Accessioning Digital Files: Tales From the Trenches at the California State Archives

Society of California Archivists Annual General Meeting, 2012

Rebecca L. Wendt
California State Archives
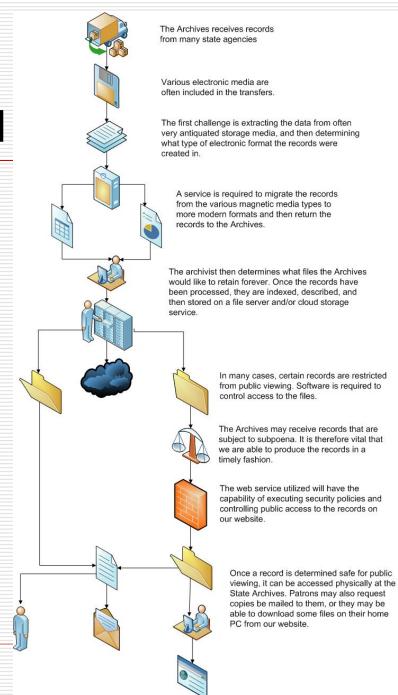
CA State Archives

- ☐ Electronic Records Program
  - ■ Electronic Records Task Force
    - ☐ Electronic Records Archivist
    - ☐ Systems Analyst
    - ☐ Archivists
  - ■ IT Department (Secretary of State)
- ☐ Appraisal Program
- ☐ State Agency Records (3 branches of government)

E-Records Program

- ☐ Vision statement
- ☐ Strategic plan
- ☐ OAIS model adopted (still in progress)
- ☐ Policies and procedures created
- ☐ Metadata standards adopted
- ☐ Experimentation with Archivematica and iRODS
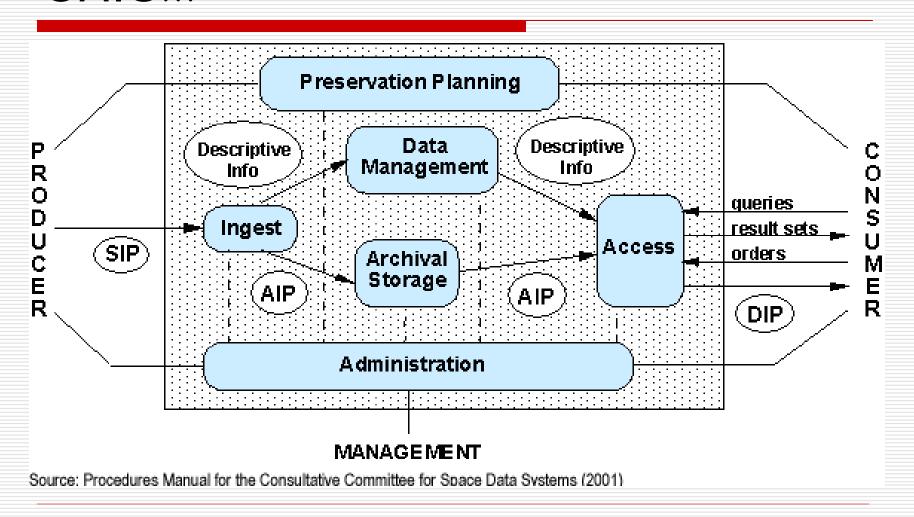
It's An Imperfect World

- ☐ We might not know what's coming
- ☐ We might not be able to handle the media/format
- ☐ The metadata may be lacking
- ☐ What preservation server storage space?!

The Archives receives records from many state agencies

Various electronic media are often included in the transfers.

The first challenge is extracting the data from often very antiquated storage media, and then determining what type of electronic format the records were created in.

A service is required to migrate the records from the various magnetic media types to more modern formats and then return the records to the Archives.

The archivist then determines what files the Archives would like to retain forever. Once the records have been processed, they are indexed, described, and then stored on a file server and/or cloud storage service.

In many cases, certain records are restricted from public viewing. Software is required to control access to the files.

The Archives may receive records that are subject to subpoena. It is therefore vital that we are able to produce the records in a timely fashion.

The web service utilized will have the capability of executing security policies and controlling public access to the records on our website.

Once a record is determined safe for public viewing, it can be accessed physically at the State Archives. Patrons may also request copies be mailed to them, or they may be able to download some files on their home PC from our website.

Appraisal, Transfer, Receiving, & Accessioning

- ☐ Reliance on records managers
- ☐ Format
- ☐ Sneakernet
- ☐ Maintaining authenticity
- ☐ Clean machine
- ☐ Triage
- ☐ **THIS IS A TRANSITION PERIOD**

Until we have a fully functional OAIS...

Source: Procedures Manual for the Consultative Committee for Space Data Systems (2001)

…A Different Workflow

- ☐ Work closely with our State Records Appraisal Program
- ☐ Work closely with records creating agencies
- ☐ Be involved throughout the process
- ☐ Educate, educate, educate
- ☐ Join larger agencies (NDSA, etc.) for help and support

Retention Schedules

STD 73

(32) CalRIM APPROVAL NUMBER											
								.10 ˉ 15 q			

ITEM # (33)	CUBIC FEET * (34)	NOTIFY CA. STATE ARCHIVES (35)	TITLE AND DESCRIPTION OF RECORDS (*Triple spaces between items*) (36)	MEDIA (37)	VITAL (38)	RETENTION				PRA (Exempt) & IPA (43)	REMARKS (44)
						OFFICE (39)	DEPT. (40)	SRC (41)	TOTAL (42)		
1.	**N/A**		**ELECTRONIC RESOURCES**	E		**VARIES**					
		NOTIFY ARCHIVES	Emails	E		Active emails retained, plus permanent retention of emails that contain significant or noteworthy analysis, deliberation, or factual content. Emails to be permanently retained include those that contain content that traditionally would have been captured in paper correspondence or paper memoranda.				✓	Retained emails to be sent to Archives at end of Administration under Gov. Code, § 6268. May contain material exempt from production under the PRA pursuant to Gov. Code, § 6254, subd. (l) and Gov. Code, § 6255, and other PRA exemptions.
1b	n/a	NOTIFY ARCHIVES	Calendars	E		Active (i.e. calendars for current employees).				✓	Calendars retained in the Governor's Office at the end of his or her term to be sent to Archives of under Gov. Code, § 6268. May contain material exempt from production under the PRA pursuant to Gov. Code, § 6255, and other PRA exemptions.
1c	n/a	NOTIFY ARCHIVES	User data (Word documents, Excel spreadsheets, databases [including correspondence tracking database, legislative database, scheduling database, etc.])	E		Active files retained, plus permanent retention of user data that contain significant or noteworthy analysis, deliberation, or factual content. Files to be permanently retained include those files that contain content that would have traditionally been captured in paper correspondence or paper memoranda.				✓	Retained user data to be sent to Archives at end of Administration under Gov. Code § 6268. May contain material exempt from production under the PRA pursuant to Gov. Code, § 6254, subd. (l) and Gov. Code, § 6255, and other PRA exemptions.

Memo to State Agencies

DEBRA BOWEN | SECRETARY OF STATE
STATE OF CALIFORNIA | CALIFORNIA STATE ARCHIVES
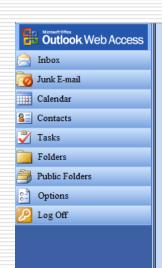1020 O Street | Sacramento, CA 95814 | Tel (916) 653-7715 | Fax (916) 653-7363 | www.sos.ca.gov

Dear Records Managers:

As you know, the State Archives collects, catalogs, preserves, and provides access to the historic records of California state government. You may be familiar with the way the Archives appraises and "flags" records listed on retention schedules that have archival value. These records may be in various media and formats: paper, microfilm, maps, microfiche, motion picture film, and electronic formats. To comply with state law and to assist us in identifying records that have archival value, it is vital that all records--including electronic records--be included in your records survey and on your records retention schedule.

We realize that electronic records present different and unique challenges. They are often in proprietary formats and may also be so voluminous and widely distributed that a full accounting may be more difficult than with traditional formats. This makes good records management a vital step in the identification and preservation of these valuable records. Here are some tips to help you with this important task:

Constant Communication

To: Wendt, Rebecca
Cc: Raul Villanueva; Daniel Claypool
Subject: Citizens Redistricting Commission - Archive Project

Good Afternoon Rebecca,

Thank you for taking my call this morning.

I did some digging and have information for you on the volume of electronic data you might expect to collect.

- SFL Data, our contractor, has approximately 14GB of collected data that would need to be transferred via a method that maintains the chain of custody. This should include all Commissioner computer images, CRC agendas, transcripts, public comment, handouts, press clippings, and maps. Primary formats are .pdf, .eml, .doc, and .xml. The Commission maps are available with census block level detail in .dbf, .shp, .cdf, and .kmz. The maps are also available in graphical format in .pdf files.
- There are also 88GB of video (.flv) and 500MB of audio (.mp3) files. These will encompass all Commission meetings and teleconferences.

I believe I estimated about 100GB of data this morning. If we account for uncollected data, I'd say you are talking about no more than 110-120GB. We've understood from the beginning of this process that our records would be of historical interest, so I've tried to set up logical data handling procedures that make the data easy to document and search. That said, we've amalgamated records from three different state agencies and there are some quirks present. I am happy to provide any expertise you need locating items of interest or sorting out our records.
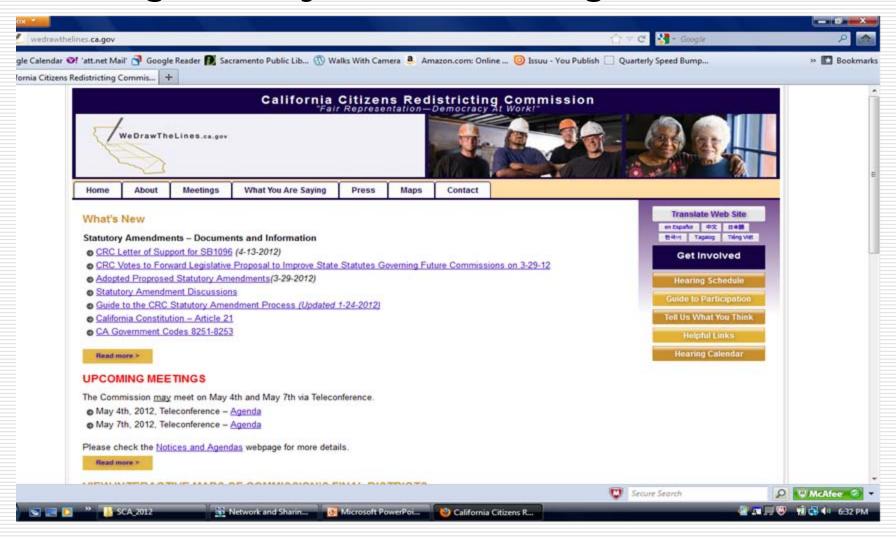
A question we have: Should the Commission be subject to litigation at some point after transferring our records to the State Archives, how would we go about gaining access to the records?

Hope the day is treating you well, and looking forward to speaking with you Friday. :)
--

Christina Shupe
Senior Operations Analyst
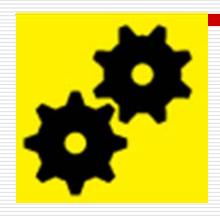Citizens Redistricting Commission
"Fair Representation - Democracy at Work!"

Working Closely: Redistricting Commission

NDSA

http://www.digitalpreservation.gov/ndsa/

Infrastructure

So what do we actually do?

- ☐ Ask for specific formats and/or notification of formats
- ☐ Ask for all documentation
- ☐ Ask for all encryption to be removed
- ☐ Ask for specific metadata
- ☐ Ask for checksums
- ☐ Ask for transfer on hard drives

What If?

- ☐ If electronic records just "show up," the same processes apply EXCEPT
- ☐ No pre-existing checksums
- ☐ We might have to look for more drives/emulators
- ☐ We might fail

Receive + Accession Records

- ☐ "Receiving" record created in our collection management system by any archivist

- ☐ Records run through "clean machine" by any archivist

- ☐ Accessioning record created in our collection management system by any archivist

- ☐ Triage decision made by *lucky* E-R Archivist: to preservation server or maintain on original media **for now**?

The "clean machine" & a multitude of drives

Dirty Laundry

- We have 20+ TB of born digital material
- 1/3 of that (maybe) is on a preservation server and backed up
- A good access component for our OAIS is lacking...we have to create use copies *On Demand*
- Certification of records not possible at this time

Triage

- ☐ This is RISKY!
- ☐ We know!

We're *obviously* not perfect BUT

- ☐ We are no longer waiting for perfection
- ☐ We ARE accessioning digital records
- ☐ We ARE increasing system capacity
- ☐ We ARE looking for partners
- ☐ We ARE looking for other funding sources
- ☐ We ARE continuing to experiment

To Infinity and Beyond (otherwise known as Next Steps)

- ☐ Work with CIOs, Records Managers, et al.
- ☐ Assist in statewide IT planning
- ☐ Raise awareness at higher levels
- ☐ Complete our OAIS
- ☐ Continue to accession digital records
- ☐ Take over the world

THANKS FOR LISTENING!

Any questions?

Rebecca L. Wendt

California State Archives

rwendt@sos.ca.gov