A USE-CASE INTRODUCTION TO SETUP, STANDARDS, PROCESS, AND PRESERVATION FOR SMALLER INSTITUTIONS ENGAGED IN STILL IMAGE DIGITIZATION

BRITTANY STRATTON, WENDY MARTIN, AND ANNA LAMPHEAR, UNIVERSITY OF TEXAS LIBRARIES
CONNECTING TO COLLECTIONS CARE (C2CC)
ONLINE COMMUNITY PRESENTATION – NOVEMBER 9TH, 2017
University of Texas Libraries

• Millions of items
• Representing the full spectrum of physical formats
• Housed in our main library, branch libraries, and off-site storage facilities
• Digitizing select materials for preservation and access
Prioritization of Digitization Projects

Digital Project

- IT
- Digital Stewardship
- Subject Specialists
- Content Management
Use case 1: UT Publications

Items from the University of Texas Bulletin series, 1886-1937, and the University of Texas Publication series, 1938-1992
Use case 2: Sharir / Sharir + Bustamante Dance Collection

An unbound scrapbook collection of press clippings, photographs, small pamphlets, letters, awards, and advertisements
# Planning for Digitization and Digital Preservation

Collect data early and consistently; assess condition before starting

---

T**exas Libraries**

<table>
<thead>
<tr>
<th>SIP</th>
<th>2017.0350.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal ID</td>
<td>fal-batch00022-2017-05-10</td>
</tr>
<tr>
<td>Owning Location</td>
<td>FAL</td>
</tr>
<tr>
<td>Title/Description</td>
<td>The Shank/Sharr - Bustamante Dance Collection: External Review Portfolio Binder: Main (black binder), 2000-2005</td>
</tr>
<tr>
<td>Project Name</td>
<td>FAL_Sharr Collection</td>
</tr>
<tr>
<td>Other Identifier</td>
<td></td>
</tr>
<tr>
<td>Date Received</td>
<td>5/16/2017</td>
</tr>
<tr>
<td>Rush</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>books/bound volumes</td>
</tr>
<tr>
<td>Format (other)</td>
<td></td>
</tr>
<tr>
<td># of Items</td>
<td>1</td>
</tr>
<tr>
<td># of Images</td>
<td>117</td>
</tr>
<tr>
<td>Contact Information</td>
<td>Rath Karr, <a href="mailto:rath@austin.texas.edu">rath@austin.texas.edu</a></td>
</tr>
<tr>
<td>Job Phase</td>
<td>Complete</td>
</tr>
</tbody>
</table>

**File Names**: twyjc-02-01-0001

**Notes**: there are a few multi-page documents that will need to be removed from their sleeve to be scanned, and not every document has its own sleeve, please see the attached spreadsheet to make sure the appropriate documents get the right filenames.

status updated the spreadsheet with correct file names in box - 05/7/2017

spreadsheet submitted by FAL did not account for the sleeve with the CED/DVD. So as not to throw off the file names for everything else, the sleeve gets a unique file name, indicated above.

scanning starts with cover of binder

**2017.07.11 given to Anna, who will return it to FAL LJM**
## Scanning Standards

### Federal Agencies Digital Guidelines Initiative

<table>
<thead>
<tr>
<th>Master File Format</th>
<th>1 Star</th>
<th>2 Star</th>
<th>3 Star</th>
<th>4 Star</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TIFF, JPEG 2000,</td>
<td>TIFF, JPEG 2000,</td>
<td>TIFF, JPEG 2000,</td>
<td>TIFF, JPEG 2000,</td>
</tr>
<tr>
<td>Access File Formats</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Resolution</td>
<td>150 ppi</td>
<td>300 ppi</td>
<td>300 ppi</td>
<td>400 ppi</td>
</tr>
<tr>
<td>Bit Depth</td>
<td>8</td>
<td>8</td>
<td>8 or 16</td>
<td>8 or 16</td>
</tr>
<tr>
<td>Color ²</td>
<td>Grayscale or Color</td>
<td>Grayscale or Color</td>
<td>Grayscale or Color</td>
<td>Grayscale or Color</td>
</tr>
</tbody>
</table>

2. Color: The type of color representation in the file.
## File Formats

<table>
<thead>
<tr>
<th>File Format</th>
<th>Compression</th>
<th>File Size</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
<td>No</td>
<td>Large</td>
<td>Requires re-encoding to render</td>
</tr>
<tr>
<td>TIFF</td>
<td>No</td>
<td>Large</td>
<td>Best for master files; highly sustainable</td>
</tr>
<tr>
<td>JPEG</td>
<td>Yes</td>
<td>Smaller</td>
<td>Good for web use and accessibility</td>
</tr>
<tr>
<td>PDF</td>
<td>Yes</td>
<td>Smallest</td>
<td>Good for combining multiple images</td>
</tr>
</tbody>
</table>
Resolution

The sum of pixels in a square inch is the PPI

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ppi</td>
<td>200 ppi</td>
</tr>
<tr>
<td>400 ppi</td>
<td>300 ppi</td>
</tr>
<tr>
<td>700 ppi</td>
<td>400 ppi</td>
</tr>
</tbody>
</table>

Image source: [www.digitalcitizen.life](http://www.digitalcitizen.life)
24 bit = 8 bit + 8 bit + 8 bit
red   green   blue

1-bit = 2 shades
8-bit = 256 shades

Bit Depth
Color Space

• Calibration is ideal if possible
• Scan in a color space, Adobe RGB (1998) if possible but sRGB can be used
• Attach an International Color Consortium (ICC) color profile

Image source: https://commons.wikimedia.org/w/index.php?curid=8359333
Process – Overhead Scanner
Camera Settings and Focus

Camera Settings

Camera Model: Canon EOS 5D Mark II
Camera Port: 1
Shutter Speed: 1/160
F Stop (Aperture Value): 8
ISO Speed: 500
JPEG Quality:
White Balance Setting: Fluorescent
Image Quality: RAW Large
Picture Style: Calibrated Color
Color Space: Adobe RGB

Umlauf, Charles


Hanging: Illinois Wesleyan University.
Check Colors
Capture Cover To Cover
Alternative Process - Camera & Tripod

Required hardware includes:

• Camera (DSLR preferred)
• Remote shutter release switch
• Tripod
• Copy Platform
• Lights
Process – Flatbed Scanner
Settings & Configuration
Targets & Positioning

Kodak Q-13
Preview Scan

Manually select the scan area to:

• Reduce scan time
• Reduce file size
• Introduce quality control
• Decrease post-processing
Post Processing

• Post processing includes image manipulation and compression
• Do not process the master file, create a copy first
• Software is necessary
UT Publications

Archival Master: convert RAW to TIFF
Publication Master: crop & compress as JPEG
Access Derivative: create PDF and run OCR
Sharir Scrapbooks

Archival Master: TIFF created at scanning
Publication Master: rotate, crop & compress as JPEG
Access Derivative: create PDF & OCR the text
Digital Preservation: Strategy

• Technical metadata
• Information packages
• BagIt specification
• LTO tape for long-term storage
• Local record keeping system
Digital Preservation: Representational / Technical Metadata

FITS: File Information Tool Set

Any file

JHOVE
FITS wrapper + XSL

DROID
FITS wrapper + XSL

NLNZ ME
FITS wrapper + XSL

ExifTool
FITS wrapper + XSL

File utility
FITS wrapper + XSL

FFIdent
FITS wrapper + XSL

FITS XML

Consolidator

FITS XML

Exporter

Standard XML

Image source: https://projects.iq.harvard.edu/fits/fits-processing
Digital Preservation: Technical Metadata (continued)

<compressionScheme>Uncompressed</compressionScheme>

<imageWidth>15832</imageWidth>

<imageHeight>12624</imageHeight>

<colorSpace>RGB</colorSpace>

<iccProfileName>sRGB IEC61966-2.1</iccProfileName>

<SamplingFrequency>400</SamplingFrequency>

<bitsPerSample>8 8 8</bitsPerSample>

<samplesPerPixel>3</samplesPerPixel>

<scannerManufacturer>i2s Digibook Scanner</scannerManufacturer>

<scannerModelName>Suprascan A0 14000 RGB</scannerModelName>

<scanningSoftwareName>Digibook Version :6, 2, 0, 383</scanningSoftwareName>
Digital Preservation: OAIS

Open Archival Information System (OAIS) reference model (ISO-STD 14721)
The OAIS Information Model includes three types of information packages:

- Submission Information Package (SIP) = detailed metadata
- Archival Information Package (AIP) = archival content files and detailed metadata
- Dissemination Information Package (DIP) = access content files and limited metadata
Digital Preservation: Package Structure

```
2017_0350 (SIP #)
  2017_0350_files (archival master files)
  2017_0350_files_metadata
  2017_0350_derivatives (publication and access files)
  2017_0350_derivatives_metadata
  2017_0350_sharepoint.txt
  2017_0350_item_spreadsheet.csv
```
Complete bag = AIP
Digital Preservation: bag-info.txt

Source-Organization: University of Texas Libraries

Organization-Address: University of Texas Libraries
The University of Texas at Austin
Post Office Box P Austin, TX 78713-8916

Contact-Name: [Name of the person responsible for the bag; can be the person creating the bag]

Contact-Phone: [International format telephone number of person responsible for the bag; can be the person creating the bag]

Contact-Email: [Fully qualified email address for the person responsible for the bag; can be the person creating the bag]

External-Description: [List of the file formats found in bag payload, using PRONOM registry extension values. Brief, human-readable summary of the content and the overall context from which the content arises. If referencing analog carrier, use controlled vocabularily terms such as PBCore InstantiationPhysical terms for AV and computer media, and Art & Architecture Thesaurus terms for other physical formats. Include items such as project or collection name and relevant dates, date ranges and identifiers. If more extensive descriptive metadata is found in bag payload, note that and give a brief description.]

External-Identifier: [Minimum is to use this element one time with one value that uniquely identifies this bag, such as an ARK or UUID. Element may be repeated with additional identifiers in a key-value pair that points to any external systems also containing this content, such as an OCLC number for the content represented by the files in the bag.]

Internal-Sender-Description: [Describe equipment and software used to generate the files that may be useful in rendering files later. Describe file tiers included in the bag payload, such as 'archival masters'.]

Internal-Sender-Identifier: [Include here, in a key value pair, repeating the element as necessary, identifiers meaningful to the internal local context. Examples include ID (handle, DOI, etc.) for the content in a local access system (IR, online exhibit, etc.), an accession number for the collection from which the files originate, or other local identifier used for recordkeeping purposes.]

Rights-Statement: [Note here any conditions or restrictions associated with the Content Information pertaining to both preservation and access.]

Bag-Group-Identifier: [Use here some convention or identifier that associates a related groups of bags]

Bag-Count: [Count representing this bag, in order, out of all of the bags referred to by the Bag-Group-Identifier. Use format # of #]
Digital Preservation: Bagger

- BagIt-compliant tool for packing and validation
- Graphical user interface
- Customizable

Image source: https://blogs.loc.gov/thesignal/2016/04/baggers-enhancements-for-digital-accessions/
Digital Preservation: Storage

• Lots of data
• LTO tape = best strategy for us
• Other options:
  • Cloud storage
  • External hard drives
• Multiple copies
• Care over time
### Digital Preservation: Record Keeping

<table>
<thead>
<tr>
<th>SIP Number</th>
<th>Title/Description</th>
<th>Source</th>
<th>Bag Group ID</th>
<th>Owning Location</th>
<th>Associated AIP UUID</th>
<th>Tape Number</th>
<th>Associated Identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0002_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0003_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0004_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0005_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0006_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0007_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0008_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0009_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0010_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0011_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0012_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0013_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0014_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0015_0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questions