State Government Digital Preservation Profiles

Pennsylvania LARM

July 2006

This page intentionally left blank.

Introduction

The state government digital preservation profiles available on this Web page are based on information collected from those state and territorial library, archives, and records management units that completed CTG's State Government Digital Information Preservation Survey. CTG administered the survey to all state and several territorial librarians, archivists, and records managers in early 2006. See baseline report for details.

The profiles are organized by state or territory and the library (**L**), archives (**A**), and records management (**RM**) units that were represented in the survey response. (For example, "**Alabama ARM**," which is the first responding unit in the drop down menu below, refers to the **a**rchives and **r**ecords **m**anagement units from Alabama in one survey response.) For some states, respondents also included agencies other than library, archives, and records management. These profiles are marked with an asterisk (*) and the other units that contributed to the survey response are specified in the profiles.

To select a different profile, click on the Introduction/Select Profile link located on the top left of the screen.

You can also view consolidated tables showing how all the state/territory responding units replied to selected questions in the survey. (This is also found in Appendix E of Preserving State Government Digital Information: A Baseline Report.)

STATE PROFILES CONTACT
G. Brian Burke, Project Manager
Center for Technology in Government
University at Albany, SUNY

Section 1. Responding Unit(s)

Name of state or territory.
Commonwealth of Pennsylvania

Name and affiliation of individual(s) who responded on behalf of the state or territory.
Linda F. Avetta, Pennsylvania Historical and Museum Commission, State Archives
717-705-6923 or lavetta@state.pa.us;

George Friedline, Department of Education, State Library
717-783-5964 or gfriedline@state.pa.us

This profile includes information provided by the following units for this state:
• Library
• Archives
• Records Management

Section 2.1a. Setting Standards for EXECUTIVE Agencies

The following units have authority for setting standards, specified below, for digital information created or maintained by EXECUTIVE agencies. Multiple boxes are checked to reflect authority that is shared or delegated.

	Authority not assigned	Library has authority	Archives has authority	Records Management has authority	Others (See below)
Setting data management standards and or guidelines for information creation (e.g., metadata, file formats).			X	X	X
Setting information technology standards and or guidelines for information creation (e.g., state approved software applications).		X	X	X	X
Setting standards for information retention and disposal (e.g., retention periods and methods of disposal) for various series/types of digital records and publications.		X	X	X	X

Other standards and the units with authority over setting them.
Data Management and IT standards are set by the Office of Administration / Office of Information Technology; however the Archives, Library & Records Management areas do provide recommendations and draft some of those standards and guidelines.

Additional information.
The Office of Administration and the State Archives work in conjunction with each other to define standards and guidelines where records are involved.

Section 2.1b. Services Provided to EXECUTIVE Agencies

The following units provide the services, specified below, to EXECUTIVE agencies. If services are provided by multiple units, the appropriate boxes are checked.

	Services not provided	Library provides services	Archives provides services	Records Manage-ment provides services	Others (See below)
Storage for digital information.	X				
Consultation and training services on digital information creation.			X	X	X
Consultation and training services on digital information management.			X	X	X
Consultation and training services on digital information preservation.			X	X	X
Consultation and training services on digital information preservation.			X	X	X
Preservation (e.g., migration, reformatting).			X	X	X
Access (e.g., search engine).			X		X
Certification (e.g., trustworthiness of system, backups sufficient).			X		X

Other services provided and the units that provide them.
As with setting standards and policy, the Office of Administration / Office of Information Technology also provides digital information creation, management, access, certification guidance, assistance and services. They aren't involved as much in preservation at this time, but we are attempting to have the preservation aspect be addressed at the time of the creation of digital records, rather than an afterthought.

Additional information.
No information provided

Section 2.2a. Setting Standards for LEGISLATIVE Agencies

The following units have authority for setting standards, specified below, for digital information created or maintained by LEGISLATIVE agencies. Multiple boxes are checked to reflect authority that is shared or delegated.

	Authority not assigned	Library has authority	Archives has authority	Records Manage-ment has authority	Others (See below)
Setting data management standards and or guidelines for information creation (e.g., metadata, file formats).	X				
Setting information technology standards and or guidelines for information creation (e.g., state approved software applications).	X				
Setting standards for information retention and disposal (e.g., retention periods and methods of disposal) for various series/types of digital records and publications.	X				

Other standards and the units with authority over setting them.
No information provided

Additional information.
No information provided

Section 2.2b. Providing Service to LEGISLATIVE Agencies

The following units provide the services, specified below, to LEGISLATIVE agencies. If services are provided by multiple units, the appropriate boxes are checked.

	Services not provided	Library provides services	Archives provides services	Records Manage-ment provides services	Others (See below)
Storage for digital information.	X				
Consultation and training services on digital information creation.	X				
Consultation and training services on digital information management.	X				
Consultation and training services on digital information preservation.	X				
Consultation and training services on digital information preservation.	X				
Preservation (e.g., migration, reformatting).	X				
Access (e.g., search engine).	X				
Certification (e.g., trustworthiness of system, backups sufficient).	X				

Other services provided and the units that provide them.
No information provided

Additional information.
It's not that services and training couldn't be made available . . . we could provide services at their request. We have no authority over the legislature, their services and / or their training.

Section 2.3a. Setting Standards for JUDICIAL Agencies

The following units have authority for setting standards, specified below, for digital information created or maintained by JUDICIAL agencies. Multiple boxes are checked to reflect authority that is shared or delegated.

	Authority not assigned	Library has authority	Archives has authority	Records Management has authority	Others (See below)
Setting data management standards and or guidelines for information creation (e.g., metadata, file formats).	X				
Setting information technology standards and or guidelines for information creation (e.g., state approved software applications).	X				
Setting standards for information retention and disposal (e.g., retention periods and methods of disposal) for various series/types of digital records and publications.	X				

Other standards and the units with authority over setting them.
No information provided

Additional information.
As with the Legislature, we have no authority over Judicial at the State level.

Section 2.3b. Providing Service to JUDICIAL Agencies

The following units provide the services, specified below, to LEGISLATIVE agencies. If services are provided by multiple units, the appropriate boxes are checked.

	Services not provided	Library provides services	Archives provides services	Records Manage-ment provides services	Others (See below)
Storage for digital information.	X				
Consultation and training services on digital information creation.	X				
Consultation and training services on digital information management.	X				
Consultation and training services on digital information preservation.	X				
Consultation and training services on digital information preservation.	X				
Preservation (e.g., migration, reformatting).	X				
Access (e.g., search engine).	X				
Certification (e.g., trustworthiness of system, backups sufficient).	X				

Other services provided and the units that provide them.
No information provided

Additional information.
As with the Legislature, we have no authority over Judicial at the State level.

Section 3. State Government Digital Information Preservation Activities

This section includes descriptions of past or current digital preservation activities in the state.

TITLE OF THE STATE GOVERNMENT DIGITAL INFORMATION PRESERVATION ACTIVITY:
Web Capture

SHORT DESCRIPTION:
During the Ridge Administration, there was a significant increase in the Web presence with emphasis on using the internet to conduct government business. In September 2001 Governor Ridge announced he would be leaving his post as Governor of Pennsylvania to move on to Washington to become the head of National Homeland Security. At that time the OA and PHMC conducted a 'Web capture' of the PA Web presence beginning with the PA Portal and capturing 4 - 5 levels for each agency, dependent on agency and Web site structure. Some agencies were using Dynamic Site Framework (DSF) and not all of the DSF sites were able to be captured. However, given just a few days notice, it was a successful capture. Lt. Gov. Mark Schweiker became the PA Governor from September 2001 through January 2003. In November of 2002 a second Web capture was taken to document e-government business on the web for the Schweiker Administration.

LINKS TO RELATED WEB SITES OR PROJECT DOCUMENTS:
The Web captures are not available via a Web site. The captures are stored on media in the Archives and can be loaded upon request in the search room.

CONTACT FOR THIS ACTIVITY:
Linda F. Avetta
717-705-6923 or lavetta@state.pa.us

TITLE OF THE STATE GOVERNMENT DIGITAL INFORMATION PRESERVATION ACTIVITY:
Digital Access Tools

SHORT DESCRIPTION:
The Pennsylvania Historical and Museum Commission (PHMC) in conjunction with the Office of Administration (OA) has developed two Web destinations for the delivery of digital images to remote researchers. The first venture was ARIAS (Archives Records Information Access System) which provides access to more than 1.5 million military service abstract cards spanning the Commonwealth's participation in national defense efforts from the Revolutionary War through World War I. The second venture resulted in the creation of Doc Heritage (Documentary Heritage) which focuses on select documents, people and events. In addition to digital images of the documents (some of which have transcripts), a companion narrative provides historical context for the researcher.

LINKS TO RELATED WEB SITES OR PROJECT DOCUMENTS:
www.digitalarchives.state.pa.us and www.docheritage.state.pa.us

CONTACT FOR THIS ACTIVITY:
Harry F. Parker
(717)783-9873 or haparker@state.pa.us

TITLE OF THE STATE GOVERNMENT DIGITAL INFORMATION PRESERVATION ACTIVITY:
E-mail Training

SHORT DESCRIPTION:
The PHMC, in conjunction with the OA, has a written policy for the management of e-mail that was introduced several years ago. Policy implementation suffered due to the lack of training. This past year the PHMC developed a training program for the management of e-mail. Although the full training program runs 2 hours, a condensed Web version of the training (for managers and supervisors) was developed by PHMC and reviewed and approved by the OA and the Office of General Counsel (OGC). The 18 minute condensed Web-based training is mandatory for managers and supervisors throughout the entire Commonwealth. The 2-hour course continues to be offered free of charge for all Commonwealth employees who may wish to register for the full training.

LINKS TO RELATED WEB SITES OR PROJECT DOCUMENTS:

E-mail management policy - http://www.oa.state.pa.us/oac/cwp/view.asp?A=12&Q=120207

To access the online e-mail training, visit:
http://www.oit.state.pa.us/oaoit/cwp/view.asp?a=738&q=199314&oacNav=|2091|

CONTACT FOR THIS ACTIVITY:
Linda F. Avetta
717-705-6923 lavetta@state.pa.us

TITLE OF THE STATE GOVERNMENT DIGITAL INFORMATION PRESERVATION ACTIVITY:
ITB Policy

SHORT DESCRIPTION:
The Pennsylvania Historical and Museum Commission is heading a team of representatives from other state agencies in developing a new 'Records Management' series of Information Technology Bulletins (ITBs) providing policies and procedures for managing the electronic records of the Commonwealth. This is part of a much larger project of the Commonwealth's Office of Information Technology to update IT standards. The overall approach of the new records management ITBs will emphasize the importance of managing records throughout their life cycle. Procedures will be developed for specific types of records and systems such as e-mail records, electronic documents or publications, Electronic Document Management Systems, and web site records. The ITBs will also include standards for the review of the development of information systems to insure records management is integrated into initial system design instead of an afterthought.

LINKS TO RELATED WEB SITES OR PROJECT DOCUMENTS:
The index to the new ITBs will be at this address when they have been completed and approved:
http://www.oit.state.pa.us/oaoit/cwp/view.asp?a=722&Q=203395&PM=1&oaoitNav=|8305|1821|&oacNav=|2095|

CONTACT FOR THIS ACTIVITY:
Linda F. Avetta, Commonwealth of PA, PHMC, State Archives
717-705-6923 or lavetta@state.pa.us

TITLE OF THE STATE GOVERNMENT DIGITAL INFORMATION PRESERVATION ACTIVITY:
Pennsylvania Documents Checklist

SHORT DESCRIPTION:
The Pennsylvania Documents Checklist contains monographs and serials received by the State Library of Pennsylvania. Many are distributed throughout the state through the Pennsylvania Documents Depository system. Born digital and Web-based documents are included and are referenced by URL. Pennsylvania publications are included in the Checklist as they are cataloged and included in the State Library's library management system and on-line public access catalog.

Preservation of public state paper documents is an ongoing responsibility of the State Library. Procedures are in place to film and digitize older paper documents that are in need of preservation. These include conversion to archival quality microfilm and World Wide Web presentation through the State Library's CONTENTdm server hosted at the Health Sciences Library Consortium in Philadelphia. However, there is currently no procedure to capture or preserve born digital public state documents.

The State Library has created an electronic 'in box' to receive digital documents, and the address was included in a recent management directive. The State Library, however, does not itself have the administrative clout to require creators to submit copies of digital documents to that email address.

LINKS TO RELATED WEB SITES OR PROJECT DOCUMENTS:
http://www.statelibrary.state.pa.us/libraries/cwp/view.asp?a=13&Q=43074

CONTACT FOR THIS ACTIVITY:
Mr. Tom Duszak, Head, Cataloging Section, State Library of Pennsylvania
333 Market Street, Harrisburg, PA 17126-1745
tduszak@state.pa.us or 717-772-5616

Section 4. Training Needs for Digital Preservation Related Activities

Level of training needed for the digital preservation capabilities, specified below.

	Training already provided	Basic training needed	Advanced training needed
Identify the type and amount of digital information throughout the state.		X	
Select and appraise state government information in digital form.		X	
Identify key stakeholders related to specific digital information (other local/state agencies, other states, private sector, etc.).		X	
Negotiate and make agreements with key stakeholders to preserve digital information.		X	
Acquire state government information in digital form for holdings.		X	
Manage state government information in digital form (metadata, reformatting, etc.).		X	
Manage the ingest of digital information into a repository.		X	
Manage the long-term storage of digital information in a repository.		X	
Develop mechanisms to monitor the long-term usability of state government information in digital form.		X	
Make state government information in digital form accessible to users.		X	
Produce a disaster and recovery planning for state government information in digital form.	X		
Manage copyright, security, and other legal issues of relevance to state government digital information.		X	
Other (See below).		X	

Other training needed for digital preservation capability.
Most items are implementation related. Our particular group isn't necessarily doing it, but implementation needs to be broader to include many agencies.

Additional information on existing training programs.
No information provided

Section 5. State Government Digital Information Currently At-Risk #1

This section includes examples of state government digital information that is at-risk of deteriorating or being altered or lost through format or technological obsolescence, policy or procedural gaps, or financial constraints.

DESCRIPTION OF AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Digital information created by Commonwealth Media Services (CMS) and various executive branch agencies.

CONDITIONS CAUSING INFORMATION TO BE AT-RISK:
With the emergence of digital photography and audio recordings, the more traditional and somewhat stable medias used to record public events (in particular the Governor's and Legislative proceedings) are being replaced by newer and more technologically complex systems. At the same time, the technology is becoming more readily accessible to the causal users throughout state government. While CMS is aware of their archival value and have made efforts to ensure their transfer to the Archives, the Commission's ability to retain their functionality is dependent upon an ever decreasing budget for such activities. Add to this the prevailing attitude that 'Every person is their own publisher/producer' and we have the potential for wide-spread loss of audio/visual records.

STRATEGIES CONSIDERED TO PRESERVE AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Currently efforts are underway to partner with CMS and the Pennsylvania Public Television Network (PPTN) to ensure the long term viability of these records. Policy will be provided through OA & PHMC once mechanisms are in place to migrate and perpetuate the content.

DESCRIPTION OF AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Electronic correspondence files created at the cabinet level.

CONDITIONS CAUSING INFORMATION TO BE AT-RISK:
Agencies have embraced technologies that provide them the ability to receive and respond to constituent correspondence and internal memorandum in electronic format. Their efforts to produce an efficient system that empowers all Commonwealth employees brings with it the risk of supplanting traditional filing systems.

STRATEGIES CONSIDERED TO PRESERVE AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
It has been an ongoing objective of the PHMC and the Office of Administration to ensure that agencies manage information created with new technologies with the same level of concern that they have for the appropriate retention of more traditional formats. Systems have been inventoried and training in the proper handling of e-mail have been initiated.

DESCRIPTION OF AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Land records in the local Recorder of Deeds Offices

CONDITIONS CAUSING INFORMATION TO BE AT-RISK:
Deeds and other records relating to the ownership of land are recorded at the county level by Recorders of Deeds. These are arguably the most important records in Pennsylvania. Most, if not all, Recorders are utilizing optical imaging technology to record these documents. Since original deeds are not retained by the office and the recorded deed is the official, legal copy, the recording techniques are of great importance. In the next few months, several test counties will be experimenting with electronic recording of deeds and related land records, including electronic signatures. These record will be integrated with those county's existing imaging systems but will expedite the recording process.

STRATEGIES CONSIDERED TO PRESERVE AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
The County Records Committee, which provides the regulations for retention of county records, including the Recorder of Deeds, requires that backup copies be retained for all permanent records including land records. These copies must be human readable, either in paper form or on microfilm that meets state standards. As far as we can ascertain, most Recorders are in compliance with this regulation. (This is our strategy for preservation of digital information at the county level and it is unlikely to change any time soon)

DESCRIPTION OF AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Geographic Information Systems (GIS)

CONDITIONS CAUSING INFORMATION TO BE AT-RISK:
As with the previous item dealing with land records, GIS contains multiple layers of 'land' graphics / records. Maps as we know them, are practically non-existent and are being replaced by GIS. Most, if not all, agencies who have land-related records are utilizing GIS. Water, waste, roadways, bridges, public utilities, protected land including environmental or historical grounds, and more are now recorded via GIS. The components are layered together and are accessible only using the technology that creates them. As the information is produced in graphical overlays, it is difficult, at best, to get a complete map that can be preserved. Also, the information is changing on a minute by minute basis. There are no published 'maps' to get an absolute picture of a certain period in time. Because of the multiple layers, sizes, and various colors, microfilming is not feasible. Storing these records in a human readable format is virtually impossible.

STRATEGIES CONSIDERED TO PRESERVE AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
As noted above, the components are layered and accessible only by using the technology that creates them. We are at a loss for how to maintain this type of information as a permanent archival record. Because of the multiple layers, sizes, and various colors, microfilming is not feasible.

DESCRIPTION OF AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
Born digital public state documents are not currently being systematically harvested or archived.

CONDITIONS CAUSING INFORMATION TO BE AT-RISK:
Technology has not advanced sufficiently to allow automatic harvesting of Web-based public documents or the automatic generation of administrative and descriptive metadata sufficient to manage the documents. Neither is there statutory authority for the State Library to enforce compliance with a management directive requiring Webmasters and document creators to submit copies of these electronic documents to the State Library, or to notify the Library of the URLs when such documents are published on intranets or the Internet. Resources for harvesting and management of a digital archive manually are not available to the State Library of Pennsylvania. It simply takes too much time to do it by hand one document at a time.

The State Library of Pennsylvania conducted an experiment in the spring of 2005 using a product from OCLC, Inc. branded Digital Archive. The results were promising, but not sufficiently automated.

Please see: Kemper, Ann, Sandra Wolf 'Access to Pennsylvania Documents. ' DTTP, Documents to the People 1 July 2005 30-34. Research Library. ProQuest. State Library of Pennsylvania, Harrisburg, PA. 14 Feb. 2006 <http://www.proquest.com/>

STRATEGIES CONSIDERED TO PRESERVE AT-RISK STATE GOVERNMENT DIGITAL INFORMATION:
The State Library is using digital collection management software (CONTENTdm) to manage its paper to digital preservation activities. We are three months into the utilization of CONTENTdm. It may be possible to use this application as part of an archiving plan to preserve Pennsylvania public digital documents. However, as stated above, automation of the harvesting and metadata generation processes will be necessary if we are to achieve preservation efforts similar to the efficiencies experienced in the acquisition and preservation of paper documents.

Examples of government digital information that was not preserved and is no longer accessible.
PA began creating information in an electronic environment as early as 1958. While most early systems continued to have more traditional formats as their 'copy of record', some files suffered losses as a result of the destruction of both the original input source and the specific mainframe based report application.

The Commission's Korean War Compensations Files suffered just such a loss. Many of the original compensation file applications were destroyed in 1972, as a result of flooding caused by Hurricane Agnes. The Department of Military Affairs was able to provide service to veterans by using a report generated by the Commonwealth's Central Management Information Center (CMIC). When the Department turned over their files, the printout of the report and 4 cartons of lists were all that had survived and as a result of some physical damage to the report, not all the information is accessible today.

Section 6. Enterprise Architecture

The following section describes the state's and corresponding units' awareness of and involvement in their state's Enterprise Architecture efforts.

Aware of state's Enterprise Architecture efforts.
Yes

Involved in state's Enterprise Architecture efforts.
Yes

Nature of involvement in the state's Enterprise Architecture efforts, if appropriate.
We are involved as much as we can be . . . by actively participating on some IT standards committees, EDMS efforts, and Information Domain teams. We are also involved in writing IT policy for the Information Domain Team.

Links to relevant documentation that describes the unit's involvement in the state's Enterprise Architecture efforts.
Enterprise Architecture policy and domain standards -
http://www.oit.state.pa.us/oaoit/cwp/view.asp?a=722&Q=203395&PM=1&oaoitNav=|8305|1821|&oacNav=|2095|

Section 7. Additional Thoughts or Comments

We are very limited on a shoe string budget. With no hardware, software, nor personnel resources to build an e-archives, we do have some preliminary steps in place to design a repository where e-information may be deposited in PDF format for use with the newly proposed PDF-a.