

CoSA & Preservica Practical Digital Preservation 2017



Preserving Digitized State Government Records

May 9, 2017

Practical Digital Preservation 2017

Welcome!

PDP Briefings

- Protecting and Preserving Long-Term Digital Information January 24
- For IT Professionals & practitioners
- The Governance of Long-Term Digital Information May 23
- For Senior Managers & Budget Administrators



Bonita Weddle
New York State Archives

PDP “Hot Topic” Webinars Tuesdays 2-3pm Eastern

- Preserving and Protecting Audio-visual Files April 11
- Preserving Digitized State Government Records May 9
- Best Practices in Digital Preservation: International Perspective June 13

PDP Online Workshops - Digital Preservation 101:

‘State Archives and Agencies Putting Digital Preservation into Action’

- Part 1: Practical Training in the Key Concepts February 14
- Part 2: Practical Training in the Key Concepts February 28
- Part 3: Case Studies March 14

Sign up today on the CoSA website – PERTTS Portal > Education -Training

Today's Presenters



Vince Brooks
Library of Virginia



Allen Ramsey
CT State Library



David Portman
Preservica



Bryan Collars
South Carolina

Agenda



- Welcome
- Best Practices for Digitization Project Management
- Concerning a brief overview of metadata in an electronic records transfer at the SCDAH
- Virginia Digital Chancery Project
- Working with Digitized Content: Technical Considerations
- Q&A

Best Practices for Digitization Project Management

- 2011 - CoSA State Electronic Records Initiative (SERI) program focuses on challenges of electronic records management and preservation
- SERI Tools and Resources subcommittee identified need for a guidance document providing best practices for digitization project management
- Committee discussion centered on desire of many of state and local government agencies to go “paperless” but is that the best use of their resources?
- Guidance document geared towards our audiences of state archives and records management programs but basic principals apply to the management of any state or local government agency digitization project

Overview of the best practices

- Four broad categories:
 - Pre-Project Planning
 - Project Preparation
 - Project Management
 - Post-Project
- Each category includes strategies and best practices of successful digitization projects

Pre-Project Planning

- Pre-project planning is often documented in a **Project Plan**.
- At the very basic foundation of any digitization project is its purpose. Why is this project being done and why is it being done now?
- The foundation of your project will be built upon these three questions:
 - What do you have?
 - What are you trying to do?
 - Why are you doing it?

Pre-Project Planning (cont)

- Stakeholders
 - Identify
 - Acquire approvals (if necessary)
 - They can provide valuable input or insulate the project from changing institutional priorities or budgets, and thus their buy-in is essential.
- Budget estimates
 - Costs
 - Grants

Pre-Project Planning (cont)

- Risks
 - Document
 - Mitigate
- Specifications
 - Technical
 - Destination
- Implementation
 - In-House
 - Outsource

Project Planning

- Prepare the records
 - Physical preparation
 - Arrangement
 - Description
 - Packaging and shipping
- Staff
 - Project Manager
 - Additional in-house staff
 - Communication
- Post Digitization Management
 - How will you manage the originals and the digital copies?

Project Management

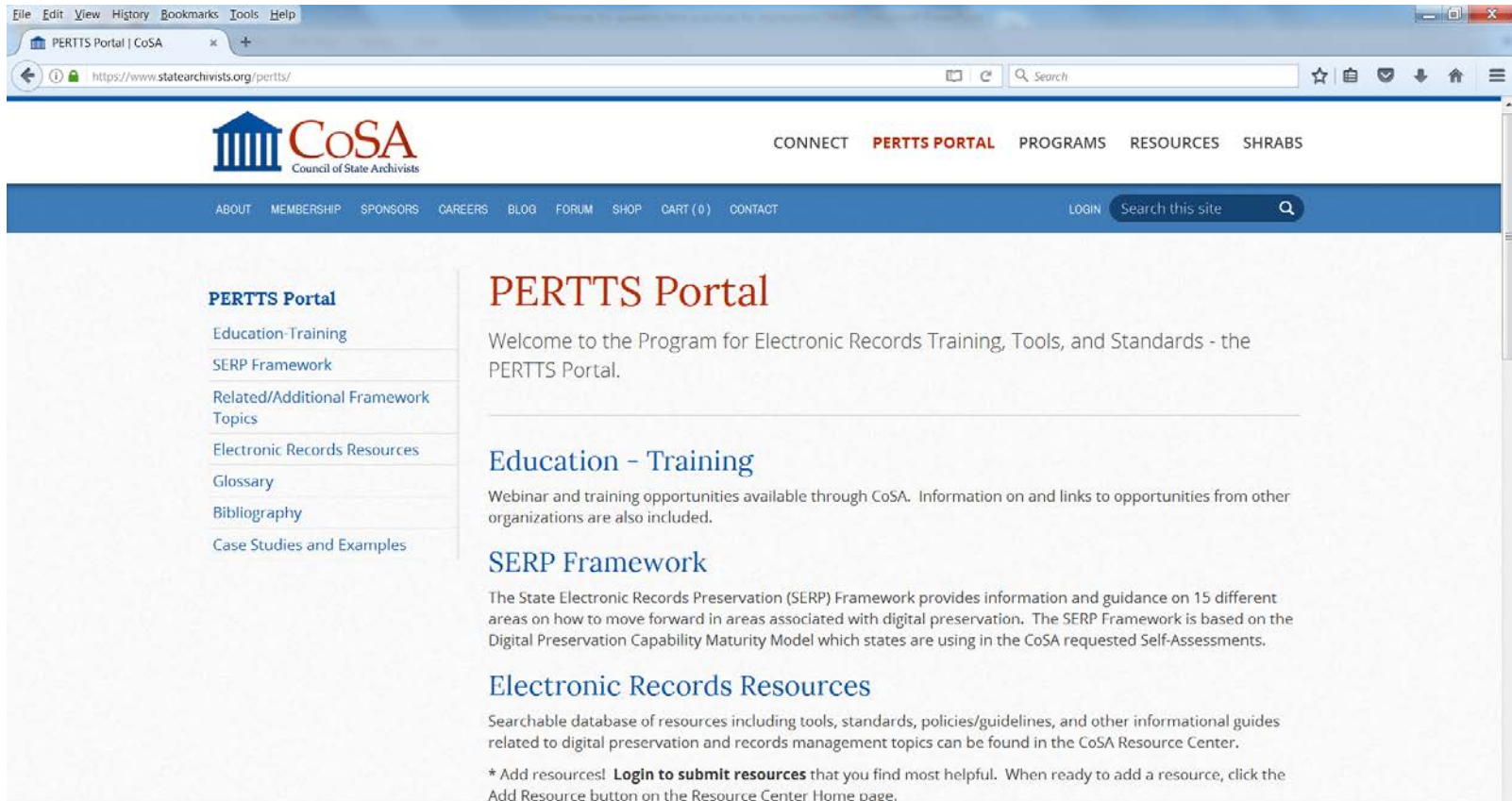
- The day has come for the project to begin! How do you actually manage this thing?
- Leadership and Direction
 - Assign a single person or very small group as the project manager(s).
- Workflow and Procedures
 - Communication
 - Description
 - Digitization Workflows
- Quality Control and Quality Assurance

Post-Project

- Accounting
 - Time spent on project
 - Financial resources (i.e. new equipment, cost share)
- Lessons Learned
 - Was the project successful?
 - What could be done better in future projects?
- Finalize Documentation
 - Final report provides a full accounting for the entire project and is not just a last report.
- Sustainability (Long Term Management)

Coming Soon!

Best Practices for Digitization Project Management



Series of Best Practices and Guidance Documents

Are there best practices or guidance documents that you would find the most helpful or necessary in your work?

Let me know!

Allen Ramsey

allen.ramsey@ct.gov

860-757-6512

METAMESS

CONCERNING A BRIEF OVERVIEW OF METADATA IN AN ELECTRONIC RECORDS TRANSFER AT THE SCDAH

Bryan Collars
Electronic Records and
Imaging Supervisor

South Carolina
Dept. of Archives and History



METADATA MYTHCONCEPTIONS

We all “knew” that when archival institutions started to receive electronic records the metadata and the records themselves would be in an arrangement and order that would make them easily accessible, right?

METADATA MYTHCONCEPTIONS

**WE WERE
WRONG**

METADATA MYTHCONCEPTIONS

Electronic records are messy, not in the manner in which paper records are messy, but rather in the manner in which they were stored by an individual, or an automated system or a proprietary software suite.

METADATA MYTHCONCEPTIONS

For Example

```
1915-018179~EDWARDS~GEORGE~L~1915-06-25~BAMBERG~ALMA~RIVERS~AFFIDAVIT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000001_00000001_1.pdf  
1915-022991~CARTER~SUDIE~1915-06-02~DARLINGTON~ELLEN~BELL~AFFIDAVIT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000002_00000002_1.pdf  
1915-026743~MCKENZIE~CHARLES~WESLEY~CLARENDON~LILLIE~DRIGGERS~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000003_00000003_1.pdf  
1915-039418~HINDS~LEON~1915-11-23~FLORENCE~ALICE~KNOTTS~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000004_00000004_1.pdf  
1915-001934~BALLENTINE~DOROTHY~LOUISE~1915-01-27~RICHLAND~LOUISE~W~ROBINSON~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000005_00000005_1.pdf  
1915-044579~GOODLETT~RUBY~ANNABETH~1915-12-16~SPARTANBURG~JAOLIA~THOMPSON~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000006_00000006_1.pdf  
1915-025302~REESE~ALICE~EARLINE~1915-07-26~SUMTER~SINDER~HOLLIN~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000007_00000007_1.pdf  
1915-021446~LAWSON~EMMA~1915-06-20~UNION~HATTIE~AGNES~PATTESON~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000008_00000008_1.pdf  
1915-021902~REEVES~LOUIE~OLIN~1915-07-20~ANDERSON~BESSIE~GILSTRAP~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000009_00000009_1.pdf  
1915-025084~VASSEY~OMAHA~JOHN~SPARTANBURG~OLA~NIX~AMENDMENT~~~~~  
@@C:\TEMP\BIRTH_1915_COMPLETE\00000010_00000010_1.pdf
```

METADATA MYTHCONCEPTIONS

| | | | | | |
|--|---|----------------------------------|--|---------------------------------|---|
| 07-06-15 M LMM 05 | Affidavit of Correction to Birth Record SOUTH CAROLINA DEPARTMENT OF HEALTH AND ENVIRONMENTAL CONTROL | | | Page 2 of 2 | |
| Enter Correct Information Concerning Person Whose Birth Record is Being Amended | REGISTRANT'S FULL NAME AT BIRTH | | | STATE FILE OR BIRTH NUMBER | |
| | GEORGE L EDWARDS | | | 139-15-018179 | |
| | BIRTH DATE | Month JUNE | Day 25 | Year 1915 | CITY OR TOWN BAMBERG, SOUTH CAROLINA |
| ITEMS TO BE AMENDED OR CORRECTED | ITEM OMITTED OR IN ERROR | | BIRTH CERTIFICATE SHOWS | | SHOULD BE |
| | CHILD'S GIVEN NAME | | OMITTED | | GEORGE L EDWARDS |
| | | | | | |
| AFFIDAVIT | I HEREBY DECLARE UPON OATH THAT THE ABOVE STATEMENTS ARE TRUE AND CORRECT. SIGNATURE OF PARENT (OR OTHER) <i>George L. Edwards</i> | | | RELATIONSHIP SELF | |
| NOTARY (AFFIX SEAL) | SUBSCRIBED AND SWORN TO BEFORE ME ON <i>See Attached</i> | | SIGNATURE OF NOTARY NOTARY COMMISSION EXPIRES | | |
| AFFIDAVIT | I HEREBY DECLARE UPON OATH THAT THE ABOVE STATEMENTS ARE TRUE AND CORRECT. SIGNATURE OF PARENT (OR OTHER) | | | RELATIONSHIP | |
| NOTARY (AFFIX SEAL) | SUBSCRIBED AND SWORN TO BEFORE ME ON | | SIGNATURE OF NOTARY NOTARY COMMISSION EXPIRES | | |
| ABSTRACT of Supporting Evidence (for health dept. use) 577 324 | DO NOT WRITE BELOW THIS LINE | | | | |
| | NAME AND KIND OF DOCUMENT (INCLUDING BY WHOM ISSUED AND DATE OF ISSUE) | | | DATE ORIGINAL DOCUMENT WAS MADE | |
| | 1 NY STATE DRIVER'S LICENSE RECORD 482536959394289-15 NEW YORK, NY | | | 11/22/1976 | |
| | 2 | | | | |
| | 3 | | | | |
| | INFORMATION CONCERNING REGISTRANT AS STATED IN DOCUMENT OF CORRESPONDING NUMBER ABOVE. | | | | |
| | 1 GEORGE L. EDWARDS DATE OF BIRTH 06/25/1915 | | | | |
| 2 | | | | | |
| 3 | | | | | |
| ADDITIONAL INFORMATION | | | | | |
| I certify that I have examined the documents referred to above, that they show no changes or erasures, and appear to be authentic. | | REGISTRAR <i>Thomas L. Davis</i> | EVIDENCE REVIEWED BY <i>Lillian M. Mount</i> | DATE FILED <i>11/13/07</i> | |

DHEC0613 (Rev. 01/1997)

METADATA MYTHCONCEPTIONS

```
<scdah:scdah xmlns:scdah="http://scdah.sc.gov/"
xmlns="http://scdah.sc.gov/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://scdah.sc.gov/
http://scdah.sc.gov/metadataSchema.xsd">
<title>00000001_00000001_1.pdf</title>
<identifier>1915-018179</identifier>
<creator>Dept. of Health and Environmental Control. --Public
Health Statistics and Information Services</creator>
<dates>
<created>1915-06-25</created>
<available>2016-01-01</available>
</dates>
<description></description>
<recordGroup>
<recordGroupTitle>Dept. of Health and Environmental
Control</recordGroupTitle>
<recordGroupIdentifier>169</recordGroupIdentifier>
</recordGroup>
<series>
<seriesTitle>Birth certificates</seriesTitle>
<seriesIdentifier>S169123</seriesIdentifier>
</series>
```

```
<type>Text</type>
<format>.pdf</format>
<accession>2016-103</accession>
<language>English</language>
<subject>GEORGE L EDWARDS</subject>
<subject>GEORGE EDWARDS</subject>
<subject>BAMBERG</subject>
<subject>ALMA RIVERS</subject>
<subject>AFFIDAVIT</subject>
<rights><rightsHolder>None</rightsHolder>
<accessRights>Public Domain</accessRights>
</rights>
</scdah:scdah>
```

DE-MISTIFYING THE METADATA

In theory the process to convert the delimited text file into individual XML files should be rather straight forward. However, reality is always a little more complicated than that though.

In total there are 12+ individual processes necessary to make the as submitted metadata file into a useable XML file.

DE-MISTIFYING THE METADATA



What did we use?

Excel

Access

Word

Notepad

Notepad ++

Gsplitter

Bulk Rename Utility

METADATA MASTERED

```
<scdah:scdah xmlns:scdah="http://scdah.sc.gov/"
  xmlns="http://scdah.sc.gov/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://scdah.sc.gov/
    http://scdah.sc.gov/metadataSchema.xsd">
<title>00000001_00000001_1.pdf</title>
<identifier>1915-018179</identifier>
<creator>Dept. of Health and Environmental Control. --Public
  Health Statistics and Information Services</creator>
<dates>
<created>1915-06-25</created>
<available>2016-01-01</available>
</dates>
<description></description>
<recordGroup>
<recordGroupTitle>Dept. of Health and Environmental
  Control</recordGroupTitle>
<recordGroupIdentifier>169</recordGroupIdentifier>
</recordGroup>
<series>
<seriesTitle>Birth certificates</seriesTitle>
<seriesIdentifier>S169123</seriesIdentifier>
</series>
```



| 07-06-15 M T.M.V. 05 | | | | Affidavit of Correction to Birth Record SOUTH CAROLINA DEPARTMENT OF HEALTH AND ENVIRONMENTAL CONTROL Page 2 of 2 | | | |
|---|--|---|--|---|--|--------------------------------|--|
| Enter Correlat Information Concerning Person Whose Birth Record is Being Amended | | | | REGISTRANT'S FULL NAME AT BIRTH GEORGE L EDWARDS | | | |
| | | | | STATE FILE OR BIRTH NUMBER 139-15-018179 | | | |
| BIRTH DATE Month Day Year JUNE 25 1915 | | BIRTH PLACE City or Town BAMBERG, SOUTH CAROLINA | | County | | State SOUTH CAROLINA | |
| ITEMS TO BE AMENDED OR CORRECTED | | | | ITEM OMITTED OR IN ERROR | | | |
| | | | | BIRTH CERTIFICATE SHOWS | | | |
| | | | | SHOULD BE | | | |
| CHILD'S GIVEN NAME | | | | OMITTED | | | |
| | | | | GEORGE L EDWARDS | | | |
| AFFIDAVIT | | | | I HEREBY DECLARE UPON OATH THAT THE ABOVE STATEMENTS ARE TRUE AND CORRECT. | | | |
| SIGNATURE OF PARENT (OR OTHER) | | | | <i>George L. Edwards</i> | | | |
| NOTARY (AFFIX SEAL) | | | | SUBSCRIBED AND SWORN TO BEFORE ME ON <i>See Attached</i> | | | |
| AFFIDAVIT | | | | I HEREBY DECLARE UPON OATH THAT THE ABOVE STATEMENTS ARE TRUE AND CORRECT. | | | |
| SIGNATURE OF PARENT (OR OTHER) | | | | SIGNATURE OF NOTARY | | | |
| NOTARY (AFFIX SEAL) | | | | SUBSCRIBED AND SWORN TO BEFORE ME ON | | | |
| | | | | SIGNATURE OF NOTARY | | | |
| | | | | NOTARY COMMISSION EXPIRES | | | |
| | | | | DO NOT WRITE BELOW THIS LINE | | | |
| ABSTRACT of Supporting Evidence (for health dept. use) <i>577 324</i> | | | | NAME AND KIND OF DOCUMENT (INCLUDING BY WHOM ISSUED AND DATE OF ISSUE) | | | |
| | | | | DATE ORIGINAL DOCUMENT WAS MADE | | | |
| | | | | 1 NY STATE DRIVER'S LICENSE RECORD 482536959394289-15 NEW YORK, NY 11/22/1976 | | | |
| | | | | INFORMATION CONCERNING REGISTRANT AS STATED IN DOCUMENT OF CORRESPONDING NUMBER ABOVE. | | | |
| | | | | 1 GEORGE L. EDWARDS DATE OF BIRTH 06/25/1915 | | | |
| | | | | 2 | | | |
| | | | | 3 | | | |
| | | | | ADDITIONAL INFORMATION | | | |
| | | | | I certify that I have examined the documents referred to above, that they show no changes or emendations, and appear to be authentic. | | | |
| | | | | EVIDENCE REVIEWED BY <i>Lillian M. Mount</i> | | | |
| | | | | DATE FILED <i>11/13/07</i> | | | |

DHEC 0613 (Rev. 01/1997)

METADATA MASTERED

=



One Happy
Archivist

LESSONS LEARNED

- Metadata may be voluminous, and it's best to work in discrete chunks
- You will mess up. Be prepared to start over again. And again.....
- Sometimes it's the data, not the metadata
- No matter how easy a transfer appears to be, it'll get complicated real fast

Thank you!!

Virginia Digital Chancery Project

- 18th-early 20th century court papers from ~120 Virginia localities
- Digitization began around 2000 (previously microfilmed)
- Extensive genealogical, local, national, and social history content
- Currently 70 localities digitized, totaling 10+ million images

Problem: Image Consistency/Work Flows

- Due to large volume of materials, contract initially awarded to two vendors
- After a few localities scanned, inconsistency in image quality became apparent
- Issues with proprietary software use, one obdurate vendor, and a high volume of re-work
- Workflows and processes largely grew out of doing the actual work
- This slowed down the project

Solution: Image Consistency

- All contracts were 1 year, renewed annually, up to 5 years
- Let contracts lapse for the two inconsistent vendors
- Used image data to justify sole-source
- Very good working relationship with current vendor for last 8 years
- Vendor works with other CHI, so they understand our objectives

Lessons Learned: Work Flows

- Some process documentation still lacking, but most worked out/written down
- Many more professional sources to review and projects to mimic now
- Lessons learned from the digital chancery project have informed other internal and external projects
- Be flexible, but develop as many of your processes as you can before taking on the project
- Time spent thinking before doing is time well-spent

This solution worked because...

Images

- We documented the vendor image problems
- We had a contractual “out”
- We had administrative support for a sole-source contract
- We found a consistent vendor who spoke our language

Work Flows

- Had to be flexible, recognize what we didn't know or hadn't thought of
- We had good communication between most players in the project
- Later federal grant applications forced us to delineate processes (to think about what we were doing)

If you try this, you should...

- Consult any and all best practices now available for similar projects
- Speak with others who have taken on like projects, especially if they are completed (better post-mortem)
- Select and convene the project team early and brainstorm not just the desired outcome but the processes
- Proposed processes should be reevaluated at various stages such as after adding new team members/work areas, piloting a project, hiring vendors, etc.

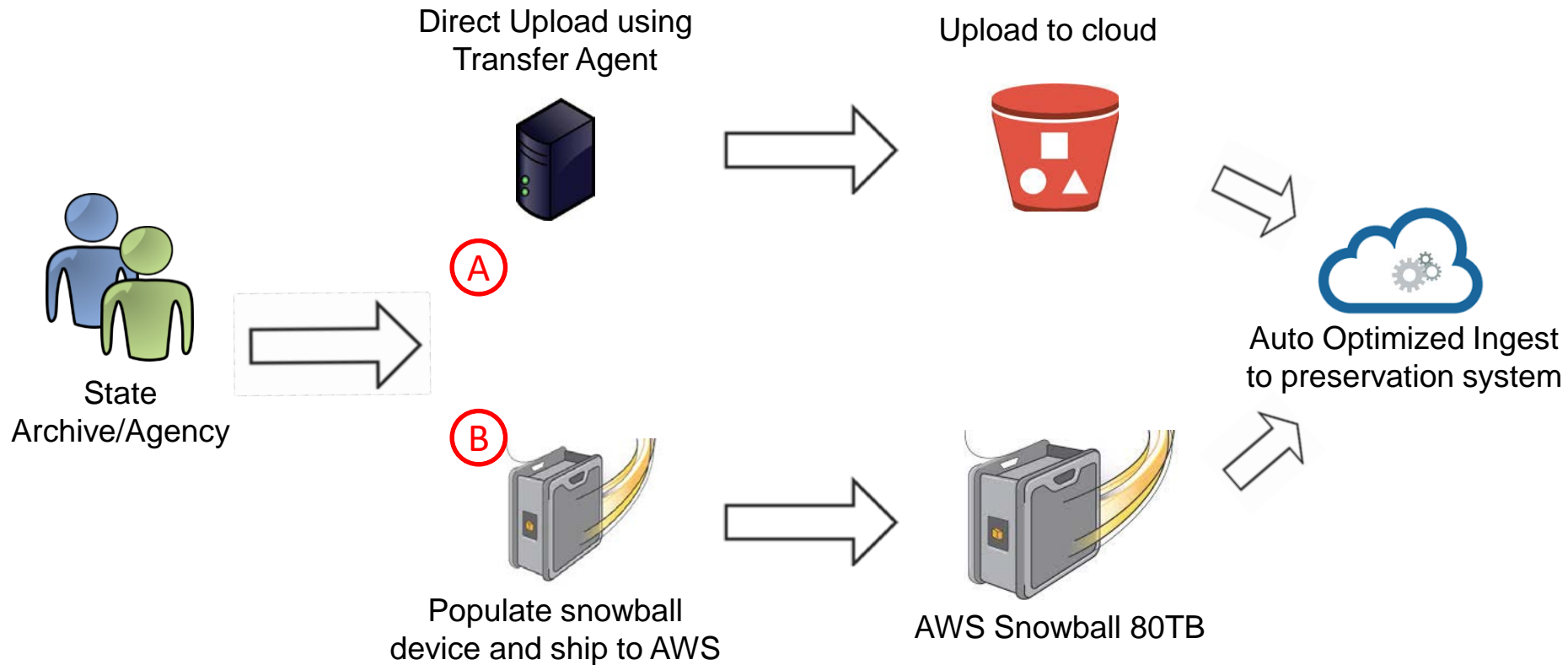
Digitization Project Output

- Multiple versions of the same file
(e.g. a high-resolution TIFF and a low-resolution JPEG)
- Simplify the preservation of digitized content - automatically bulk unpack and ingest the entire output
- Combine related sets of output files into a single record with associated metadata

Benefits of this approach

Save time and effort when compared to manually ingesting and preserving each file and file variant

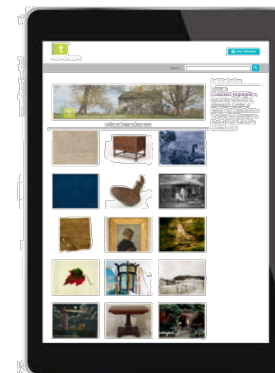
Joining Up Digitization & Preservation



Delivering Value with Public Access

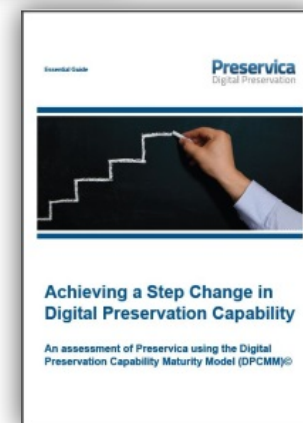
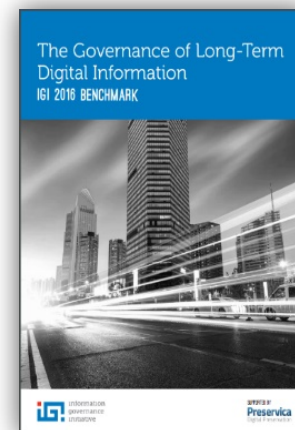


- Transparency via Access
 - Internal & external audiences
- Controlled access with search capability



Next Steps & Resources

- CoSA/Preservica Practical Digital Preservation Program:
<https://www.statearchivists.org/pertts/education-training/cosa-preservica-practical-digital-preservation/>
- Preservica Resources Center
<http://www.preservica.com/resources/>
- Preservica Webinars
 - Live Demo: May 11 @ 10am Eastern
Register at www.preservica.com
- Upcoming Events
 - CoSA: Boise, ID (10-13 July)
 - NAGARA: Boise, ID (12-15 July)
 - SAA: Portland, OR (23-29 July)



Questions?



Thank You!

www.preservica.com

info@preservica.com

@preservica

@dPreservation

www.statearchivists.org/