



Protecting the Bits: Preserving your digital records

Mark Myers,
Senior Electronic Records Specialist,
Texas State Library and Archives
Commission

Texas Digital Archive (TDA)



- Created in 2015
- Repository for electronic archival state records held by TSLAC
- Preserves and provides access to electronic records



The Texas Digital Archive (TDA) manages, preserves, and facilitates access to the electronic records collections of the Texas State Library and Archives Commission, including those transferred by State agencies or digitized by the State Archives. **All records visible in this portal are unrestricted and available for public use.**

Click here for tips on [Locating records in the TDA](#)

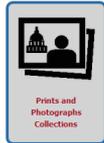
Collections



Governor's Records --
Rick Perry,
2000-2014



Manuscripts
Collections



Prints and
Photographs
Collections



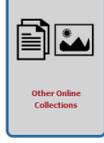
Texas Legislature



Texas State Agencies



Sam Houston
Regional Library and
Research Center



Other Online
Collections



Online Exhibits



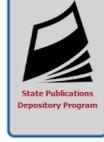
Special Subject
Interests



Artifacts Collection



Texas Judiciary



State Publications
Depository Program

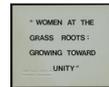
Featured Materials



An early group of women
registrants. From the
William Deming
Howaday collection



Erminia Thompson
Folsom Lecture on
Women's Suffrage



"Women at the grass
Roots: Growing toward
unity" sign

For more information concerning our holdings, including materials not available online, visit the [Archives & Reference](#) page.

To return to the TSLAC main website, please [click here](#).

When using images and other records from the Texas Digital Archive for publication or display, please refer to the citation information found in the item's metadata or in the descriptive guide for that collection. This statement follows the form: "Courtesy of Texas State Library and Archives Commission (and cite the collection name and record title or image number)."

Currently the TDA has:

- ~ 8.3 million records
- < 60 TB
- Born-digital & Digitized material
- Records from all branches of TX state government
- Wide variety of file types:
 - Text
 - Photos
 - Audio
 - Video & Film
 - Databases
 - Artifacts (kind of)

TSLAC's Mission

- Preserve the record of government for public scrutiny
- Secure and make accessible historically significant records and other valuable resources, both for print and electronic documents



Protecting

Electronic State Records

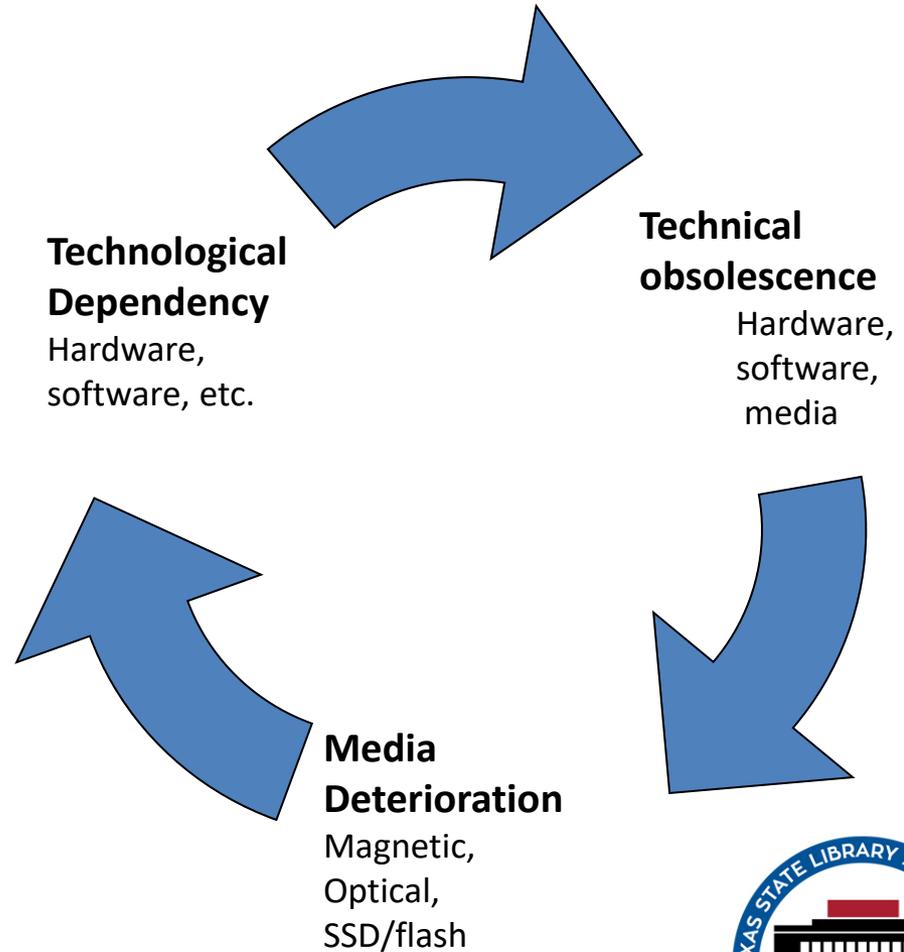
Using “protecting” to mean:

- Securing
- Preserving
- Making them accessible



Preservation Challenges

- Unlike some paper records, digital records do not survive without constant attention
- Biggest challenge facing electronic records - **CHANGE**



When Electronic Records Are at Risk

- Software is obsolete or getting to end of life
- Dependency on specific hardware or system
- Non-standard/specialty formats used
- Systems being replaced
- Records no longer in use
- Records are on removable media



Get control over your records!

- Starts with Records Management
 - Records Retention Schedules
 - Identify what records you have and what purpose they serve
 - Establish appropriate retention periods to manage volume
 - Dispose of records at the right time
 - Protect records that need to be retained
 - Recordkeeping systems
 - Organize and Categorize Records
 - Centralized control
 - Managed Access and storage
 - Policies and Procedures
 - People know how to use the system
 - Train people on the culture of management



Securing records in the TDA



Step one – getting the records



What do we take?

- The same kind of records that you transfer in paper
 - Records with an “A” designation on the schedule
 - Records with an “R” designation (need to be reviewed by an archivist prior to transfer)
- After active administrative use ends
- “Original” format – especially with images, audio, video
 - Or the “official” or “published” for text-based records



What do we take?

- Can take records in any format, however:
 - Proprietary formats that require special/expensive license
 - Encrypted records – other forms of DRM
 - Older “legacy” formats
 - Legacy media
- We need to talk first



Transferring Electronic Records to the Texas State Archives

The Texas State Archives *is now accepting transfers* of electronic records in the Texas Digital Archive (TDA). The TDA is a repository built around international standards for the long-term preservation of electronic records.

Benefits of the Texas Digital Archive

- Saves taxpayer money. By transferring archival electronic records to the TDA, agencies will reduce their data storage and migration costs
- Reduces agency staff time spent responding to Public Information Act requests
- Monitors and ensures the authenticity, integrity, and security of electronic records over time
- Allows for the migration of records to new formats, avoiding information loss through format obsolescence
- Uses digital forensic tools to recover information from obsolete media
- Increases the transparency of government operations by providing a web portal (<https://tsl.access.preservica.com/>) that allows the public to search, view, and download records online

The TDA will take state agency records that have been scheduled as "A" (Archival) or "R" (Review) on an approved records retention schedule. "R" records will need to be reviewed by an appraisal archivist prior to transfer.

- State agencies: For more information about the types of records for transfer see the [FAQ's for State Agencies](#) on the TSLAC website
- Legislators: For more information about the types of records for transfer see the [FAQ's for Legislators](#) on the TSLAC website

For questions about transferring electronic records to TSLAC, an agency records management officer (RMO) should contact either:

- [Laura Saegert](#), Assistant Director for Archives

or

- [Mark Myers](#), Senior Electronic Records Specialist

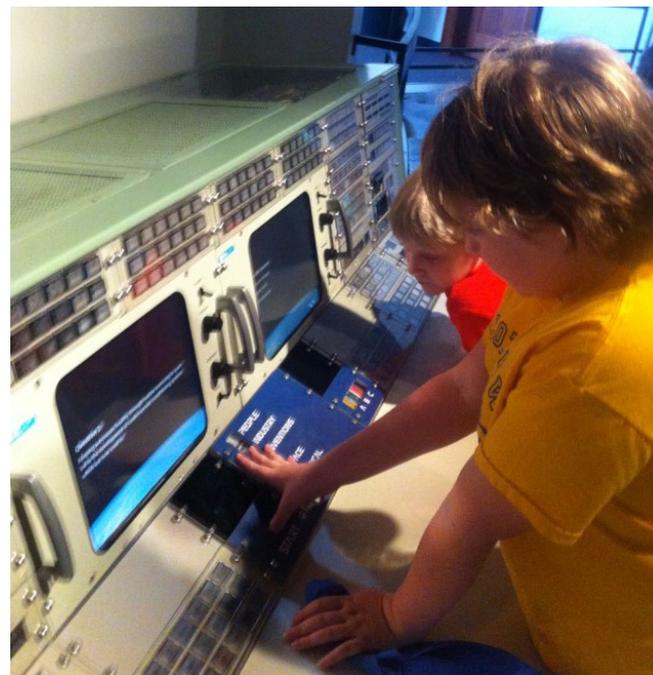
[back to top](#)

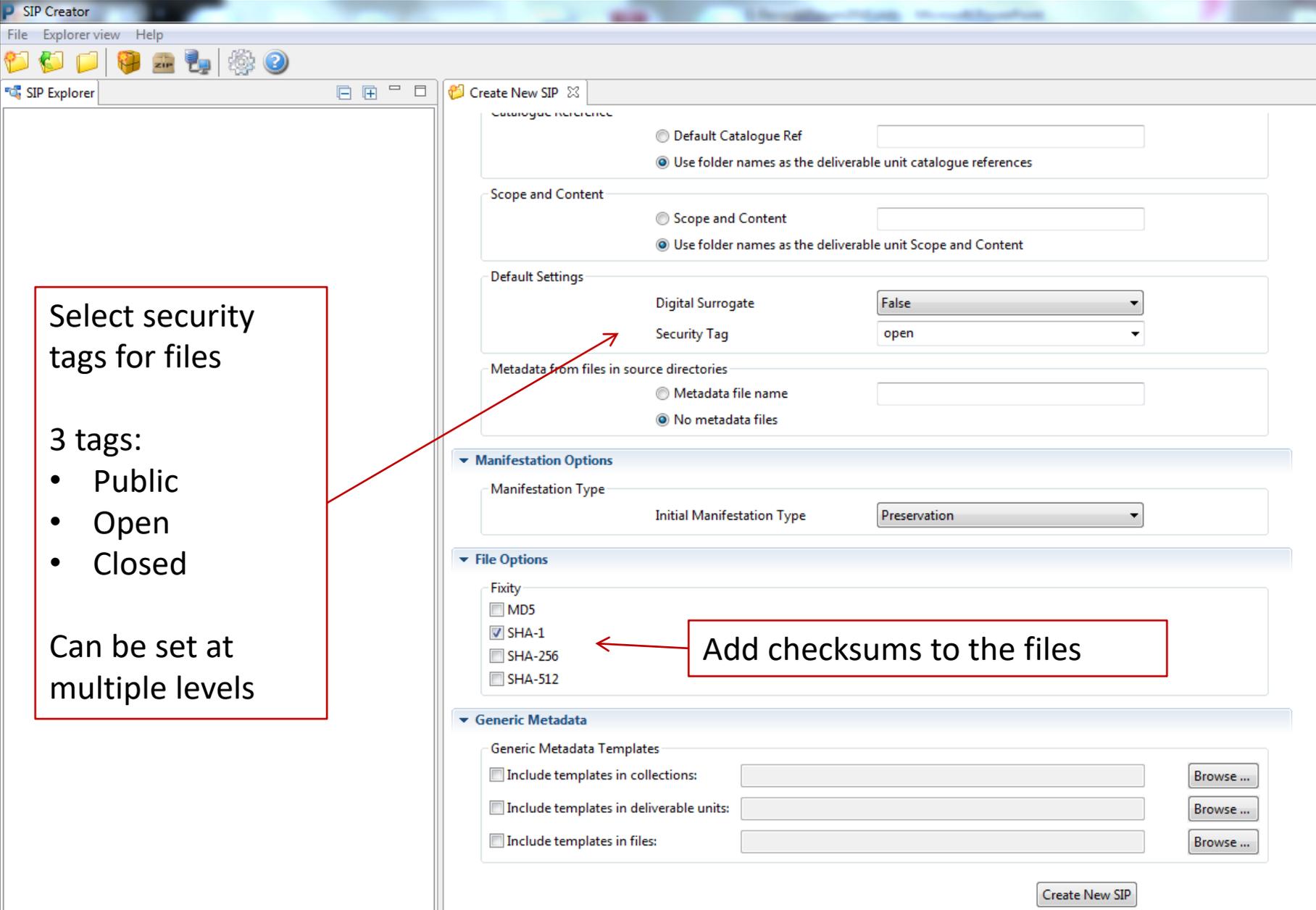
Agency RMO may also complete and return the [Electronic Records Survey](#) form for general information about the types of electronic records at the agency.

For more information about the transfer process please refer to: [Procedures for Transferring Electronic Records](#)

Step two - build the foundation

- TDA is built in the Amazon Government Cloud
 - Amazon S3 online storage
 - Glacier near line storage
- Encryption in transit and at rest
- Rolling integrity checks





Select security tags for files

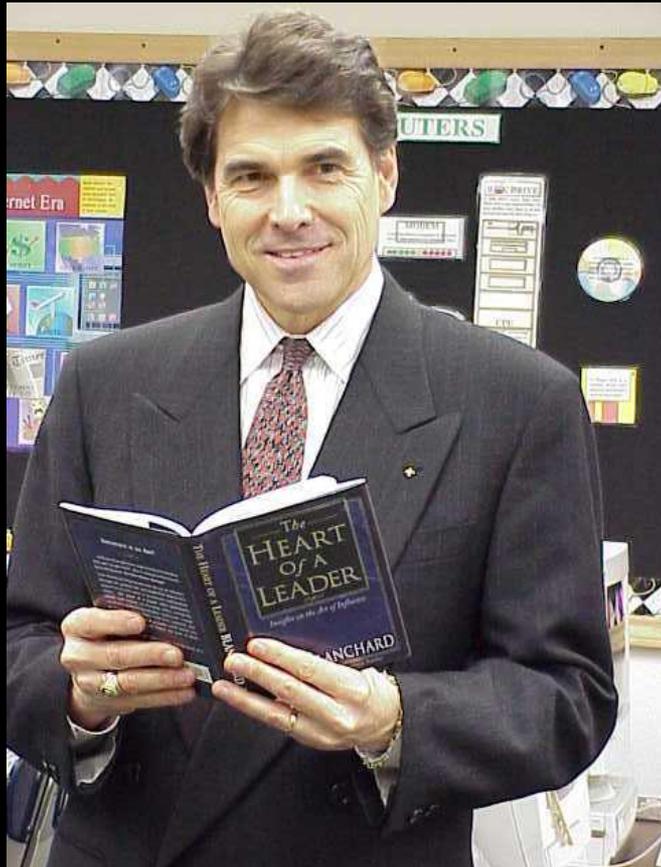
3 tags:

- Public
- Open
- Closed

Can be set at multiple levels

Add checksums to the files

Integrity = Protecting the bits



Example of Bit Rot

Preservation

Not Simply Storage
Records



Texas Digital Archive (TDA)

Functional Requirements

- Ingest electronic records from state agencies in any format
- Characterize, validate and track digital formats over time
- Monitor integrity and security of digital files
- Migrate to sustainable formats for preservation
- Collect and store all appropriate metadata about the records
- Allow for proper access to records that meets the needs of our users
 - Provide records in access formats based on user needs
- Follow appropriate archival standards for preservation



Long-Term Preservation: Sustainable Formats

Types of Standards

- State/Enterprise
- National/
International
Standards
(ANSI, ISO)
- Industry
Standards
- “De facto”
Standards

Formats expected not to change,
or change slowly

- Widely supported & used
- Easily Transferable
 - Compatible with other applications
 - Forward/Backward Compatibility
 - Version Control
- Non-proprietary or
open/published standards



fox

Ingest Access Preservation Data Management Administration Help

Start Waiting Running **Completed** Reports Manage

Workflow Details

Workflow Context	Auto Ingest (standard)
Workflow Definition	Standard Ingest Workflow (Auto Selection)
Workflow ID	16109
Workflow State	Completed
Date Started	29.10.15 09:06:12
Date Finished	29.10.15 09:43:43
Number of Files	847
Total Size	27 MB
Collection Code	009
Submission name	Speeches, 2001-2014 0.5 cubic ft. and 27.4 MB (848 files, 14 folders)
Top Level Record	speeches_and_remarks

Back

Step Progress

State	Name	Progress	Started	Finished	Messages
✓	Identify Packages	██████████	29.10.15 09:06:12	29.10.15 09:06:19	
✓	Import from Transfer Area	██████████	29.10.15 09:06:19	29.10.15 09:06:28	
✓	Process Package	██████████	29.10.15 09:06:28	29.10.15 09:06:31	
✓	Virus Check	██████████	29.10.15 09:06:31	29.10.15 09:06:43	
✓	Fixity Check	██████████	29.10.15 09:06:43	29.10.15 09:06:46	
✓	Metadata Integrity	██████████	29.10.15 09:06:46	29.10.15 09:06:49	
✓	Content Integrity	██████████	29.10.15 09:06:49	29.10.15 09:06:52	
✓	SIP Validation	██████████	29.10.15 09:06:52	29.10.15 09:06:55	
✓	SIP Validation with Database Crosscheck	██████████	29.10.15 09:06:55	29.10.15 09:06:58	
✓	Characterise	██████████	29.10.15 09:06:58	29.10.15 09:08:07	View
✓	Generate Storage Routing Map	██████████	29.10.15 09:08:07	29.10.15 09:08:13	
✓	Store Files	██████████	29.10.15 09:08:13	29.10.15 09:10:16	
✓	Store Metadata	██████████	29.10.15 09:10:16	29.10.15 09:11:40	
✓	Delete from Transfer Area	██████████	29.10.15 09:11:40	29.10.15 09:11:43	
✓	Store Metadata File	██████████	29.10.15 09:11:43	29.10.15 09:11:46	
✓	Update Search Index	██████████	29.10.15 09:11:46	29.10.15 09:12:37	View
✓	Thumbnail Creation	██████████	29.10.15 09:12:37	29.10.15 09:43:42	

Ingest Workflow – how we bring a file into the repository.

- Runs a series of “micro-services” on the files
- Most important for long-term preservation being the “Characterize” step.

File Edit View History Bookmarks Tools Help

Explorer - Properties

Tessella Limited (GB) | https://us.preservica.com/explorer/explorer.html#prop:108&2baec1b-6088-49d3-bf49-1ac78460c9e2&0

Welcome | TSLAC | Angelina | Texas State ... | Outlook | Preservica site | Preservica Access | PreservicaUserGroup | PreservicaDemoSite

Preservica Digital Preservation

Mark Myers (0) English

Dashboard Ingest Access Preservation Data Management Administration Help

Root > Texas Governor Rick Perry > Office of the First Lady > Biographical information, 2008-2014, 1.72 MB (31 files) > First Lady Anita Perry (FLAP) various bios (biographies), 2008-2011 [335 KB, 13 files] > CURRENT FLAP BIO.doc

CURRENT FLAP BIO.doc

Description Technical Metadata History

File Properties

Name	Value
Creating Application	Microsoft Office Word
Creation Date	Thu Dec 18 21:54:00 UTC 2008
Encrypted	false
Number of Pages	1
Creator	michelle.sneed
Word Count	290
Character Count	1653

Formats

Name	PUID	Version
Microsoft Word Document	fmt/40	97-2003

Fixity

Name	Value
SHA-1	08f9288e6116dee1e856912ace2bce886829b9ce

Characterize step identifies the file format

Plugged in to an international registry of file formats

Filter Apply

File Formats (1603)

Property Groups (11)

Migration Pathways (663)

PUID	Name	Version
pth/172	Open Office Convert DOC 6.0/95 to ODT 1.2	
pth/124	Open Office Convert DOC 6.0/95 to PDF 1.4	
pth/170	Open Office Convert DOC 6.0/95 to PDF/A	
pth/173	Open Office Convert DOC 97-2003 to ODT 1.2	

24 of 34 461-480 of 663

Software (297)

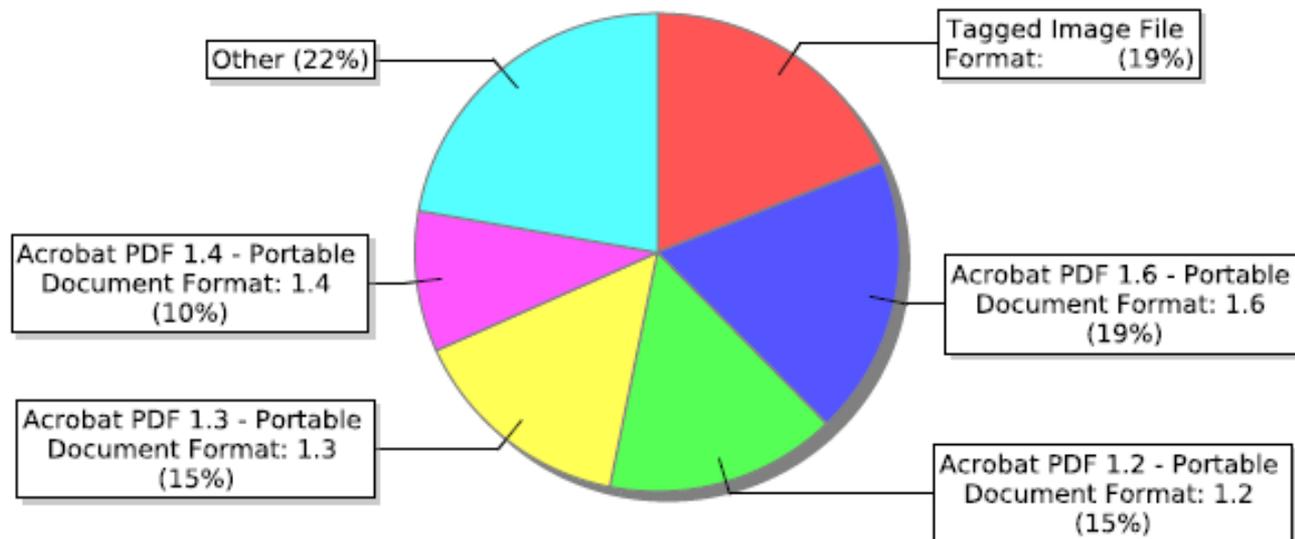
Tools (94)

Format

Microsoft Word Document

Version	97-2003
PUID	fmt/40
Description	With the release of Word 97, Microsoft revised the native binary word processing format, which is based on its generic OLE2 Compound Document Format. The format is proprietary and Microsoft does not make details of its structure public. The information here is derived primarily from OpenOffice.org's reverse-engineered documentation of the format and should not therefore be regarded as definitive. A Word document is stored as a 'WordDocument' stream within a Compound Document Format file. The format remained unchanged with the releases of Word 2000, 2002 and 2003. An alternative extension of .wbk refers to a backup file of a Word document, however there is no material or structural difference between a .wbk file and the .doc file it is a backup of.
File Extensions	wbk doc
MIME Type	application/msword
Release Date	
Withdrawn Date	
Format Group	Document / Microsoft Word
Internal Signatures	
Properties	
Property Extraction	
Created By	
Validated By	
Rendered By	
Embedded Objects Extracted By	
Has Priority Over	

File formats - detail



File format	Version	Format puid	Frequency
Tagged Image File Format		fmt/353	863653
Acrobat PDF 1.6 - Portable Document	1.6	fmt/20	861995
Acrobat PDF 1.2 - Portable Document	1.2	fmt/16	706656
Acrobat PDF 1.3 - Portable Document	1.3	fmt/17	692193
Acrobat PDF 1.4 - Portable Document	1.4	fmt/18	435345
Microsoft Word Document	97-2003	fmt/40	110383
Exchangeable Image File Format	2.2.1	fmt/645	89839
Acrobat PDF/A - Portable Document	1a	fmt/95	88626
JPEG File Interchange Format	1.01	fmt/43	87409
Raw JPEG Stream		fmt/41	77102

System allows us to normalize to a variety of file formats

The screenshot shows the Preservica web interface. The breadcrumb trail is: Root > Texas Governor Rick Perry > Office of the First Lady > Biographical information, 2008-2014, 1.72 MB (31 files) > First Lady Anita Perry (FLAP) various bios (biographies), 2008-2011 [335 KB, 13 files]. The main table lists 13 files, all of which are Microsoft Word Documents. A red circle highlights the 'Description' column, and a red arrow points from a text box to the 'Manifestation' column, which is set to 'Preservation 1 (Original)'. The text box contains the text: 'These are MS Word documents marked as the Original'.

#	Name	Description	Size	Type	Manifestation
	CURRENT FLAP BIO.doc	Microsoft Word Document	23.5 KB	File	Preservation 1 (Original)
	First Lady Anita Perry bio 200.doc	Microsoft Word Document	27 KB	File	
	First Lady Anita Perry bio TCW.doc	Microsoft Word Document	26.5 KB	File	
	First Lady's Intro- 4-08.doc	Microsoft Word Document	25 KB	File	
	FLAP bio for S. America 03 10.doc	Microsoft Word Document	26.5 KB	File	
	FLAP description for Senate Ladies Club director 12 08.doc	Microsoft Word Document	22 KB	File	
	FLAP's BIO REV. 0908.doc	Microsoft Word Document	31 KB	File	
	FLAP Web bio.doc	Microsoft Word Document	27.5 KB	File	
	FLAP Web bio revised.doc	Microsoft Word Document	27.5 KB	File	
	FL INTRO REVISED 11 08 FINAL.doc	Microsoft Word Document	23 KB	File	
	FL Intro REVISION B.doc	Microsoft Word Document	23 KB	File	
	Priorities.doc	Microsoft Word Document	27.5 KB	File	
	Rewritten FL bio 1108.doc	Microsoft Word Document	25 KB	File	

Name: CURRENT FLAP BIO.doc
File Type: Microsoft Word Document
Type: File Size: 23.5 KB Reference: 82baec1b-6088-49d3-bf49-1ac78460c9e2

These are MS Word documents marked as the Original

File Edit View History Bookmarks Tools Help

Explorer - First Lady Anita ... Registry

Tessella Limited (GB) | https://us.preservica.com/explorer/explorer.html#browse:78&469ea88c-7836-4c1f-939a-308766ee16bf&2

Google

Welcome | TSLAC | Angelina | Texas State ... Outlook Preservica site Preservica Access PreservicaUserGroup PreservicaDemoSite

Preservica Digital Preservation

Mark Myers (0) English

Dashboard Ingest Access Preservation Data Management Administration Help

anita perry

Root > Texas Governor Rick Perry > Office of the First Lady > Biographical information, 2008-2014, 1.72 MB (31 files) > First Lady Anita Perry (FLAP) various bios (biographies), 2008-2011 [335 KB, 13 files]

Sync Tree

Tools View

#	Name	Description	Size	Type
	CURRENT FLAP BIO.pdf	Acrobat PDF 1.4 - Portable Document Format	33.388 KB	File
	First Lady Anita Perry bio 200.pdf	Acrobat PDF 1.4 - Portable Document Format	36.057 KB	File
	First Lady Anita Perry bio TCW.pdf	Acrobat PDF 1.4 - Portable Document Format	37.174 KB	File
	First Lady's Intro- 4-08.pdf	Acrobat PDF 1.4 - Portable Document Format	30.087 KB	File
	FLAP bio for S. America 03 10.pdf	Acrobat PDF 1.4 - Portable Document Format	35.746 KB	File
	FLAP description for Senate Ladies Club director 12 08.pdf	Acrobat PDF 1.4 - Portable Document Format	16.513 KB	File
	FLAP's BIO REV. 0908.pdf	Acrobat PDF 1.4 - Portable Document Format	45.226 KB	File
	FLAP Web bio.pdf	Acrobat PDF 1.4 - Portable Document Format	42.194 KB	File
	FLAP Web bio revised.pdf	Acrobat PDF 1.4 - Portable Document Format	37.829 KB	File
	FL INTRO REVISED 11 08 FINAL.pdf	Acrobat PDF 1.4 - Portable Document Format	29.814 KB	File
	FL Intro REVISION B.pdf	Acrobat PDF 1.4 - Portable Document Format	29.816 KB	File
	Priorities.pdf	Acrobat PDF 1.4 - Portable Document Format	35.844 KB	File
	Rewritten FL bio 1108.pdf	Acrobat PDF 1.4 - Portable Document Format	30.084 KB	File

Manifestation Presentation 2 (Active)

Filter Apply Show accrued files Show children Show deprecated

Name: CURRENT FLAP BIO.pdf
 File Type: Acrobat PDF 1.4 - Portable Document Format
 Type: File Size: 33.388 KB Reference: 91f325d4-a278-4d7c-9e81-c46bcae3786c

After migration now have a "Presentation" format in PDF.
 Presentation format is the one that gets pushed out to the public.

From: Karen Chalkley Turcotte
To: Dwight Harris; Steve Robinson; Tom Tye; Judy Binscoe
Cc: Pamela Ward; Steve Spilman; Butch Held; Silvia Machado
Subject: Escapes Beto House
Date: Monday, August 30, 1999 2:13:23 PM
Importance: High

Saturday 8/28/99 at approximately 10:45 P.M., four youth escaped from Beto House. They broke out a window in the bedroom and left the facility.

Two of the youth were apprehended in Pharr at approximately 3:00 A.M Sunday. According to the police report, this is what occurred:

A police officer observed a car parked on the side of the road. He pulled up behind it and turned on his lights. Four males ran from the car. Two females, Jennifer Moreno and Nadia Pena who were in the vehicle were questioned. These two females are ERJC employees who had been volunteering at Beto House. Ms. Moreno and Ms. Pena had a visit with two of the youth, earlier that Saturday.

When the officer questioned the women, they stated that the youth had forced their car off the road and told them to let them in and drive away. The women stated they refused to drive away and that is when the officer arrived and the four boys ran off. When two of the youth were apprehended, the women stated they could not be sure if either of them were the four males involved in the car jacking and refused to press charges of car jacking or robbery.

When the youth were questioned, allegations of a sexual relationship between one of the female staff and a youth were made. The youth stated that they had all four escaped together. They paged Ms. Pena and she picked them up in Mission, Texas. She drove them to the location on the road where the officer had pulled up behind the car. She dropped them off on the road side and said she would return in 45 minutes. She did so with Ms. Moreno and that is when the officer arrived. The police brought the personal items the youth had escaped with back to the House. The items were in the trunk of Ms. Pena's car.

Butch Held and Steve Spilman have been notified and the police will be investigating further and have expressed a desire to pursue charges against the women.

Carmen Gonzalez has filed the appropriate alleged mistreatment report due to sexual contact allegations.

Please contact me if you require additional information.
Karen Turcotte

June 03, 2014 Continued
Tuesday

we will try again, if you want to

just shoot me a couple or three possible dates -

bob turner

On Wednesday 05/21/2014 at 11:31 am, Mike Morrissey wrote:
I am stuck today

Sent from my iPhone

On May 21, 2014, at 11:09 AM, "bob turner" wrote:

Would it work better - say at 1 -- would you have made your "pitch" by then?? Bob.
(man - the job has to come 1st -- I can't feed your hungry family). Bob.

From: Mike Morrissey

Sent: Wednesday, May 21, 2014 10:58 AM

To: bob turner

Subject: Re: possible lunch dates May 16? 21, 22?

At this point I think I'm stuck

Sent from my iPhone

On May 21, 2014, at 8:57 AM, "bob turner" wrote:

Does it still look possible.? Bob.

From:

Mark Myers

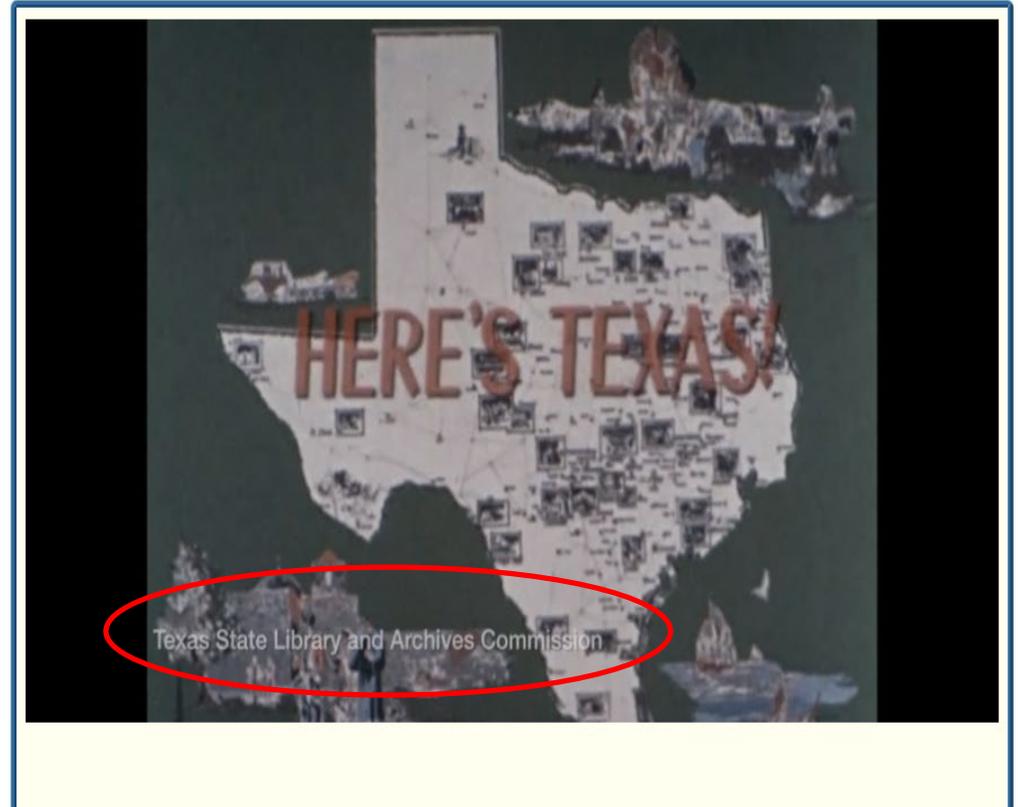
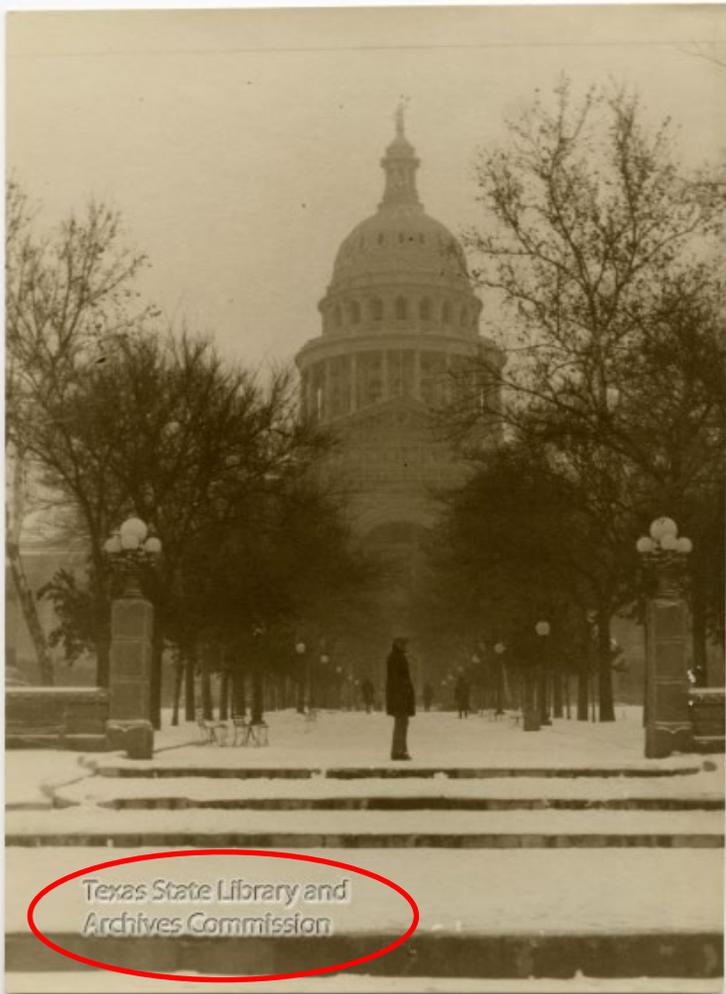
181

5/29/2015 3:36 PM

Redaction of an Outlook calendar entry in a PST file

Redaction of an email





We also watermark images and films for extra protection.
The watermarked image is the presentation copy



Providing Public Access to electronic records



Access = Preservation



Public Access Portal

- Preservica Universal Access
 - Free – provided as part of preservation system
 - WordPress front-end
 - Linked directly to back-end Preservica repository
 - Can control access through security settings
 - Display/search
 - Customizable look and feel
 - Browse, read, search



113 total results

In complete archive

Sort by Relevance



REFINE YOUR SELECTION

Format group

- Document
- Image
- Spreadsheet

Item or folder title

- Woodlawn
- Capitol in the snow, Austin, about 1920
- Snow in Austin, Texas (6th and Congress Avenue).
- Calendar of the papers of Mirabeau Buonaparte Lamar. compiled and edited by Michael R. Green.

Collection

- Texas Historical Commission History Programs Division records
- Graham (R. Niles) collection
- Mabel H. Brooks photograph collection
- Places collection
- Representative Sylvester Turner records, Texas House of Representatives
- Mirabeau Buonaparte Lamar papers

Type

- Image
- Photographs
- Catalogs

	spbld.dat
	Texas Credit Union Commission 03-03-05.pdf
	Capitol in the snow, Austin, about 1920
	Capitol in the snow, Austin, about 1920
	06-09 Texas Quarter.pdf
	06-09 Texas Quarter.pdf
	tcbld.dat
	Snow in Austin, Texas (6th and Congress Avenue).
	05-20 Texas Quarter Launch Advisory.pdf
	05-20 Texas Quarter Launch Advisory.pdf
	ombld.dat

Search

- Searches everything:
 - Metadata (seen/unseen)
 - Full-text searching of text files
- Facets and filters let you narrow the search



Content viewers

East Sixth Street, Austin, undated

Object Type: File
In Record: [Scrapbook 3](#)

East Sixth Street, Austin, undated Details Title: East Sixth Street, Austin, und...	Citation Information: 1932/005-21... Identifier: 1932005_214	Date Range of Creation: undated Geographic Coverage: Austin (Te...	Medium: Black-and-white photogra... Copyright Information: The copyi... Type: Image Language: English
---	--	--	--

[Show more ▼](#)



Photos

2008-12-05-ChristmasGreeting-youtube.mp4

Object Type: File
In Record: [2008-12-05-ChristmasGreeting-youtube](#)

Archival Description	TARO Identifier: um:tao:tslac:501...	Descriptive Rules: DACS	Other Finding Aid: http://www.lib...
	Finding Aid Title: Texas Governor ...	Title: Video recordings	Preferred Citation: (Ident...

[Show more ▼](#)



Video

2010-12-02-Greeting-Phone-82ndLegislativeSession.mp3

Object Type: File
In Record: [Phone Holds](#)



Audio



Benefits of transferring records to the TDA

- Relieves the preservation burden from the agency
- Saving taxpayer money – Reducing agency storage and migration costs
- Reducing agency staff time for PIA requests
- Increases transparency by making more records accessible
- Preserves and protects electronic state records a secure trusted repository



Want to transfer electronic records to the TDA?

- First step – Contact us!!
- Complete our electronic records survey
- Identify the series to transfer
- Negotiate the timing and method transfer





WHO YOU GONNA CALL?

- Laura Saegert – Assistant
Director for Archives

- lsaegert@tsl.texas.gov

- 512-463-5500

-Mark Myers – Senior Electronic
Records Specialist

- mmyers@tsl.texas.gov

- 512-463-5434

-If you have worked with an appraisal
archivist you can contact them as well.



Questions?

TSLAC Wants Your



Electronic Records

