

**Information Technology Policy
Commonwealth of Pennsylvania
Governor's Office of Administration/Office for Information Technology**

ITP Number:	ITP-INFRM004	
ITP Title:	Management of Web Records	
Issued by:	Deputy Secretary for Information Technology	
Date Issued:	September 21, 2007	Date Revised: November 18, 2010
Domain:	Information	
Discipline:	Records Management	
Technology Area:	Information	
Revision History Date:	Description:	
11/18/2010	ITP Refresh	

Abstract:

The purpose of this Information Technology Policy (ITP) is to provide more specific guidance, in accordance with policies established in Management Directive (MD) 210.5, "Records Management," regarding the retention and disposition of Web resources that meet the criteria of a record.

The Commonwealth's presence on the Internet is an important method for providing information and services to citizens and the business community. The Web is a venue for the transaction of business and the delivery of services, both public and private. Record of this activity is necessary for accountability in cyberspace as it is in the more traditional material world.

While planning for a Web site, analysis of the content of the Web site needs to take place to determine which elements are records and the associated retention periods that have been established for each series of records. Web records may include content generated externally by customers, suppliers, or partners, who may submit or exchange information through an agency Web site. Web records may be generated internally and include documents such as policy statements, contracts, or other information on which agency customers or partners rely upon to conduct business. Such internal records may include publications that are Web-only versions and are to adhere to policy set forth in MD 205.23, "Submission of Commonwealth Publications to the State Library for Collection and Distribution to Other Designated Libraries." Whether generated externally or internally the contents of Web sites are to be analyzed and scheduled. Web content management systems are being employed to manage aspects of Web content control such as versioning, security, and work flows, but they also need to be implemented to manage the retention of records.

When planning a Web site in order to meet business needs, an agency needs to make sure its site is trustworthy. Trustworthy Web sites involve numerous qualities. A trustworthy site is one that is reliable and in which the content can be trusted as full, accurate, complete, and unaltered; it is authentic and has been created by the entity for which it is identified; and it is usable - its content can be easily located and retrieved.

An untrustworthy Web site may lead to risks. These risks include the inability to document or validate transactions, the inability to reconstruct views of Web content that were created dynamically, and the loss of versions of policies, procedures or other content that were only posted to the Web with no other copies retained.

Other ITPs and policies may be referenced that impact the construction and maintenance of a Web site and include guidance on accessibility, search engines, file name standards, hardware and software standards, privacy concerns, and Right-to Know policies.

General:

This ITP applies to all departments, boards, commissions and councils under the governor's jurisdiction. Agencies not under the governor's jurisdiction are strongly encouraged to follow this policy.

Policy:

All Web content, including but not limited to forms, policy, charts, graphics, announcements, is to be inventoried and analyzed to define records series, and maintained and disposed of in accordance with approved records retention and disposition schedules.

The life cycle of Web records (reference ITP INFRM001, *Life Cycle of Records*) is to be determined prior to or during creation and managed according to established policies and procedures.

These policies are to be based upon the business needs of the agency to either retain content or destroy content. Destroying content which is no longer useful reduces costs, confusion, and contradictions. Likewise, maintaining content systematically and with appropriate metadata reduces the cost of locating information when it is needed and allows the use of more automation in managing records and thus reduces costs.

No Web applications or content are to be developed without taking into consideration the life cycle, retention, and disposition of the resulting records. If Web content is generated via a Web application, ITP INFRM005, *System Design Review of Electronic Systems*, applies.

Retention and Scheduling of Web Records

Some Web content may already be governed by existing records retention and disposition schedules. For example, Web records may be covered under general records retention and disposition schedules and, in particular, the General Administrative Schedule. Some Web records may be duplicate copies of records maintained in another format or system and, thus, already scheduled on existing agency-specific records retention and disposition schedules.

Some Web content may only be partially governed by existing schedules or may be entirely new records. For example, databases that operate with a Web interface are to be scheduled as part of the process of scheduling the records within the database, mentioning that they are Web interactive. In some cases the Web interface or data entry forms related to such a database are to be new records and are to be scheduled separately if they govern the content of the database. In particular, versions of a Web form used to input data into a database are to be retained as long as the data is retained as part of the documentation of how that data was generated.

When determining retention periods for Web site-related records, as with other records, the agency needs to assess how long the information is needed to satisfy business needs and provide government accountability. If specific Web content is available in places other than the Web site, consider whether the existence of the information in other records affects the retention needs for the Web records. If the information is unique to the Web site, the Web version is, by definition, the official record copy and needs to be scheduled.

Several options exist when analyzing a Web site being added to the records retention and disposition schedule, depending on risk factors. An agency may want to breakdown the Web site; however, this is not recommended because most Web sites have a variety of content types and each is to be evaluated separately as to possible records management risks. A Web site may be analyzed as a whole, minus one or two portions that may have different characteristics, such as an e-government application. An agency may also need to break down the Web site by function for evaluation and scheduling purposes. This could result in a large number of independent records series. Whatever specific process is used, the entire Web site is to ultimately be inventoried with records series defined and scheduled.

As an example, some Web sites may have very dynamic content that may be a record (i.e., Web pages that reflect the status of various systems). If the information is used as the basis for decisions, then a log is to be created and maintained as the record. If the information is similar to a status report, then it is to

be kept only as long as it is useful. All sorts of transitory announcements are entered or displayed on Web sites (i.e., jobs announcements and conferences) and, in most cases, it is important to keep a copy of both the announcement and the date the copy was posted and removed. Such records are to be scheduled. In many cases the Web announcement is not to be the official copy, but it may still be important to retain the knowledge of when it was posted and removed from the Web site for a period of time.

The inventory process is to be coordinated with the agency Web content managers and the agency records coordinator, as well as the Web site manager/Web master for implementation.

Records Management Requirements for Web Sites

- All agency Web sites are to be inventoried to define records series and establish records retention and disposition schedules for all records on the sites.
- When records on a Web site duplicate records stored in other formats or systems, a clear determination is to be made regarding which copy is the duplicate and which is the official record copy that is subject to the records retention and disposition schedule.
- The duplicate- and information-only Web pages are to be destroyed when they are no longer needed as indicated on the General Administration Records Retention and Disposition Schedule.
- As Web content management systems are implemented, they are to be integrated with records management systems, if possible, or have records management requirements specified in the content management system as records are created or ingested. For example, expiration dates for documents are to correspond to retention periods (unless the document on the Web site is a duplicate as indicated in the previous paragraph).
- When record series of official record copy containing long-term and archival records exist on a Web site, special arrangements are to be made to preserve them. They may be maintained in a separate system or format or removed to such a system after their active life on the Web site. If maintained on the Web site itself, security and access rights as well as backup provisions are to be more rigorous. For Web publications, if a permanent copy is to be kept on the agency servers, the State Library is to accept a persistent URL (i.e., a permanent link to the document) instead of the document itself according to the policy set forth in MD 205.23. The link is to be permanently maintained with redirects to its current location from the old address.
- Version control is especially important with policies, contracts, and other documents, available via the Web, which are updated or revised. This is critical for any document in which visitors to the Web site may base business decisions or actions. The date and time that specific records are posted and removed may be needed for business or administrative needs. Revised or rescinded policies are to be retained according to the Records Retention and Disposition Schedule(s). Policies will generally be archived and subject to permanent retention. It is recommended that agencies remove revised or rescinded policies from public-facing Web sites in order to prevent those policies from being mistaken with current policies. Agencies are to make provisions to keep copies of each version of policies along with the dates in which the policies were issued/revised/rescinded utilizing a configuration management system, document management system, agency file shares, or hard copy. Likewise, Web-based forms or other documents related to the entry of information into a database are to be maintained as long as the content generated by the forms is scheduled to be retained. For Web publications, all Web-only document versions are to be forwarded to the State Library of Pennsylvania as set forth in MD 205.23.

Refresh Schedule:

All standards identified in this ITP are subject to review and possible revision annually or upon request by the Enterprise Architecture Standards Committee.

Exemption from This Policy:

In the event an agency chooses to seek an exemption, for reasons such as the need to comply with requirements for a federally mandated system, a request for waiver may be submitted via the Commonwealth of PA Procurement and Architectural Review (COPPAR) process. Requests are to be entered into the COPPAR Tool located at <http://coppa.oe.pa.gov/>. Agency CIO approval is required. Contact your agency [CoP Planner](#) for further details or assistance.

Questions:

Questions regarding this policy are to be directed to RA-ITCentral@pa.gov.

References:

MD 205.23: Submission of Commonwealth Publications to the State Library for Collection and Distribution

MD 210.5: Records Management

ITP-INFRM001: The Life Cycle of Records: General Policy Statement

ITP-INFRM005: System Design Review of Electronic Systems