnesota Electronic Records Task Force Final Report = <http://conservancy.umn.edu/handle/11299/174097>
  - Will be documented in our Digital Preservation Plan

# We Want to Hear from YOU!

- **What Do YOU Do at YOUR Library?**
  - Who are YOUR Villains?
  - Do YOU have a Plan?
  - Tell us More!



**THANK YOU!!**

**THANKS!**

**PLAN!**

**DIGITIZE!**

**DUPLICATE!**

**POW!**

Questions? Contact us at: [archives@mnsu.edu](mailto:archives@mnsu.edu)