# Capturing and Preserving Your Organization's Digital History

Chris Prom

Assistant University Archivist and Professor

University of Illinois Archives

prom@illinois.edu

http://e-records.chrisprom.com

2016 National Archives Conference

for Fraternities and Sororities

# Exercise One

- Work with a neighbor
- Describe something your organization produces in digital format that has long term value
- List one or more challenges that you think might make it difficult to make it accessible 100 years from now.

# Some Common Processing Challenges

- Variety of formats (difficult to render in future)
- Software and hardware dependence
- Bulk of records
- Context easily lost
- Varying information practices (file naming
- Need to adapt existing practices . . .
- While learning new theories and techniques
- (The seeming) complexity of standards
- Understanding which tools to use

# The Meta-problem

- "Certainly there are a range of critical issues confronting special collections and archives in this age of digital production. But of most immediate importance, it would seem to me, is the active acquisition of born-digital material without formal plans for the ongoing management and preservation of these materials."

  —Ben Goldman, *Bridging the Gap*, 12

# Fundamental Question

- What do we **really** need to do to:
  - capture and preserve 'digital objects'
  - Retain them as evidence
  - Show they are authentic
- It's less than you might think . . .

# Top Ten Ways Your Organization can Capture and Preserve its Digital Heritage!

# Number 10

# Decide what to preserve

# Survey your organization

- Methods:
  - Develop simple survey and ask people to complete it
  - survey materials on websites, cloud services and shared drives

# Decision Making Criteria

- Good: By format or media type
- Better: By creating person or group
- Best: By function or activity in the organization
- These are not mutually exclusive and other factors may apply

# Discussion

What digital materials best document your organization and are most likely to be of historical value?

# Number 9

# Know Thyself (and Stewardship Requirements)

# Basic Archives Programs Have

- Appraisal process (does x have continuing value?)
- Transfer method
- Storage 'solution' (however imperfect)
- Descriptive system
- Preservation policies and planning
- Access tools
  - —see Ben Goldman, *Bridging the Gap*

# General Process

- Understand the format and structure of what you are trying to preserve
  - Know your capacities/start small, think big!
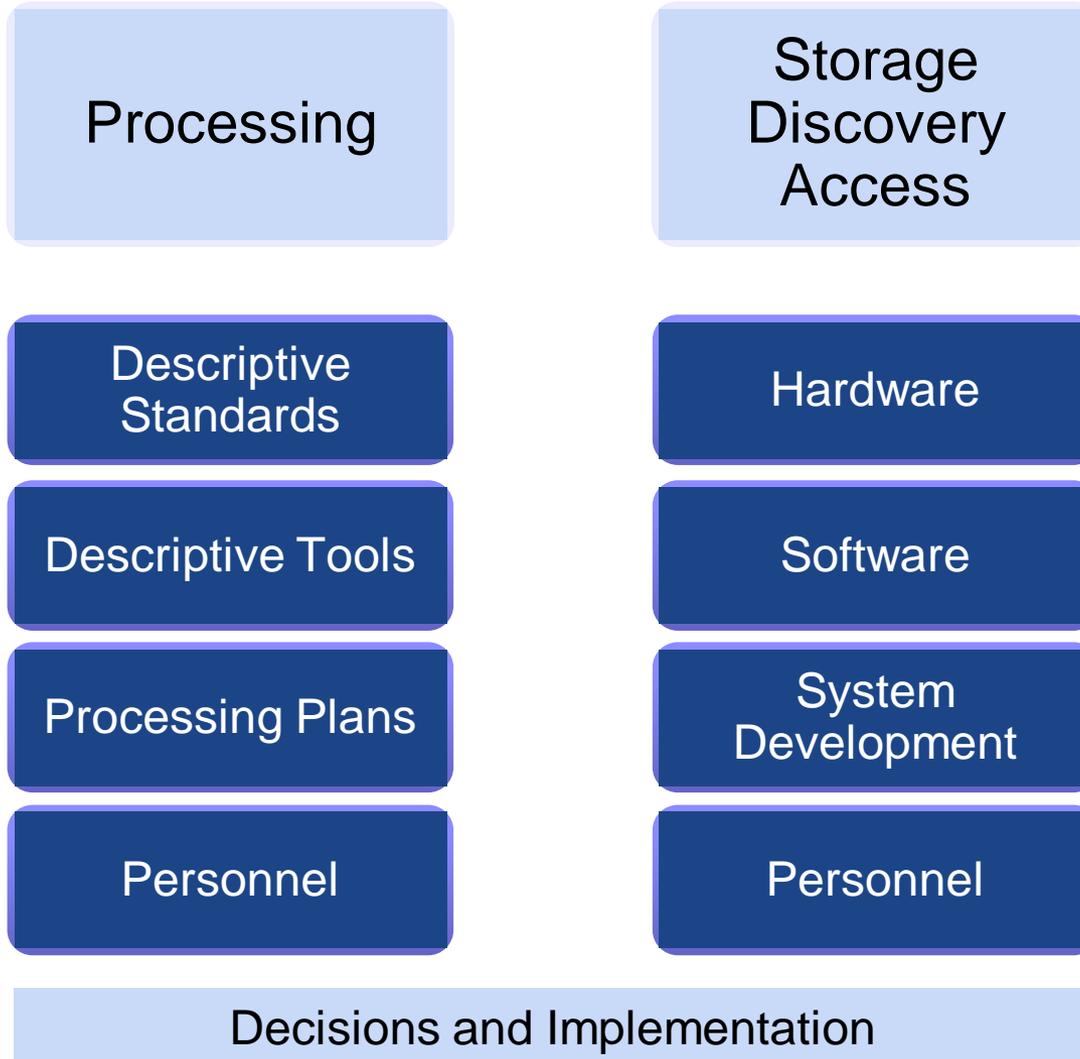- Match tools to tasks, within your existing technology framework

# Summary of Requirements

| Taking Custody | Processing |
|---|---|
| <ul><li>Transfer/copy</li><li>Secure legal custody</li><li>Gather contextual info</li><li>ID conservation issues</li><li>Stabilize records<ul><li>Identify formats, extent, structure</li><li>Achieve rudimentary intellectual control (prepare SIP)</li></ul></li></ul> | <ul><li>Develop processing plans<ul><li>Establish 'intellectual arrangement'</li><li>Note relationships</li></ul></li><li>arrange records<ul><li>Identify and record preservation and access issues</li><li>Organize records</li></ul></li><li>Describe records</li><li>Transfer to storage, discovery, access systems</li></ul> |

Processing

Storage
Discovery
Access

Descriptive Standards

Hardware

Descriptive Tools

Software

Processing Plans

System Development

Personnel

Personnel

Decisions and Implementation

# Number 8

Develop a simple, flexible workflow

UNIVERSITY LIBRARY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

# Do it Yourself Repository

## Define a set of local policies and procedures for file ingest, processing and storage.

- Use simple e-records survey forms and submission agreements to identify and accession records. Keep them with your other accession records (in paper or electronic form) and also put a copy inside the nascent a Submission Information Packet consisting of:
  - Submission Agreement
  - Metadata
  - Content

- At a minimum, track fixity information (e.g MD5 checksums) , for the original files using one of the ingest tools, in the metadata folder 'archival information packet.' Put fixity info in a tab delimited file produced by something like Jacksum or Karen's Directory Printer. This is part of your PDI (Preservation Description Information). Make sure you use the same format for each set of checksums you produce, so you can easily migrate it later.
- Use a batch program on top of FITS to generate metadata about each file. Store the resulting output it in your metadata folder.
- Decide whether or not you want to undertake preservation actions (such a migration) for particular types of files
- If you do, develop a preservation and access plan; then,
- Implement either proprietary or open source tools for preservation actions (e.g conversion/migration). Do not discard any original files! For each set of records you accession, place the originals into an 'originals' folders in your 'Content" Folder. Place the converted files in a 'Processed' folder. These two folders, linked together, will become the 'content information' of your 'archival information packet.' The processed folder will eventually serve as the content information of the 'dissemination information packet.'
- For each preservation action you undertake on the content information or any part of it (e.g. migration to a new format, renaming, deleting, or reorganizing files), keep track of your actions at an aggreate level in. Describe the actions you undertook in a short narrative statement, and keep this information in text file named "PreservationActions.txt" in the root of the 'Metadata" folder. This record serves as part of your preservation description information (PDI). (Unfortunately, there are no easy to use tools currently available to track file level peservation actions taken outside of systems such as RODA, etc.) You might consider saving log output, if available, from your conversion applications and putting it in the Metadata folder.
- Run a batch FITS process and an MD5 checksum generator against the processed folder, and put the results into your Metadata folder.
- For each 'archival information packet,' use your existing catalog system to track collection-level descriptive metadata. Track AIPS at aggregate level (don't worry about single files) before doing detailed descriptive records for part of the AIP. Detailed description of the collection can be done using standard archival cataloging tools.
- As a "dark archives," of the AIP, store all e-records in one folder, containing the subfolders for 'Metadata' and 'Content', for each set of records you are preserving . The main folder must use a unique identifer that is the

18

# Arrangement and Description

Conduct Research

Survey Collection

Create Processing Plan

Arrange Materials

Describe Materials

# Number 7

# Choose your tools carefully

# Basic Tool Types

- Capture
- File Processing (de duplication etc)
- Identification
- Digital Preservation
- Storage
- Description

# COPTR Tools Registry

- http://coptr.digipres.org/Main_Page

# Basic Tool Tips

- Use those at your comfort level
- Avoid tools that use closed or 'functionally closed' formats/technologies
- Open source vs. Proprietary
- photo sharing websites: look for
  - Bulk download
  - uses and preserves embedded metadata

# Exercise Two

- Work with a Neighbor
- List Three pieces of software or services you already use that might help you capture or preserve your organization's digital materials
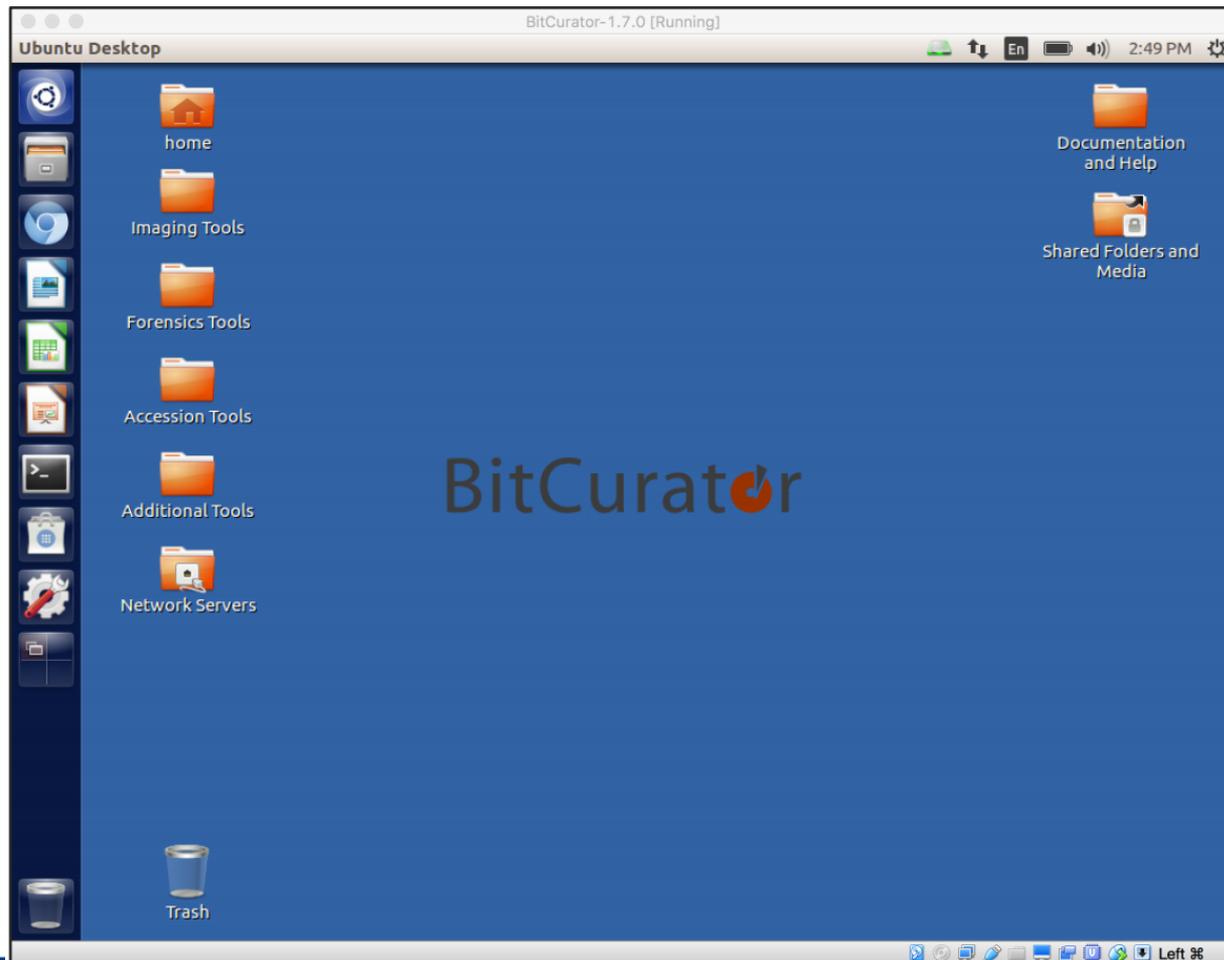
# Number 6

# Give people easy deposit pathways

# Transfer Methods

- Email
- File Sharing Service (Box, Dropbox)
- On device
- Disk imaging (for experts!)

# Forensic Software (expert tool!)

# Taking Custody

- Track what you receive
- "Accession Log Spreadsheet"
- **http://goo.gl/Oy81Ds**

# Automated capture

- Ideal for transactional records and 'lifestreaming' applications (social media)
- Twitter: TAGS (Twitter Archiving Google Sheet)
- Websites:
  - Archive IT, https://www.httrack.com/, wget.
- Facebook and others: ThinkUp
- Social Feed Manager (Experts Only!)

# MailStore Home (for Email)

- Free for personal use

- Downloads to Local Computer

- Must be run manually

- Dependent on scheduling and backup

- [http://www.mailstore.com/](http://www.mailstore.com/)

# EPADD (EMail Process, Appraise, Discover, Deliver)

- Connects to and downloads email
- Allows to arrange, remove, (redact)
- Demo
  - http://localhost:9099/epadd/index.jsp

# Number 5

# Have safe places
# (to put things)

# Storage Options

- 3-2-1 rule
- Options
  - Local
    - requires careful configuration, RAID along not sufficient
  - Non preservation cloud providers
    - e.g. Amazon Glacier
  - Preservation-oriented cloud
    - Duracloud
    - Beware online Backup services!

# Photos

- [https://www.smugmug.com/](https://www.smugmug.com/)
- [http://www.resourcespace.org/](http://www.resourcespace.org/)
- [https://github.com/guardian/grid](https://github.com/guardian/grid)

[http://www.embeddedmetadata.org/social-media-test-results.php](http://www.embeddedmetadata.org/social-media-test-results.php)

# Digital Preservation Services

- DSpace Direct
- Preservica
- ArchivesDirect

# Number 4

# Group like things together

# Key Points

- Records are unique:
    - organic nature,
    - aggregations sharing common provenance (origin) should be kept together and separate from records of other provenance
- original order
- Arrangement means identifying groupings; order can be imposed but usually not necessary

# Archives levels can apply digitally

- Collection
- Series
- File
- Item
  - Example:
    - [http://archives.library.illinois.edu/archon/index.php?p=digitallibrary/digitalcontent&id=7893](http://archives.library.illinois.edu/archon/index.php?p=digitallibrary/digitalcontent&id=7893)

# Processing Plans

- AIMS White Paper
  - http://dcs.library.virginia.edu/aims/white-paper/
- Includes factors such as
  - Formats, extents of material
  - organization
  - technical and preservation problems
  - recommended arrangement and access methods

# "Archival Information Packet"
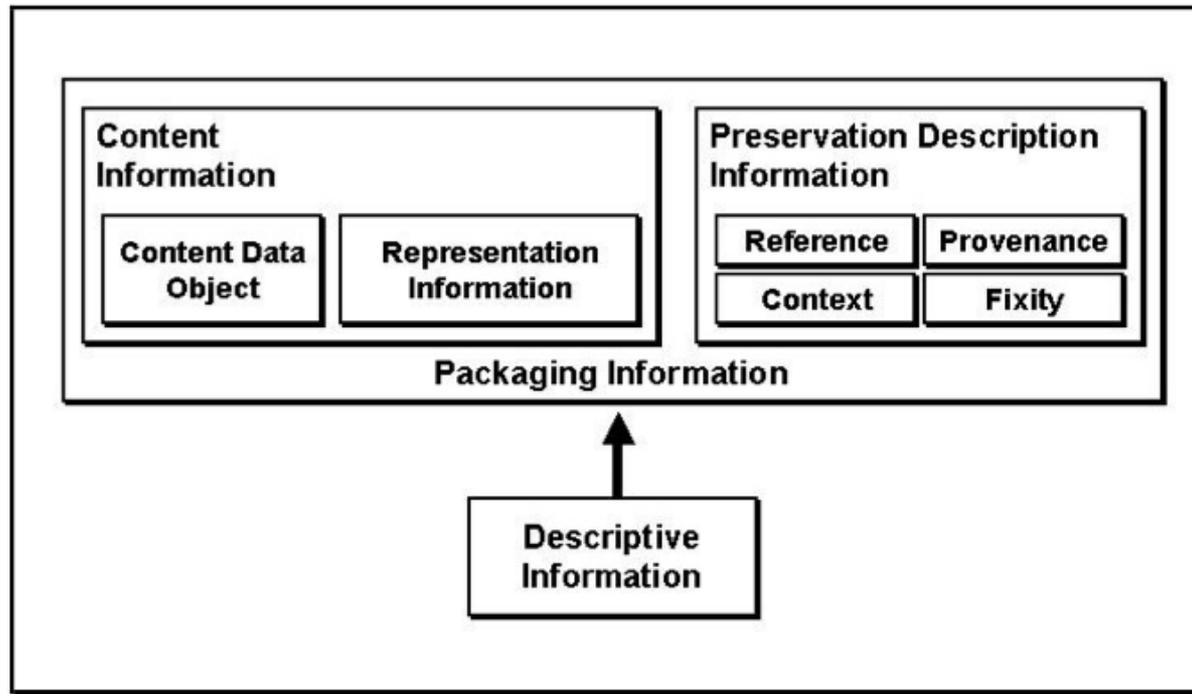


Lavoie, Brian. The Open Archival Information System Reference Model: Introductory Guide. *DPC Technology Watch Report* 04-01. London: Digital Preservation Coalition, 2004.

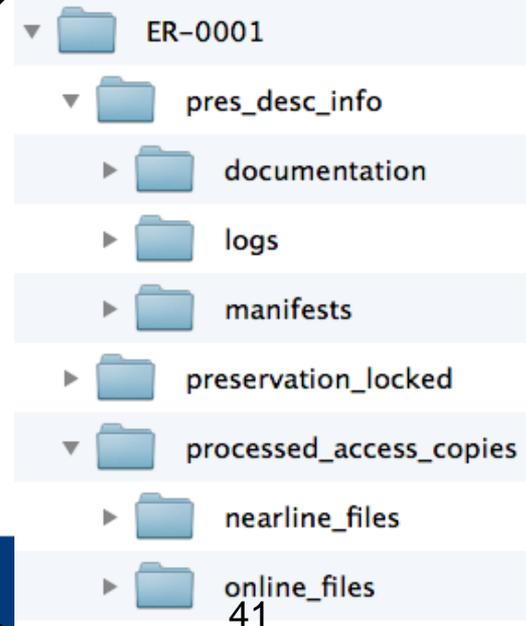# Archival Information Packet

Catalog Record for ER-0001

_____

_____

_____

ER-0001
- Content Information
    - Data object(s)
    - Rep. Information
- Preservation Description Information

=



- ▼ 📁 ER–0001
    - ▼ 📁 pres_desc_info
        - ▶ 📁 documentation
        - ▶ 📁 logs
        - ▶ 📁 manifests
    - ▶ 📁 preservation_locked
    - ▼ 📁 processed_access_copies
        - ▶ 📁 nearline_files
        - ▶ 📁 online_files

# Example

- Records from Center for Democracy in Multiracial Society
  - http://archives.library.illinois.edu/archon/index.php?p=digitallibrary/digitalcontent&id=6854

# Number 3

Describe what you have

# Archival Descriptive Systems

- ICA AtoM
- ArchivesSpace
- Archon
- Also built in to hosted preservation services listed later.

# Archival Description

- Describing Archives: A Content Standard (DACS) = what to record
  - Use professional judgment / practice
    - Choosing an appropriate level of description
    - Choosing and using data elements consistently
    - Describing creators
    - Recording access points
  - http://saa.archivists.org/store/describing-archives-a-content-standard-dacs/223/

# Key Description Fields from DACS

- reference code/ID
- title
- dates
- extent
- name of creator
- scope and contents (formats and subjects)
- conditions governing access
- relationship of parts to whole

# Exercise Three

Thinking of some materials you would like to preserve

- Develop a
  - Title,
  - Extent
  - Scope/Contents Note

# Number 2

always remember:

# "Access drives preservation"

Brewster Kahle, founder of the Internet Archive,
at UNESCO Memory of World Conference, 2012

# AND . . . . . .

# The Number one way

# Work with Partners!!!

# Resources

http://www.digitalpreservation.gov/

http://digitalpowrr.niu.edu/

Ellen and SLC Program

# Capturing and Preserving Your Organization's Digital History

Chris Prom

Assistant University Archivist and Professor

University of Illinois Archives

prom@illinois.edu

http://e-records.chrisprom.com

2016 National Archives Conference

for Fraternities and Sororities