

Guidelines for Managing Web Site Content: Appendix A

Object Driven Approach Implementation Strategy 2

Snapshots

A snapshot usually involves creating a full and accurate record copy of an agency's public web resources at a particular point in time. If an agency decides to create period snapshots of a web site, the snapshots should be scheduled on a retention schedule. The snapshot should be maintained for the length of the retention period.

When taking snapshots of collections of web resources, it is desirable to ensure (as far as possible) the continuing processability of the website and its component pages. This means that agencies should try to retain the capability to replicate the content, layout and functionality of the site across technological platforms without loss of data integrity.

This strategy is particularly useful for static resources or collections of static objects that are essentially an agency's electronic publication(s).

A snapshot is an object-driven approach and should not be used to keep records of highly interactive dynamic sites or resources that are databases or transactional services. A deficiency of this approach is that a snapshot only provides a picture of a website at a particular point in time. If snapshots are captured in the absence of other records of web-based activity, it will be impossible to reconstruct the site together with its functionality at any other point in time. Since this method does not enable the agency to determine exactly when particular web resources were available, agencies that use the snapshots strategy should also create and maintain logs of changes made to web resources between snapshots.

Procedures for creating and capturing a snapshot

A snapshot should include all aspects of the website to ensure that a fully functional site can be reconstructed. For example, the snapshot should also include scripts, programs, plug-ins and browser software, that is, all components that make the snapshot fully functional. The snapshot should be captured with sufficient descriptive metadata.

It may be necessary to make some modifications once the snapshot is created. For example, a CGI script for site counters will need to be disabled. If site counters are not disabled, there will be no accurate or authentic record of the number of visitors to the site at the time the snapshot was created. In effect, the record is no longer a snapshot of the site.

Responsibilities

The main responsibilities to assign include:

- Determining if this is an appropriate option and whether it should be supplemented by other record management strategies;
- Determining how frequently copies of web resources should be created;
- Creating the snapshot;
- Capturing and maintaining of sufficient metadata for the length of the retention

period;

- Selecting an appropriate storage medium and undertaking data management tasks and quality control check.

Website administrators or information technology staff may already carry out the task of creating 'back-ups' of the website as part of normal data management activities. However, because these back-up copies are created for the purpose of data management activities, they are usually overwritten regularly with more recent versions, or deleted. They are not captured or maintained for record management purposes. To be used as a viable records management strategy, it is necessary to intervene and establish processes and procedures to ensure that snapshots are created, captured and maintained over time for as long as required. The above responsibilities represent the minimum list of responsibilities that should be documented in agency's procedures and assigned to records management practitioners, website administrators and information technology staff.

Tracking changes

This strategy involves tracking changes to the web resources over time and creating a log of changes or activity. The activity log needs to be maintained to satisfy requirements for accessibility for as long as needed. Used in combination with snapshots of the web resources, this approach can be a reliable option for static sites.

The main problem arising from this option is the creation of insufficient metadata of the activity log, resulting in the inability to interpret the log over time. It is vital that metadata requirements are specified and sufficient metadata is captured.

Procedures for creating and capturing activity logs

Suggested data elements that can be captured in an activity log include:

- Title or name of posting;
- Version number;
- Author or content manager responsible for creating of the object;
- Links embedded in the posting;
- Date of initial posting;
- Date of modification;
- Date of replacement or withdrawal; and
- Disposal information.

This is not a complete list and agencies should review and adapt it to ensure their requirements are satisfied.

In the case of a static website, the log should capture changes to individual pages, documents or objects on the website. Changes to scripts, plug-ins, forms used to present information etc will also need to be captured as they will affect the functionality of the records.

It may be possible to use emerging web technologies to track changes. Web robots, spiders or crawlers are automated programs that visit sites for the purpose of indexing sites for search engines. These programs may be useful for tracking changes.

Responsibilities

The main responsibilities to assign include:

- Determining the list of data elements that should be captured in an activity log;
- Establishing procedures and processes to ensure the activity log is created, updated and maintained over time;
- Capturing and maintaining the activity log, including the capture and maintenance of sufficient metadata;
- Selecting an appropriate storage medium and undertaking data management tasks; and
- Identifying preservation implications and ensuring the records are accessible for as long as required.

[Appendix B](#)

[Go to Table of Contents](#)