OSSArcFlow
Researching Archival Workflows for Born-Digital Content

Sam Meister
@samalanmeister
Open Repositories 2018
Bozeman, MT

#ossarcflow | ossarcflow_public@educopia.org
Project Goals

investigate, model and test a range of workflows for libraries and archives to curate born-digital content.
Research Questions

▪ How can institutions combine tools to support workflows that meet local institutional needs?
▪ How can institutions implement “handoffs” between different function-based systems?
Background

- Derived from BCC community discussions
- Sam Meister, pilot research on workflows (2014-16)
- Shared concerns among OSS Service Providers
- Applied research needed to examine field-wide challenges
Activities

- **Document** workflows for 12 partner institutions
- **Develop** tools (scripts) to facilitate hand-offs
- **Design** training modules that will promote use of documentation and scripts
- **Create** “Implementation Guide”
Project Partners

Atlanta University Center, Robert W. Woodruff Library
District of Columbia Public Library
Duke University
Emory University
Kansas Historical Society
Massachusetts Institute of Technology
Mount Holyoke College
New York Public Library
Rice University
Stanford University
New York University
Odum Institute
Project Team

PIs
Christopher “Cal” Lee (UNC SILS)
Sam Meister (Educopia Institute)
Katherine Skinner (Educopia Institute)

Team
Kelly Stewart, Sarah Romkey (Artefactual Systems)
Laney McGlohon, Christine DeBella (ArchivesSpace)
Kam Woods (UNC SILS)
Jessica Meyerson / Alex Chassanoff (Educopia Institute)
Courtney Vukasinovic / Caitlin Perry (Educopia Institute)
Andrew Rabkin (UNC SILS)
Yinglong Zhang (UNC SILS)
OSS Systems

Forensic disk imaging
File system analysis and reporting
Identification of PII

BitCurator

Supports core collection management
Authority control
Event tracking & reporting

ArchivesSpace

Lower barrier to digital preservation
Standards compliant - OAIS+
Microservice approach

EDUCOPIA INSTITUTE
Project Partners

Desired Outcomes from OSSArcFlow

- Identify and test tools that will help us to implement our digital preservation plan
- Collaborate with experts in the field and other institutions grappling with similar issues
- Move towards greater consistency in our workflows across collections
- Ways to automate and streamline existing digital curation workflows
Project Advisors Desired Outcomes from OSSArcFlow

● **Understand**
  ○ how people use/want to use these systems in their curation workflows?
  ○ how users think about these OSS tools in relation to other curation tools?
  ○ how service providers can better support the use of these systems in conjunction with other curation tools?
  ○ the institutional drivers that lead institutions to different uses of our software

● **Develop better sense of community standards**

● **Strengthen collaborative bridges (human and tech)**
Timeline

Year 1: July 2017 - June 2018

July - Sept.
- Document “as is” workflows - incl org context survey
- Develop workflow template

- Setup partner OSS test instances

Jan. - March
- Document “aspirational” workflows

April - June
- Publication of “as-is” workflows
Timeline

Year 2: July 2018 - June 2019

July - March
- Refine aspirational workflows
- Generate and test system hand-off scripts
- Solicit feedback on Implementation Guide

Jan. - March
- OSS community feedback on development roadmap

April - June
- Final release of partner and non-partner workflows, methods and scripts, Implementation Guide
Methods

- Surveys
- Semi-structured interviews (who, what, when)
  - Current
  - Aspirational
- Visual Modeling
- Comparisons across
  - Systems
  - Organizational Contexts
Artifacts: Workflow Representations

Rep #1 Procedural Narrative

**TRANSFER/ACCESSIONING**
- Digital content is acquired:
  - [Digital Archivist or Archives Staff] IF on physical storage media THEN
    - [Archivist] takes picture of the media
  - IF filesystem is known THEN:
    - Triaged from physical media using either Guymager or Forensic Toolkit (FTK) Imager
  - IF filesystem is unknown THEN:
    - Triaged from physical media using Kryoflux

Rep #2 Tabular Steps

<table>
<thead>
<tr>
<th>phase</th>
<th>step</th>
<th>description</th>
<th>hardware</th>
<th>software</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-accession</td>
<td>high-level content analysis, donor negotiation</td>
<td>[Michael and/or Curator] Appraisal meeting/donor negotiation using a survey instrument that gathers information about what the collection is and what it contains.</td>
<td>survey instrument (Google sheet? Excel spreadsheet?)</td>
<td></td>
</tr>
<tr>
<td>transfer/accession</td>
<td>photographic documentation</td>
<td>[Archivist] or [Digital Archivist or Archives Staff] takes a picture of the media.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>transfer/accession</td>
<td>forensic examination</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rep #3: Visual Diagram

- Gather information about material
- Obtain sample context
- Meet with or talk to acquiring institution donor
- Set expectations, collectives
- Photograph physical media
- Document soft facts/digital抓获

Form digital material

Digital file, e.g., cloud transfer

Physical media

No transfer/digital capture

No information gained
Artifacts: Workflow Representations

OSS Systems

Curation Phase

Staff role
Artifacts:
OSS System Starter

User Personae

Digital Archivist or Digital Curation Librarian

Needs
- Ability to document and track digital curation activities
- Provides training to other staff on digital curation
- Accountable to the organization for implementing

Tasks
1. You want to reuse metadata you have created in the process of exporting metadata from ArchivesSpace.
2. You have just received a collection with low capacity to be forensically imaged. Use tools within the BitCurator to back up data from the media.
3. Imagine you are looking for a way to reuse the metadata from previous microservices. Examine the contents of the AIP created to identify 1-3 ways you could reuse this data in your current processes.

"As a manager, I want a dashboard view of our progress toward our annual goal."

System Starter Tasks

ArchivesSpace: Accessioning

Your organization has just received a new collection of digital material. Use ArchivesSpace to document information about the transfer process and a high-level description of the contents.

Please click “continue” when you are done.

What You Need:
- Make sure you record your experiences in the OSS System Task Worksheet available here: https://go.gi/q2vXH4

As you experiment, consider other users of these systems including additional staff roles (archivist, administrator, etc.) and end users.

New User Stories
Preliminary Findings:

- Transition from "silos" of digital activity to holistic approach towards digital content at key points
- Tracking content across systems and tools
- Reducing time spent transforming system 1 output -> system 2 input
Partner Comparison - Tools

**NYPL**
- BitCurator
- Archivematica
- ArchivesSpace
- Filemaker Pro
- Forensic Toolkit
- ePADD

**Rice University**
- BitCurator
- Archivematica
- ArchivesSpace
- DSpace
- DROID
- ExifTool
- Quickhash / Fixity
<table>
<thead>
<tr>
<th></th>
<th><strong>NYPL</strong></th>
<th><strong>Rice University</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff Members</td>
<td>4 staff members</td>
<td>4 staff members</td>
</tr>
<tr>
<td>Time Distribution</td>
<td>100% time working on digital curation</td>
<td>20% time working on digital curation</td>
</tr>
<tr>
<td>Rotating Strategy</td>
<td>Rotating archivist strategy</td>
<td></td>
</tr>
</tbody>
</table>
Partner Comparison - Workflows

NYPL

Rice University
Partner Comparison - Pain Points

**NYPL**

“Definitely care about Aspace and Amatica talk to each other and share metadata, and push that metadata to researchers.”

**Rice University**

“It would be nice if there was a way for BCE to talk to Aspace. But it’s a quarantine system. Or if there was a way to track information – a lot of spreadsheets. In Aspace or Archivematica.”
Current Activities

Data Analysis

- Analyze and synthesize gaps and challenges across workflows
- Identify pain points and potential metadata handoffs
Next Steps

Requirements Gathering

- Create development tasks for system hand-offs
- Generate scripts to address potential system hand-offs
Community Engagement

Public Release of Current As-is Workflows (late June!)

OSSArcFlow Quarterly Newsletter

Community Forums
Stay in Touch!

Project Webpage:
educopia.org/research/ossarcflow

OSSArcFlow Public Listserv:
https://groups.google.com/a/educopia.org/d/forum/ossarcflow_public