Digital Preservation Webinar Series: Identify
Introductions

- **Sarah Wegley**  
  Digital Metadata and Web Services  
  Governors State University  
  University Park, IL

- **Julie Mosbo**  
  Preservation Librarian  
  Texas A&M University  
  College Station, TX

- **Paul Go**  
  Systems Librarian\Library Technology Manager  
  Illinois Institute of Technology  
  Chicago, IL

  Library of Congress  
  Digital Preservation Outreach and Education (DPOE) Program
DPOE’s Mission

“The mission of the Digital Preservation Outreach and Education (DPOE) program of the Library of Congress is to **foster** national outreach and education to **encourage individuals and organizations to actively preserve their digital content**, building on a collaborative network of instructors, contributors, and institutional partners.”
Modules

Identify - what digital content do you have?
Select - what portion of that content will be preserved?
Store - what issues are there for long term storage?
Protect - what steps are needed to protect your digital content?
Manage - what provisions are needed for long-term management?
Provide - what considerations are there for long-term access?
Objectives

• Learn why an inventory is necessary.
• Learn how to do an inventory for digital preservation.
• Review sample inventories.
• Discuss inventory considerations.
• Prepare for the next stage in the digital preservation process - SELECT.
Why do we identify content?

Do you have an inventory?
How will an inventory help?

All content

Possible to preserve

Actually preserved
An explicit inventory is the best way to identify content.
Creating the Inventory Sheet

- Category
- Title and Description
- Dates
- Location
- Extent
- Format
<table>
<thead>
<tr>
<th>Category</th>
<th>Title and Description</th>
<th>Date</th>
<th>Location</th>
<th>Extent</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Title and Description</td>
<td>Date</td>
<td>Location</td>
<td>Extent</td>
<td>Format</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------</td>
<td>------------</td>
<td>---------------------------------------</td>
<td>----------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>University Archives Digitization Project</td>
<td>The Creative Woman - a feminist journal created at GSU</td>
<td>1977-1992</td>
<td>Internet Archive, GSU Digital Collections - CARLI, Archive DVD</td>
<td>48 journal issues</td>
<td>PDF</td>
</tr>
<tr>
<td>University Archives Digitization Project</td>
<td>Campus Construction - photos and letters from the architect concerning the project’s progress.</td>
<td>1972-1974</td>
<td>GSU Digital Collections - CARLI, Archive DVD</td>
<td>106 photos, 27 letters</td>
<td>PDF, jpg</td>
</tr>
<tr>
<td>Born Digital</td>
<td>GSU View - university newsletter</td>
<td>2007-current</td>
<td>GSUVault, Archive hard drive</td>
<td>76 documents as of 11/2013</td>
<td>PDF versions of the original webpages</td>
</tr>
<tr>
<td>Collection Alias</td>
<td>Size (K=Kilobytes, M=Megabytes, G=Gigabytes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_const</td>
<td>777M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_ethno</td>
<td>202M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_histdoc</td>
<td>208M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_pres</td>
<td>85M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_sculpt</td>
<td>129M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_tcw</td>
<td>155M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gsu_wee</td>
<td>1.4G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CARLI provides statistics for Digital Collections at: http://collections.carli.illinois.edu/carlistats/
<table>
<thead>
<tr>
<th>Category</th>
<th>Title and Description</th>
<th>Date</th>
<th>Location</th>
<th>Extent</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAKTrust Institutional Repository</td>
<td>INA Wachsmann</td>
<td>1970s-1980s</td>
<td>IRODS, DiSC_Upload DiSC_Work</td>
<td>14 binders 6,479 slides</td>
<td>TIFF, JPEG</td>
</tr>
<tr>
<td>Collection Alias</td>
<td>Size (K=Kilobytes, M=Megabytes, G=Gigabytes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRODS_Wachsmann_binder 1</td>
<td>4.26 G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRODS_OMG</td>
<td>638.21 G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRODS_TAMU University Press Books</td>
<td>456 M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The inventory’s **content** is more important than style and format.

*Keep this in mind throughout the project.*
Inventory Considerations

• Scalable
• Current
• Available
• Documented
• Serviceable
Scalable

• A living document
• Add rows *and columns* as new digital content is discovered and new information about that content is pertinent.
• For example, adding a new column to track content that is required to be preserved at a certain level if that becomes required.
Current

• Old inventories are unhelpful.
• Keeping the inventory current helps make sure it is useful.
• Revisiting the inventory on a regular basis can ensure this.
Available

• Inventory should be on a shared directory or cloud service (such as Google Docs).
• Never keep as a single copy on one person’s computer.
Documented

• A documented inventory is able to be referenced later.
• A *living document* should reflect current inventory.
• A documented inventory can be shown to administrators and used for budgeting.
Serviceable

• Make sure the document is useful both as an inventory and a starting off point for other projects.
• An inventory can lead to other projects.
Key Takeaway

Following the SCADS model will help you identify your scads of digital objects.
Sources of digital content

• Scanned photos or documents
• Digital cameras
• Hard drives
• Thumb drives/flash drives
• Floppy disks/zip drives/other media

• CDs/DVDs/video tapes
• Shared network drives
• Digital Collections - collaborative sites hosted on external servers
• Institutional repositories
• Records management tools
Planning Your Inventory Process

- Who will be participating?
- What content and level of detail will you be documenting?
- Where will you store the inventory?
- When is it due?
- How will you know when you are done?
Factors for choosing the level of detail

• Resources available
• Extent of content
• Content’s nature and location
• Timeframe for completion
Key Takeaway

The purpose of the inventory is to assist in later modules. Aim for breadth of content over depth of detail.
Identification Results

Identify potential digital content you may need to preserve.

Treat the inventory as a management tool that grows as your preservation program grows.

Identify gaps in the digital content to inform your acquisition process.

Use the inventory as a basis for the next step in the digital preservation process: SELECT
Questions?

• To learn more about the DPOE curriculum
  http://www.digitalpreservation.gov/education

• To contact the Library of Congress about digital preservation
  http://www.digitalpreservation.gov/contact