Collaboration through Analysis: A journey in digital content management workflow analysis

Best Practices Exchange 2019
Columbus, OH

Krista Sorenson, Digital Projects Librarian
N.C. Government & Heritage Library
PROJECT GOALS

1. Establish a precedent of active assessment
2. Foster an open and collaborative environment
3. Address long-standing inefficiencies, frustrations, and issues
4. Position ourselves for adaptability
Government & Heritage Library, State Library of North Carolina

Library processes & manages:
- State Publications – Print & Digitized
- Born Digital State Publications
- North Caroliniana – Print & Digitized

Per month, we process:
- 200-300 Born Digital titles
- 150-200 Digitized titles

Preservation Storage:
- Local & DuraCloud
- 5+ TB of preservation storage
INFORMATION GATHERING

- Catalogers
- State Agency Liaison Librarian
- Metadata Creators
- Supervisors
- Digital Projects Librarian
- Digital Collections Manager
- Systems Librarian

Summer 2017
Met with Digital Staff
(1.5 months)
INFORMATION GATHERING

Every Person Answered:

- What is your role relating to digital collections / digital projects?
- What is an overview of your general workflow(s)?
- What tools / programs do you use?
- What standards / guidelines do you use?
- What else would you like us to know?
INFORMATION GATHERING

Roles & Tasks

Clearinghouse
- Acquisition
- State Pubs Liaison/Selection
- Born Digital Processing (metadata)
- Digitization/Digital Team
- Digitization of materials (State/Non-state)
- In house / Off site
- Adding Items to CONTENTdm
- Preservation management
- Digital Collection Creation/Updating & troubleshooting

Cataloging
- Born Digital/State Pub Record Creation/Updating
- Non State Pub Record Creation/Updating
- Metadata Cleanup & Controlled Vocab Maintenance

Priorities for Year

1. Workflow Overview
2. One “document”/Location for all Workflow Documentation/guidelines
3. Serial “Re-Imagining” that...
   - Avoids giant Compound objects
   - Makes Serials manageable long-term
   - Optimizes usability & functionality
   - Satisfies Standards/Guidelines
4. Workflows, like...
   - Include updating
   - Foldering, org. + Processing on S&K drive
   - CINCH
   - CDM tools (admin, OCLC Connectome, Project Client)
   - Tracking
   - Digital Preservation
   - Filename
5. Metadata - Controlled Vocab, mapping, collaborations w/ Archives
WORKING GROUP

- Determine digital content management functional requirements
- Evaluate current tools
- Analyze digital workflows
- Troubleshoot and address immediate issues, concerns

Fall 2017
Formed a Working Group
WORKING GROUP

- Group Members
  - Digital Projects Librarian
  - Digital Collections Manager
  - Head, Content & Info Delivery
  - Head, Content Management & Access
  - Systems Support Librarian

- Ad Hoc Support
  - Liaison Librarian
  - Cataloger / Metadata Manager
  - Systems Integration Librarian
WORKING GROUP: GOALS & SCOPE

- Streamline workflows
- Ensure early application of preservation
- Identify opportunities for automation
- Determine standards
- Develop new workflows
- Update and/or create documentation
FUNCTIONAL REQUIREMENTS & TOOL EVALUATION

Fall 2017 – Spring 2018 (9 months)
## FUNCTIONAL REQUIREMENTS

<table>
<thead>
<tr>
<th>ID</th>
<th>Functional Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Track Digitization Status / Process</td>
</tr>
<tr>
<td>02</td>
<td>Acquire Born Digital Files - more direct &amp; verifiable, the better</td>
</tr>
<tr>
<td>03</td>
<td>Duplication Alerts</td>
</tr>
<tr>
<td>04</td>
<td>Schedule Alerts - Claims Report</td>
</tr>
<tr>
<td>05</td>
<td>Authenticity Check</td>
</tr>
<tr>
<td>06</td>
<td>Fixity - Generate Checksums (MD5, SHA-1)</td>
</tr>
<tr>
<td>07</td>
<td>Track Chain of Custody</td>
</tr>
</tbody>
</table>
## FUNCTIONAL REQUIREMENTS

<table>
<thead>
<tr>
<th>ID</th>
<th>Functional Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>08</td>
<td>Metadata Creation:</td>
</tr>
<tr>
<td></td>
<td>&gt; Preservation - xml, json</td>
</tr>
<tr>
<td></td>
<td>&gt; Description - ability to add to pres</td>
</tr>
<tr>
<td>09</td>
<td>Batch File Renaming</td>
</tr>
<tr>
<td>10</td>
<td>Virus Check</td>
</tr>
<tr>
<td>11</td>
<td>Grab Files &amp; Metadata from Vendors</td>
</tr>
<tr>
<td>12</td>
<td>Structure SIPS / Folders</td>
</tr>
<tr>
<td>13</td>
<td>Crosswalk / move Descriptive Metadata from OCLC</td>
</tr>
<tr>
<td>14</td>
<td>Way to report various statistics</td>
</tr>
</tbody>
</table>
## FUNCTIONAL REQUIREMENTS

<table>
<thead>
<tr>
<th>ID</th>
<th>Functional Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Track Digitization Requests</td>
</tr>
<tr>
<td>16</td>
<td>Record whether Born Digital vs Digitized</td>
</tr>
<tr>
<td>17</td>
<td>Tracking System for all items</td>
</tr>
<tr>
<td>18</td>
<td>Create OCLC Record for new records</td>
</tr>
<tr>
<td>19</td>
<td>Metadata Management</td>
</tr>
<tr>
<td>20</td>
<td>Ingest Content to Content Management System</td>
</tr>
<tr>
<td>21</td>
<td>Ingest Content to Preservation System</td>
</tr>
</tbody>
</table>
## TOOL EVALUATION

<table>
<thead>
<tr>
<th>ID</th>
<th>Tools</th>
<th>Open Source</th>
<th>Free</th>
<th>Proprietary</th>
<th>Hosted</th>
<th>Built In house</th>
<th>Applies to Functional Req</th>
<th>Satisfies Functional Req</th>
<th>Possible Future Use for Functional Req</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Born Digital Access Database</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>03, 07, 16, 17, 19</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>b</td>
<td>Digitization Queue</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>01, 08, 14, 15, 16, 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>CONTENTdm Project Client</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>08, 19, 20</td>
<td>08, 19, 20</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>OCLC Connexion</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>08, 13, 18, 20</td>
<td>08, 18</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>ReNamer</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>08, 09</td>
<td>08, 09</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>HashMyFiles</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>06, 08</td>
<td>06, 08</td>
<td>03</td>
</tr>
</tbody>
</table>
# TOOL EVALUATION

<table>
<thead>
<tr>
<th>ID</th>
<th>Tools</th>
<th>Open Source</th>
<th>Free</th>
<th>Proprietary</th>
<th>Hosted</th>
<th>Built In house</th>
<th>Applies to Functional Req</th>
<th>Satisfies Functional Req</th>
<th>Possible Future Use for Functional Req</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>Adobe Acrobat</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>08, 12</td>
<td>08, 12</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>CINCH</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>02, 03, 05, 06, 08, 10, 11</td>
<td>[see notes]</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>TeraCopy</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>05, 12, 21</td>
<td>05, 12, 21</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>DuraCloud Sync Tool</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>05, 21</td>
<td>05, 21</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>Cardinal / Evergreen</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>08, 19</td>
<td>[see notes]</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>Microsoft Excel</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>01, 07, 14, 15, 16, 17, 19</td>
<td>01, 07, 14, 15, 19</td>
<td>16</td>
</tr>
<tr>
<td>ID</td>
<td>Tools</td>
<td>Open Source</td>
<td>Free</td>
<td>Proprietary</td>
<td>Hosted</td>
<td>Built In house</td>
<td>Applies to Functional Req</td>
<td>Satisfies Functional Req</td>
<td>Possible Future Use for Functional Req</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------</td>
<td>-------------</td>
<td>------</td>
<td>-------------</td>
<td>--------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>o</td>
<td>Adobe Photoshop</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>FTP / FileZilla</td>
<td>x</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>05, 11, 19</td>
<td>05, 11, 19</td>
<td></td>
</tr>
<tr>
<td>q</td>
<td>Notepad++</td>
<td>x</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>19</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>IA Downloads and MARC Metadata extraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>02, 08, 11, 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>Dropbox</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>02</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>E-mail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>02, 07, 15</td>
<td>02, 07, 15</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>Fujitsu Scanning Software</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TAKE AWAYS

- Reliance broken tools and approaches
- Time
- Inaccurate assumptions
- Ignorance of institutional history ("why")
- "That’s how we’ve always done it"
WORKFLOW ANALYSIS

- 2-4 hour interviews with each staff member who works with digital content management
- Walkthrough + Clarifying Questions for:
  - Tools they use
  - Scenarios / Tasks & their Processes
- Discussed pain points and “why”

Fall 2018
Conducted workflow analysis interviews
(4 months)
WORKFLOW ANALYSIS

- Combined notes from each interview
- Transferred notes into flowcharts
  - Whiteboards to Digital
- Color Coded different scenarios
  - Born Digital vs Digitized
  - State Publication vs General Collection
  - Digitized In House vs Off Site

Winter 2019 - Present
Charted Workflows
(4 months)
CATALOGING
BORN DIGITAL STATE PUBLICATIONS
PRESERVATION
TAKEAWAYS (THUS FAR)

- Slow process that takes commitment
- Breakdowns already emerging
- Miscommunication = repeated/unnecessary work
- Tool based approaches vs. functional, goal based approaches
- Abundance of tedious tasks
- People are frustrated
OUTCOMES (THUS FAR)

- More open dialogue and feedback
- A shared context
- Better understanding of others’ work
- Buy-in and investment outside “my job”
- Increased willingness to explore new solutions
STAY CONNECTED

Krista Sorenson
Digital Projects Librarian
krista.sorenson@ncdcr.gov
(919) 814-6804