

# ELECTRONIC RECORDS INGEST

---

## Examples & Models

# PRESENTERS

---

## **Veronica Martzahl**

Massachusetts Archives  
Electronic Records Archivist

## **Mark Myers**

Texas State Library and Archives Commission  
Electronic Records Specialist

# ACKNOWLEDGEMENTS

---

This webinar is made possible by a grant from the National Historical Publications & Records Commission (NHPRC)



NATIONAL  
ARCHIVES

---

NATIONAL HISTORICAL  
PUBLICATIONS  
& RECORDS COMMISSION

# TERMS AND MA EXAMPLES

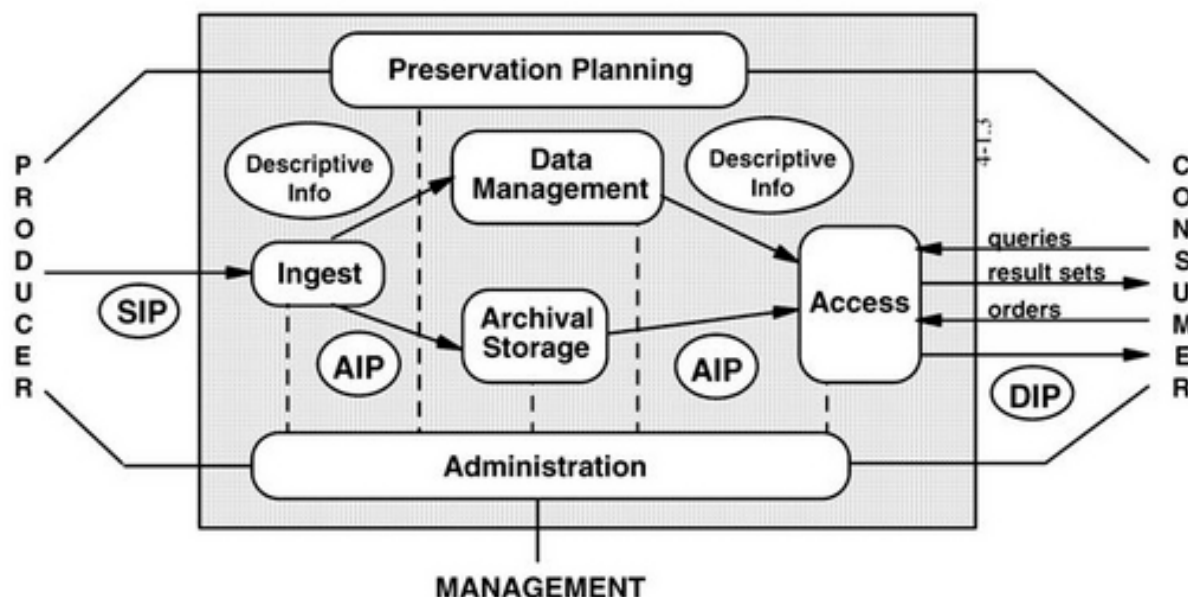
---

**Veronica Martzahl**

Massachusetts Archives

# INGEST

n. ~ In the Open Archival Information System (OAIS) model, processes related to receiving information from an external source and preparing it for storage.<sup>1</sup>



<sup>1</sup> Pearce-Moses, Richard. A glossary of archival and records terminology. "Ingest"  
<http://www2.archivists.org/glossary/terms/i/ingest>. Accessed September 8, 2014.

# REPOSITORY?

---

- **Your Archives**
- **Network Storage**
- **Surrogate Repositories**
- **Digital Preservation Repository**

# EXAMPLES OF DIGITAL PRESERVATION REPOSITORIES

---

**Preservica**  
Digital Preservation



**@archivematica**® DURACLOUD™



PROPYLON 

# PAIMIS AND PAIS

---

ISO 20652: Space Data and Information Transfer Systems -  
Producer-Archive Interface - Methodology Abstract Standard

Under development: PAIS – Producer-Archive Interface  
Specification

Guidance for establishing your submission agreement and getting  
the information you need for the Submission Ingest Package (SIP)  
you need for ingest.



# MASSACHUSETTS ARCHIVE

---

**Ingest = Getting it into our custody**

**Repository = Managed Network Storage**

## Steps

- Transfer
- Stabilization
- Monitoring

# TRANSFER

---

## Mechanism?

- **External Hard Drive**
- **FTP (File Transfer Protocol)**

## Requirements?

- **File types**
- **Metadata**

# STABILIZATION

---

- Virus scan
- Checksum(s)
- File migration
- Documentation
  - Read Me file
  - Basic Arrangement and Description

# LEVEL OF DESCRIPTION

---

1. Are the files restricted?
2. Is individual access to a file appropriate/needed for reference or other access services?
3. Are the files going into a repository with a public interface or are they going into a dark archive?
4. Is the office or collection high profile and/or does it have a history of intensive processing?

# MONITOR

---

- Make sure storage media is up to date
- Check checksums
- Review file formats

# TUFTS DIGITAL COLLECTIONS AND ARCHIVES

---

**Ingest = Into a Preservation Repository**

**Repository = Fedora**

## **Steps**

- Transfer
- Stabilization
- Arrangement and Description
- Monitoring

# TRANSFER

---

## Mechanism?

- **TAPER – Tufts Accessioning Program for Electronic Records**

## Requirements?

- **Transfer form**
- **Xythos dropbox**

# TRANSFER FORM

The screenshot shows a web browser window with the address bar displaying `dca.lib.tufts.edu/taper/`. The browser's address bar and tabs are visible at the top. The page header features the Tufts University logo on the left and the text "Digital Collections & Archives" in a large, light blue font. To the right of the header is a search bar with a "GO" button and radio buttons for "this site", "tufts.edu", and "people". Below the header, the page is divided into a left sidebar and a main content area. The sidebar, titled "TAPER Home", contains a message "You are not signed into TAPER." and a "Submission Tools" section with links for "Homepage" and "Transfer Agreement Form". The main content area features a large black and white photograph of three men in suits sitting at a table, looking at documents. Below the photograph, the text "Tufts Accessioning Program (TAPER) Home" is displayed, followed by the sentence "This is the homepage of all the TAPER tools." and two blue links: "Transfer Agreement Form" and "View Preservation Rules". At the bottom of the page, a footer contains a navigation bar with links: "Tufts | InsideTufts | Feedback | Directions | Find People | Contact Us" and a copyright notice: "Copyright © 2008 Tufts University".



# PRE-INGEST INFORMATION

---

Office of Origin

Authorized representatives

Creator

Producer

Email contact information

Record type

Dates

Copyright

Access

Formats

Arrangement and Description

Retention Period

Descriptive Standard

Respect de Fonds

Connection to dropbox

# STABILIZATION

---

- Virus scan
- Checksum(s)
- File migration

# ARRANGEMENT AND DESCRIPTION

---

## CIDER

- Archival Collection Management System
- Exports metadata to create FOXML object for FEDORA Repository
- Individual file vs. Aggregate description (Tufts Digital Library or Dark Archive)

# MONITORING

---

- Multiple FEDORA instances
- CIDER metadata

# DEVELOPMENTS

---

- **Hydra administrative interface to streamline the ingest process**
- **Admin interface also allows other library professions to do some of the metadata work for the archives (ie subject cataloging of Senior Theses)**

# IN SUMMARY

---

## ■ Goals

- Get the files into our custody
- Document, document, document
- Keep the files safe, regardless of the environment we are working in
  - Able to prove what you have is what you received
  - Able to show no deterioration has occurred
  - Able to find and access the files

# KENTUCKY AND TEXAS EXAMPLES

---

**Mark Myers**

Texas State Library and Archives Commission



# KDLA E-ARCHIVE

---

**Ingest = Locating the records & Getting them into the repository**

**Repository =**

- 1996-2008 = Managed Network Storage &
- 2008 = 2012 = Dspace &
- 2012-Present = Preservica

**Steps =**

- Acquisition (locating the records)
- Transfer
- Arrangement & Description
- Monitoring = Preservation



# ACQUISITION METHOD #1: DIG IT UP YOURSELF!

---

## Harvesting from agency websites.

- Manual process
- Have to hunt around on site
- Prioritized list of agencies
  - Outgrowth of state publications program
  - Followed retention schedule process
- Acquired tools along the way

# WEB HARVESTING/COLLECTING TOOLS



## Grab-A-Site (Blue Squirrel)

- <http://www.bluesquirrel.com/products/grabasite/>
- Pros:
  - Allowed for crawling an entire website and grabbing files of a specific type (.doc, .jpg, .pdf)
  - Also used to archive whole websites (Governor, Lt. Gov., First Lady)
- Issues:
  - Not automated
  - No metadata – websites captured as HTML and other formats
  - Not sure that it's still being maintained. Version 5.0 has been out for over 10 years.



## Archive-It ([www.archive-it.org](http://www.archive-it.org))

- KY - Became full partner in 2010; TX - partner since 2007
- Pros:
  - Allows for AUTOMATED harvesting and metadata addition/collection
  - Can navigate the morass that is the Internet: Content Mgmt systems, Social Media, etc
- Issues:
  - Can't harvest for just objects (pdf, doc, jpg) – what if you don't want the entire website?
  - Records are located on Archive-It site



## Preservica (<http://preservica.com/>)

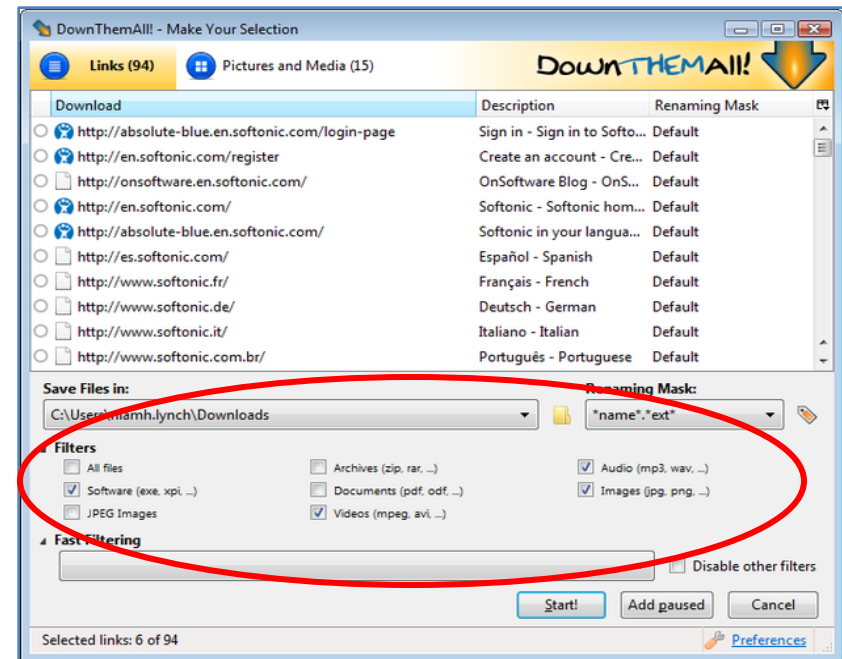
- Allows for web harvesting directly into repository
  - New feature in current version of Cloud Edition
  - Also allows for capture from known FTP sites (Enterprise Edition)
- Uses Heritrix web crawler
- Issues:
  - Same as Archive-It – records are wrapped up in WARC file



# WEB HARVESTING/COLLECTING TOOLS

## Download Them All

- Browser extension for
  - Firefox: <https://addons.mozilla.org/en-US/firefox/addon/downthemall/>
- Allows for grabbing all the files of a certain type off a single page
- Filter by file type
  - Documents (pdf, odf, doc)
  - Customizable – can add to filters
- Allows for renaming
- Issues:
  - Single page
  - No metadata, just object

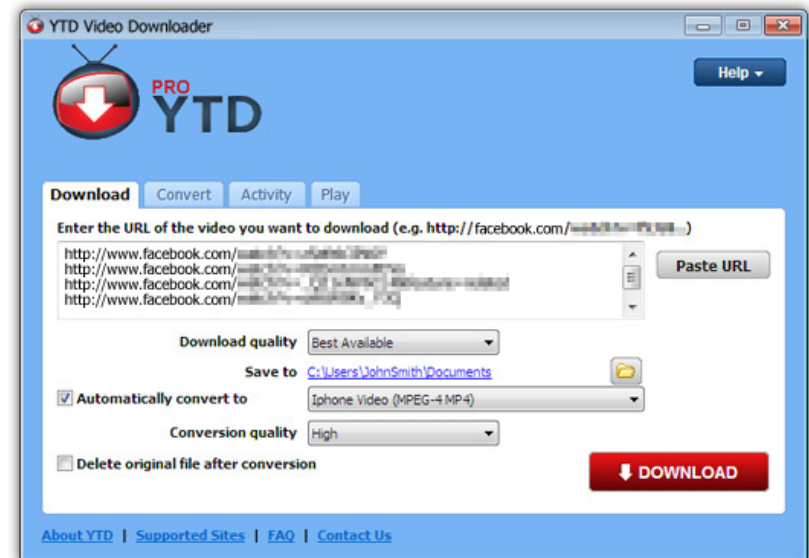


# WEB HARVESTING/COLLECTING TOOLS

## YTD Video Downloader

[http://download.cnet.com/YTD-Video-Downloader/3000-2071\\_4-10647340.html](http://download.cnet.com/YTD-Video-Downloader/3000-2071_4-10647340.html)

- Allows for identification and capture of media files on a page
- Download and convert audio/video to other formats
- Allows for starting/stopping
- Freeware & Paid (annual subscription)
  - Auto-conversion is a PRO feature
  - Free version will convert from already downloaded files
- Issues:
  - Single page, not whole site
  - No metadata, just object
  - Legality???



# ACQUISITION METHOD 2: Talk to Agencies ( or “Records Producers” in OAI-speak)

---

## Communicate with agencies:

- Ask for electronic records!
- Coordinate harvesting
  - Tell us where records are
  - Remove impediments to harvesting
  - Tell us when records will be taken down (preferably before)
- Coordinate transfers
  - Send us the records
  - Policies/Procedures
  - Method
  - Metadata

# ACQUISITION METHOD 3:

## Make (get) them do it for us

---

- **Direct deposit into DSpace**
  - Good for single/small transfers
  - Homogenous documents
    - Serials
    - Minutes
    - Pubs
  - Get's the creator to add metadata through templates
  - Archivist can approve/reject/edit
  - Issues:
    - Never quite got off ground
    - Not good for large transfers

# TRANSFER 1 - INSTRUCTIONS

---

## Instructions = Policies & Procedures

- **New Administrative Regulation for State Publications**
  - Allow for electronic transfer (instead of paper)
  - Mandates PDF format (state standard for published material)
- **Records Transfer Guidance**
  - Coupled with paper process
  - Outlined the methods
  - Outlined “desired” formats

<http://kdla.ky.gov/records/Documents/Records%20Transfer%20Procedures,January,%202010.PDF>

# Recommended Data Formats for Preservation Purposes in the KDLA Digital Archive

<http://kdla.ky.gov/records/recmgmtguidance/Documents/File%20Format%20table.PDF>

Media	High Confidence	Medium Confidence Level	Low Confidence Level	Notes/Comment
Text	<ul style="list-style-type: none"> <li>- Plain text (encoding: US ASCII, UTF-8, UTF-16 with BOM)</li> <li>- PDF/A-1 (*.pdf)</li> <li>- XML (XSD/XSL/XHTML, etc.; with included or accessible schema and character encoding explicitly specified)</li> </ul>	<ul style="list-style-type: none"> <li>- Plain text (ISO8859-1 encoding)</li> <li>- PDF (*.pdf) (embedded fonts)</li> <li>- Rich Text Format (*.rtf) version 1.x</li> <li>- OpenOffice (*.sxw)</li> <li>- <b>Microsoft Word (*.doc)*</b></li> <li>- WordPerfect (*.wpd)#</li> <li>- HTML 4.x (include a DOCTYPE declaration)</li> <li>- SGML</li> </ul>	<ul style="list-style-type: none"> <li>- PDF (external font)</li> <li>- All other text formats not listed here</li> <li>- <u>DiVu</u> (alternative format to PDF. Uses a different compression to make a smaller file. Published standard. Created by AT&amp;T; owned by producers of <u>MrSID</u> GIS format. – Used by USGS and other GIS and Washington State</li> </ul>	
Raster Image	<ul style="list-style-type: none"> <li>- TIFF (uncompressed)</li> <li>- PNG (*.png)</li> <li>- JPEG (raw)?</li> </ul>	<ul style="list-style-type: none"> <li>- BMP (*.bmp)</li> <li>- <b>JPEG/JFIF (*.jpg)</b></li> <li>- JPEG2000 (prefer uncompressed) (*.jp2, *.jpx)</li> <li>- <b>TIFF (CCITT Group 3/4, JPEG, PackBits compression)</b></li> </ul>	<ul style="list-style-type: none"> <li>- <u>MrSID</u> (*.sid)</li> <li>- TIFF (with LZW compression or in Planar format)</li> <li>- GIF (*.gif)</li> <li>- <u>FlashPix</u></li> <li>- <u>PhotoShop</u> (*.psd)</li> <li>- All other raster image formats not listed here</li> </ul>	<ul style="list-style-type: none"> <li>- "Raw" JPEG are those images that have not been resized.</li> <li>- Depends on compression format.</li> <li>- Uncompressed</li> </ul>
Audio	<ul style="list-style-type: none"> <li>- AIFF (uncompressed) (*.aif, *.aiff)</li> <li>- WAVE (LPCM only) (*.wav)</li> </ul>	<ul style="list-style-type: none"> <li>- Standard MIDI (*.mid, *.midi)</li> <li>- Windows Media Audio (*.wma) *</li> <li>- MP3 (MPEG 1/2, Layer 3) (8.mp3)</li> <li>- SUN Audio (uncompressed)</li> </ul>	<ul style="list-style-type: none"> <li>- <u>AIFC</u> (*.aifc)</li> <li>- <u>NeXT SND</u> (*.snd)</li> <li>- <u>RealNetworks</u> 'Real Audio' (8.ra, *.rm, *.ram)</li> <li>- WAVE (compressed) (*.wav)</li> </ul>	<ul style="list-style-type: none"> <li>- MP3 is a non-documented compressed version of MPEG</li> <li>- the bare MPEG is open (v. 1 &amp; 2 are</li> </ul>



# TRANSFER 2 – METHODS

---

- **Email**

- Good for small number of files (attachment)
- Limitation by (total) file size

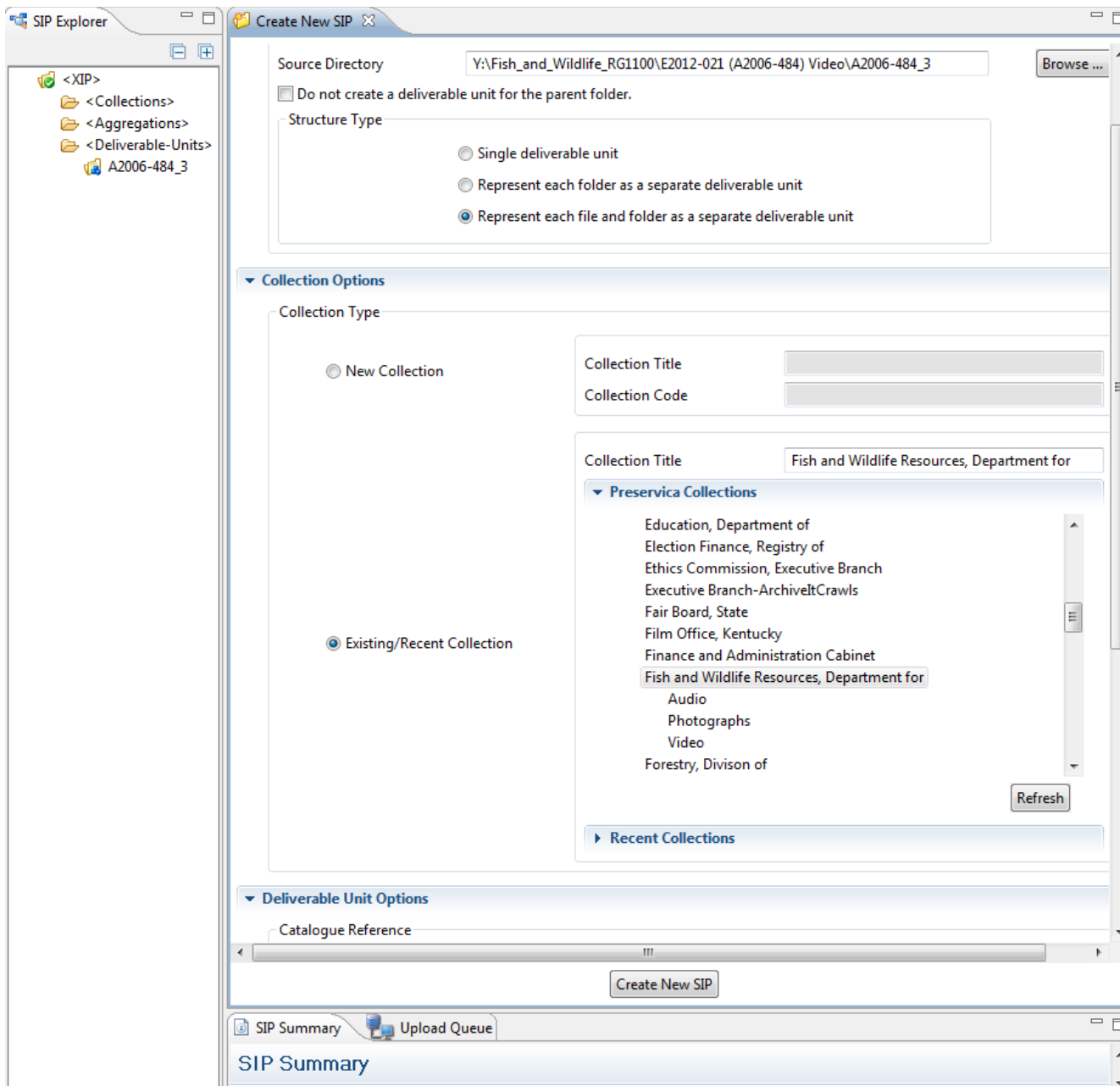
- **On Media (CD/DVD, Flash Drives, External Hard Drive)**

- **FTP (File Transfer Protocol) or Website**

- (FTP) Good for incredibly large files
- (Website) Good for large number of files already out there

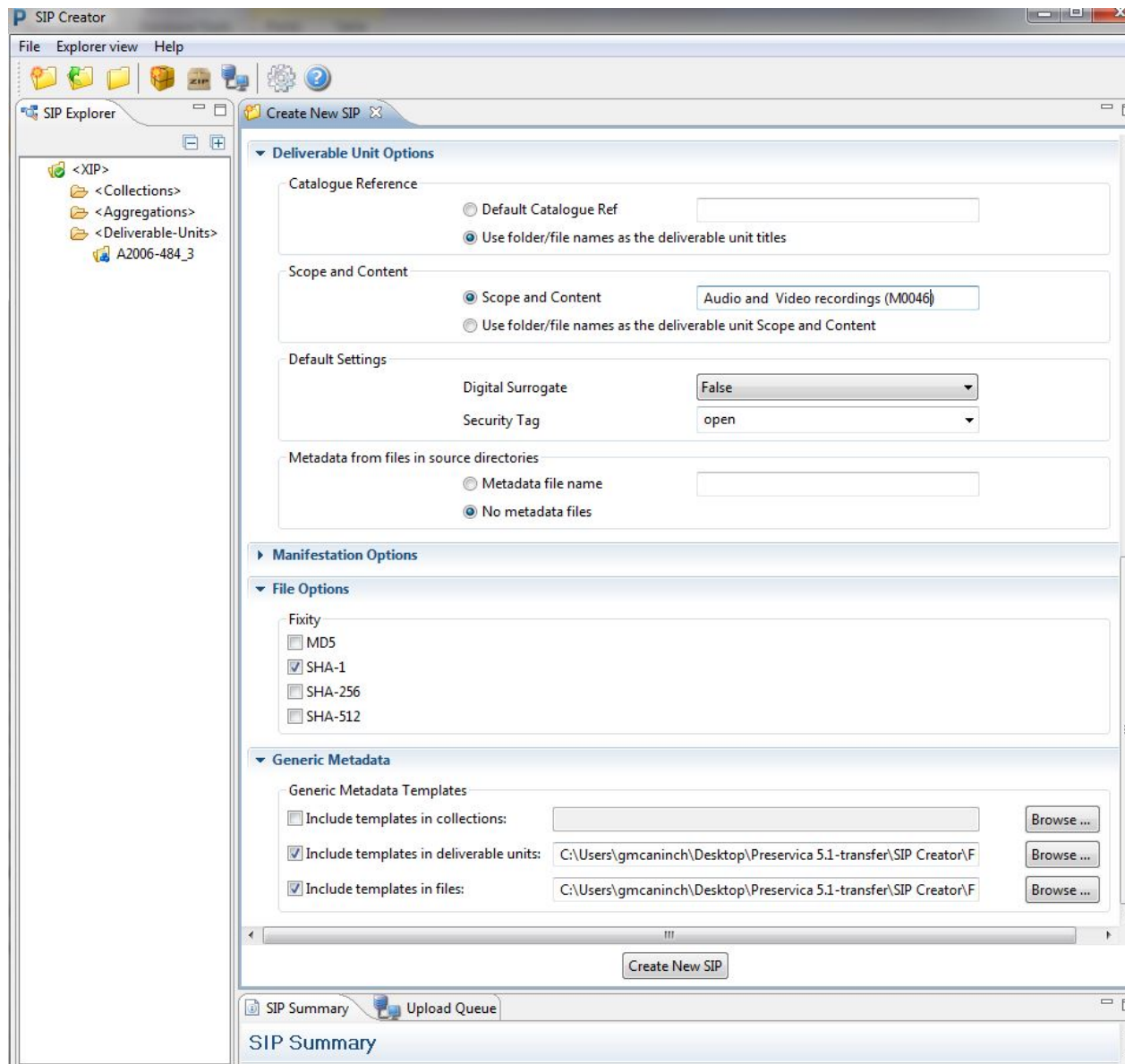
- **Direct Deposit**

- Across the LAN (GIS Snapshots)
- DSpace



# Preservica SIP Creator

- Used to ingest files into Preservica
- Allows for pulling together a SIP package



# Preservica SIP Creator

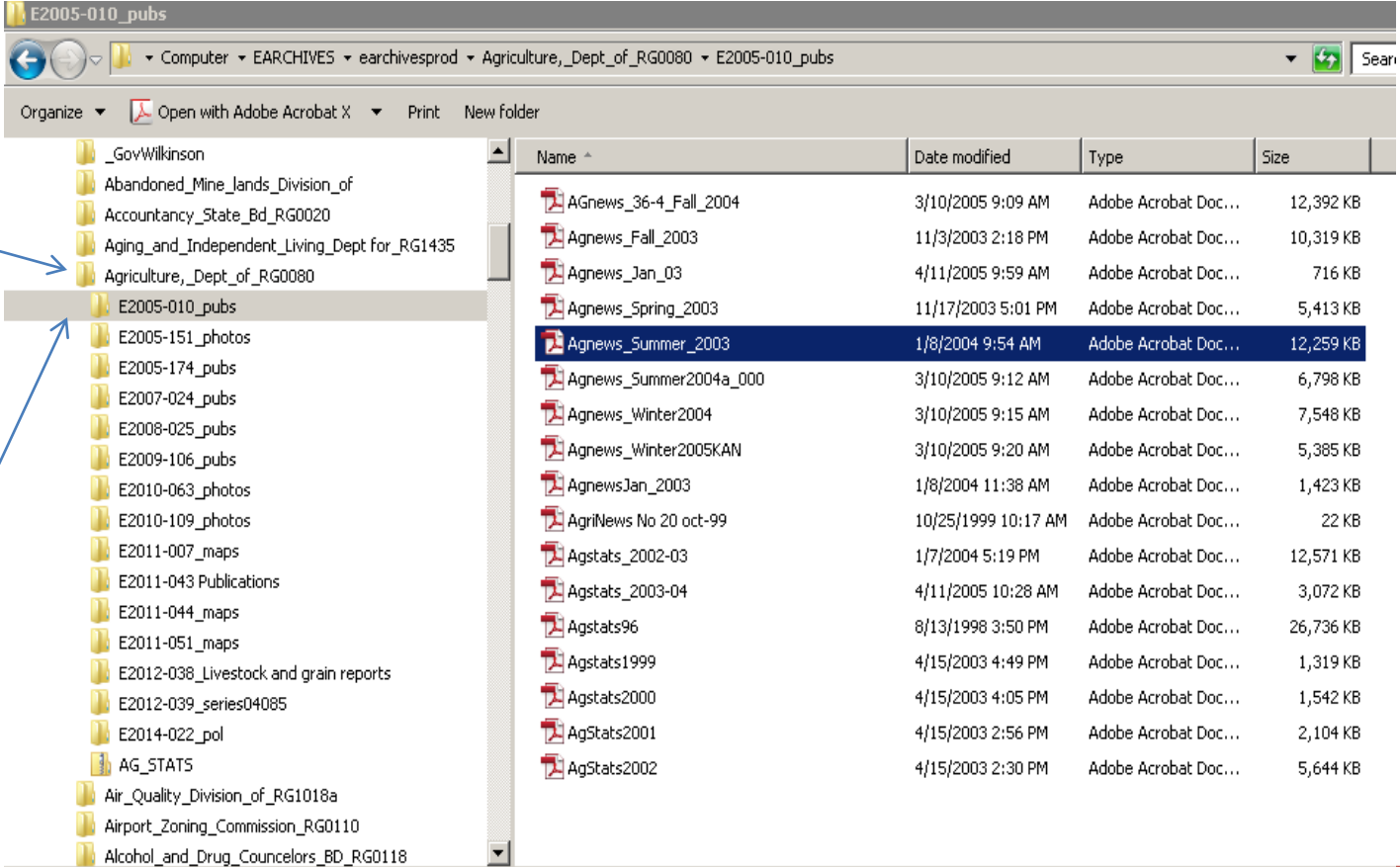
- Apply Integrity Checks
- Attach metadata to the package

# ARRANGE & DESCRIBE:

## Place files on file server (local storage)

By Agency

By Accession (SIP)



Name	Date modified	Type	Size
Agnews_36-4_Fall_2004	3/10/2005 9:09 AM	Adobe Acrobat Doc...	12,392 KB
Agnews_Fall_2003	11/3/2003 2:18 PM	Adobe Acrobat Doc...	10,319 KB
Agnews_Jan_03	4/11/2005 9:59 AM	Adobe Acrobat Doc...	716 KB
Agnews_Spring_2003	11/17/2003 5:01 PM	Adobe Acrobat Doc...	5,413 KB
Agnews_Summer_2003	1/8/2004 9:54 AM	Adobe Acrobat Doc...	12,259 KB
Agnews_Summer2004a_000	3/10/2005 9:12 AM	Adobe Acrobat Doc...	6,798 KB
Agnews_Winter2004	3/10/2005 9:15 AM	Adobe Acrobat Doc...	7,548 KB
Agnews_Winter2005KAN	3/10/2005 9:20 AM	Adobe Acrobat Doc...	5,385 KB
AgnewsJan_2003	1/8/2004 11:38 AM	Adobe Acrobat Doc...	1,423 KB
AgriNews No 20 oct-99	10/25/1999 10:17 AM	Adobe Acrobat Doc...	22 KB
Agstats_2002-03	1/7/2004 5:19 PM	Adobe Acrobat Doc...	12,571 KB
Agstats_2003-04	4/11/2005 10:28 AM	Adobe Acrobat Doc...	3,072 KB
Agstats96	8/13/1998 3:50 PM	Adobe Acrobat Doc...	26,736 KB
Agstats1999	4/15/2003 4:49 PM	Adobe Acrobat Doc...	1,319 KB
Agstats2000	4/15/2003 4:05 PM	Adobe Acrobat Doc...	1,542 KB
AgStats2001	4/15/2003 2:56 PM	Adobe Acrobat Doc...	2,104 KB
AgStats2002	4/15/2003 2:30 PM	Adobe Acrobat Doc...	5,644 KB

# Accession metadata!!!!

Microsoft Access

File Home Create External Data Database Tools

Views Clipboard Sort & Filter Records Find Window Text Formatting

Electronic records Accessions

GN1245b Cabinet:  AGENCY: Governor's Office UNIT: Budget Director, Office of State

Unit not in Record Group:

Method of Transfer: Optical Disk Download site:

File Characteristics Total number of files: 362 Accession Date: 6/19/2014 Accession Number: E2014-031

Series	Disk	Folder Name	File total	Format	Span date
00327	1	Budget requests: Education & Workforce	53	PDF	2014
00327	1	Budget requests: Energy and Environment	2	PDF	2014
00327	1	Budget requests: Finance	11	PDF	2014
00327	1	Budget requests: General Government	73	PDF	2014

Record: 14 of 16 No Filter Search

Volume in MB: 1000

Acknowledged ☐

Technical Contact:  Technical e-mail:  Technical phone:

Records officer e-mail:  Rec officer phone:

Form completed by: Mcaninch Date form completed: 6/19/2014 Schedule date: December 2010

Adobe Portable Document Format - File specifications ☐ Security Deactivated ☐ Fonts embedded

pdf versions:

Scanned Text, Photo and Video File Specifications Location: e-archivesprod\Budget\_Office\_of\_the\_RG1245b

Compression TIF:  Image resolution:  Location alt:

Color compression: (JPEG or video):  JPEG version:

Indexing:  Copyright:

Bit rate:

Navigation Pane

Complete  
Accession  
Form

# KDLA'S ELECTRONIC RECORDS ACCESSION FORM

---

## Metadata captured:

- **Agency Info** - Includes Record Group number which links out to other (paper) accession databases
- **Accession number** – “E” (for electronic)/year/000-999
- **Method of Transfer** – Optical disk, Email, Download from website, FTP. Download site url if applicable
- **Total number of files**
- **Volume in MB**
- **For each series:**
  - Series Number
  - Disk (1 of x)
  - Folder name
  - Number of Files
  - Format (PDF, DOC, JPG, TIF)
  - Date span
- **Technical Contact info** – person responsible for electronic record
- **Records Officer**
- **Other technical/preservation metadata based on file type**
  - Only PDF and Digital images

# KDLA'S ELECTRONIC RECORDS ACCESSION FORM

---

## Issues:

- All manual – not automated
  - Only internal – we have to fill out
  - Manual form on website, but staff still have to enter into database
  - No extraction of metadata from files/objects
- Doesn't capture file names – only total number of files
  - No fixity information (checksums)
  - Only sight verification – open folders and count number of files
- Mixed media an issue – folders with multiple file types
- Many times don't have technical contact info
- Technical/preservation information is often incomplete
  - Again, manual – no validation/extraction of technical information
- Mirrors Paper system

# ARRANGEMENT & DESCRIPTION: Enter DSpace

[e-Archives Home](#) → [Agency List](#)

## Agencies in the e-Archives

Select an agency to browse its series.

- **[Agriculture, Department of \(Commissioner of Agriculture\) \[RG 0080\]](#)**
  - [Annual or Summary Reports \(M0022\) - Agriculture](#)
  - [Maps \(M0051\) - Agriculture](#)
  - [Photographic File - State Agency Activities \(M0047\) - Agriculture](#)
  - [Publications Created By or For Agency \(M0007\) - Agriculture](#)
  - **[Market News Branch](#)**
    - [Weekly Livestock and Grain Market Report \(04085\)](#)
- **[Auditor of Public Accounts \[RG 0240\]](#)**
  - [Audits of State Agencies \(04053\)](#)
  - [Performance Audits \(04054\)](#)
  - [Policies and Procedures \(M0003\) - Auditor](#)
  - [Publications Created By or For Agency \(M0007\) - Auditor of Public Accounts](#)
  - [Special Examination Reports \(05010\)](#)
- **[Court of Justice \(Judicial Branch\)](#)**
  - **[Administrative Office of the Courts \[RG 0640a\]](#)**
    - [Annual or Summary Reports \(M0022\) - Administrative Office of the Courts](#)
    - [Maps \(M0051\) - Courts](#)
    - [Publications Created By or For Agency \(M0007\) - Administrative Office of the Courts](#)
  - **[Kentucky Bar Association \[RG 0335\]](#)**
    - [Publications Created By or For Agency \(M0007\) - Kentucky Bar Association](#)
  - **[Racial Fairness Commission](#)**
    - [Publications Created by or for Agency \(M0007\) - Racial Fairness Commission](#)
  - **[Supreme Court of Kentucky \[RG 2687\]](#)**
    - [Minutes of Meetings \(M0008\) - Supreme Court of Kentucky](#)
- **[Economic Development \(Cabinet for\) \[RG 0737\]](#)**
  - [Annual or Summary Reports \(M0022\) - Economic Development](#)
  - [Maps \(M0051\) - Economic Development](#)
  - [Policies and Procedures \(M0003\) - Economic Development](#)
  - [Publications Created By or For Agency \(M0007\) - Economic Development](#)
  - **[Bluegrass State Skills Corporation \[RG 0412\]](#)**
    - [Annual or Summary Reports \(M0022\) - Bluegrass State Skills Corporation](#)
    - [Publications Created By or For Agency \(M0007\) - Bluegrass State Skills Corporation](#)
- **[Office of Research and Information Technology](#)**
  - [Maps \(M0051\) - Office of Research and Information Technology](#)

### Search e-Archives

[Advanced Search](#)

### Browse

All of e-Archives  
[Agencies & Series](#)  
[Date](#)  
[Creator](#)  
[Titles](#)  
[Subjects](#)

### My Account

[Login](#)  
[Register](#)



## Cropland Data Layer Maps for the Commonwealth of Kentucky [2008]

Agriculture, Kentucky Department of; U.S. National Agricultural Statistical Service

URI: <http://hdl.handle.net/10602/11466>

<http://dlaasnp1.state.ky.us:8080/xmlui/handle/10602/11466>

Date: 2008

### Abstract:

Downloaded from the U.S. National Agricultural Statistical Services in KML and PDF format

[Show full item record](#)

### Files in this item



Name: Cropland Data Layer ...

Size: 6.798Mb

Format: PDF

Description: pdf version

[View/Open](#)



Name: CropScape\_CDL\_130 ...

Size: 995bytes

Format: Unknown

Description: KML version

[View/Open](#)

### This item appears in the following Series

- [Maps \(M0051\) - Agriculture](#)

### Search e-Archives

- ☒ Search e-Archives  
☐ This Series

[Advanced Search](#)

### Browse

All of e-Archives

[Agencies & Series](#)

[Date](#)

[Creator](#)

## Dspace Simple Item Record



### Show simple item record

dc.contributor.author	Agriculture, Kentucky Department of	
dc.contributor.author	U.S. National Agricultural Statistical Service	
dc.date.accessioned	2012-01-03T20:51:39Z	
dc.date.available	2012-01-03T20:51:39Z	
dc.date.issued	2008	
dc.identifier.other	E2011-044	
dc.identifier.uri	<a href="http://hdl.handle.net/10602/11466">http://hdl.handle.net/10602/11466</a>	
dc.identifier.uri	<a href="http://dlaasnp1.state.ky.us:8080/xmlui/handle/10602/11466">http://dlaasnp1.state.ky.us:8080/xmlui/handle/10602/11466</a>	
dc.description.abstract	Downloaded from the U.S. National Agricultural Statistical Services in KML and PDF format	
dc.language.iso	en_US	en_US
dc.subject.other	Geospatial	
dc.title	Cropland Data Layer Maps for the Commonwealth of Kentucky [2008]	en_US

### Search e-Archives

- ☒ Search e-Archives  
☐ This Series

[Advanced Search](#)

### Browse

All of e-Archives

[Agencies & Series](#)

[Date](#)

[Creator](#)

[Titles](#)

[Subjects](#)

This Series

[Date](#)

[Creator](#)

[Titles](#)

[Subjects](#)

### My Account

[Login](#)

[Register](#)

## Dspace Full Item Record



### Files in this item



Name: Cropland Data Layer ...

Size: 6.798Mb

Format: PDF

Description: pdf version

[View/Open](#)



Name: CropScape\_CDL\_130 ...

Size: 995bytes

Format: Unknown

Description: KML version

[View/Open](#)

### This item appears in the following Series

- [Maps \(M0051\) - Agriculture](#)

[Show simple item record](#)

# MONITOR = PRESERVATION

Workflow Context	Web Harvest Workflow
Workflow Definition	Heritrix Ingest Workflow (Full)
Workflow ID	299
Workflow State	Completed
Date Started	31.07.12 19:41:46
Date Finished	31.07.12 19:57:22
Number of Files	1
Total Size	696 KB
Collection Code	code
Submission name	State government organization 2006
Top Level Record	State government organization 2006

[Back](#)

## Step Progress

State	Name	Progress	Started	Finished	Messages
	SelectUrl	<div></div>	31.07.12 19:41:46	31.07.12 19:55:55	
	ConfigureCrawl	<div></div>	31.07.12 19:55:55	31.07.12 19:56:17	
	Url Crawler	<div></div>	31.07.12 19:56:18	31.07.12 19:56:37	
	Create Website XIP	<div></div>	31.07.12 19:56:37	31.07.12 19:56:40	
	Virus Check	<div></div>	31.07.12 19:56:40	31.07.12 19:56:43	<a href="#">View</a>
	Fixity Check	<div></div>	31.07.12 19:56:43	31.07.12 19:56:46	
	Metadata Integrity	<div></div>	31.07.12 19:56:46	31.07.12 19:56:49	
	Content Integrity	<div></div>	31.07.12 19:56:49	31.07.12 19:56:52	
	Characterise	<div></div>	31.07.12 19:56:52	31.07.12 19:57:01	<a href="#">View</a>
	Store Files	<div></div>	31.07.12 19:57:01	31.07.12 19:57:04	
	Store Metadata	<div></div>	31.07.12 19:57:04	31.07.12 19:57:07	
	Store Metadata File	<div></div>	31.07.12 19:57:07	31.07.12 19:57:10	
	Update Search Index	<div></div>	31.07.12 19:57:10	31.07.12 19:57:13	
	Thumbnail Creation	<div></div>	31.07.12 19:57:13	31.07.12 19:57:22	

## Preservica preservation Processes

- Virus Scan
- Characterization and Format validation
- Normalization & Creation of DIP's

Original and  
Derivative managed  
together

- Integrity Checks & Monitoring
- Migration actions

# SUMMARY

---

- **Main Points:**

- Find the records
- Bring them under control = Instruction/Guidance
- Apply appropriate metadata
  - Accession info
  - Administrative
  - Preservation/technical
- Make the records accessible (arrangement, description)
- Maintenance/Preservation

# SUMMARY

---

- **Finding the records – communication is the key**
  - Know where the records are
  - How to get the records to archive
- **Automate the process as much as possible**
  - Metadata capture
  - Harvesting
  - Making systems talk to each other

# SUMMARY

---

- **Keep it Simple**

- Agencies aren't going to repackage, rename, rearrange, use special tools/process

- **Keep it secure**

- But – need to ensure integrity through the transfer process
- Make sure you received what was sent
- Objects haven't changed

# CONTACT INFORMATION

---

**Veronica Martzahl**

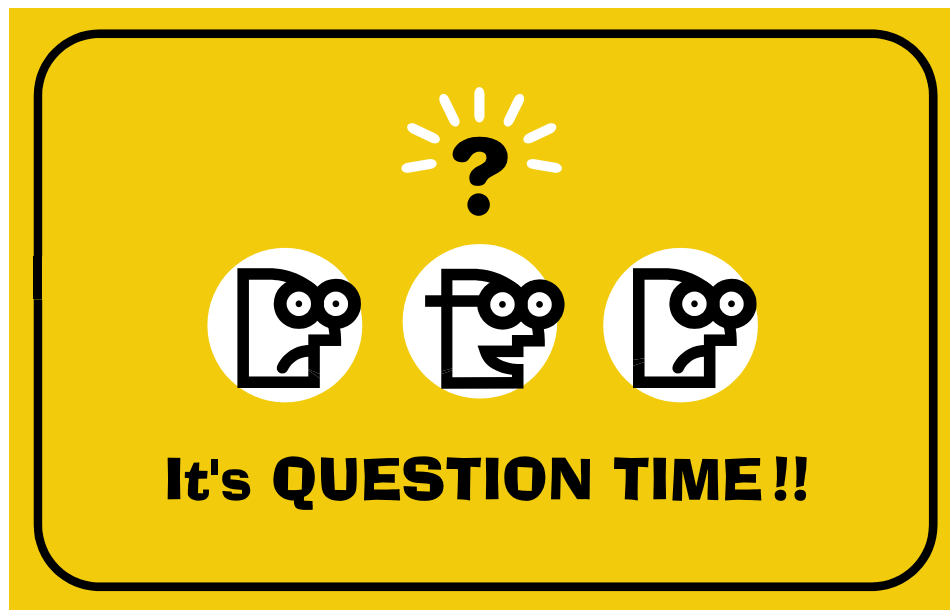
[veronica.martzahl@sec.state.ma.us](mailto:veronica.martzahl@sec.state.ma.us)

**Mark Myers**

[mmyers@tsl.texas.gov](mailto:mmyers@tsl.texas.gov)

# QUESTIONS & COMMENTS

---



# WRAP-UP

---

- Post-webinar evaluation will automatically open in your web browser when you exit the session.
- Next SERI Educational Webinar is November 5 @ 2:00pm
  - Topic: Electronic Records Storage
- Complete webinar schedule is available on CoSA's website:  
[http://www.statearchivists.org/CoSA\\_Webinars.htm](http://www.statearchivists.org/CoSA_Webinars.htm)
- All webinar slides available from the SERI webinar page:  
[http://www.statearchivists.org/seri/STEP/SERI\\_Educational\\_Webinars.htm](http://www.statearchivists.org/seri/STEP/SERI_Educational_Webinars.htm)