

Policies of Managing Web Resources at the Canadian Government: A Records Management Perspective

Natasha Zwarich

Ph. D. Student

natasha.zwarich@mail.mcgill.ca

Eun G. Park

Assistant Professor

eun.park@mcgill.ca

School of Information Studies

McGill University

Montreal, Canada

Abstract

As the importance of Web resources has increased greatly to organizations, this study examines the Web management policies and guidelines of five Canadian government agencies, studies the definitions and characteristics of Web resources from a records management perspective and presents the current policies relevant to the appraisal, retention, and preservation of Web resources in Canadian government agencies.

1. Introduction

A variety of institutions including government agencies provide online information services for the

public and conduct their main operations through Web application programs. As a result, a large number of Web documents are generated everyday and those institutions need to constitute traces of operational activities. Resources linked with or created by the Web, some of the largest documents ever created, are a unique phenomenon. One research study reports that more than four billion public Web pages exist, with an additional 550 billion connected documents (Lyman, 2002).

2. Background

2.1 Definitions and Characteristics of the Web

A Web site can be defined as a collection of electronic resources that are made available in a particular domain of the Internet for the purpose of information communication and/or the conduct of business transactions (Treasury Board of Canada, 2002). It shares a common domain name, normally belonging to a single organization or a defined group of organizations. It retains a Uniform Resource Identifier as its virtual location and many hierarchically related links to its main domain content, often referred to as the home page. It provides a body of interlinked information resources that users may navigate using browser technology.

International recordkeeping laws and stan-

dards do not specifically address the Web as electronic records. However, all recognized definitions of the term “records” embrace or, at least, do not exclude electronic records. For example, *ISO 15489 Information and Documentation – Records Management* is applicable to electronic records because it “applies to the management of records, in all formats or media, created or received by any public or private organization in the conduct of its activities” (2001, p. 7). Records are evidence of business and administrative transactions and are essential resources for accountability and organizational memory. The content of these records can determine the acts of the users who access them. Applying traditional records management principles to Web site maintenance is still a valid approach to ensure whether records published on the site are controlled and can meet the requirements for organizational accountability.

Records are evidence of business and administrative transactions and are essential resources for accountability and organizational memory. The rise of electronic government, e-commerce, and general use of Web sites to communicate official ordinances has resulted in an increasing number of records existing solely online, and being published only in that form. Traditional paper-based record-keeping infrastructures commonly fail to identify, con-

trol, and capture records published in electronic format, particularly when responsibility for creation, control, dissemination and maintenance of records is widely distributed (Pennock & Kelly, 2006). This is a problem because records must meet certain requirements in order to qualify as evidence. Web resources, like other electronic records, meet the characteristics of trustworthy records: *reliability, authenticity, integrity and usability* (ISO 15489, 2001, p. 7). A reliable web site is one whose content can be trusted as a full and accurate representation of the transactions or activities to which it attests and therefore can be depended upon in the course of subsequent transactions or activities. An authentic web site is one that is proven to be what it purports to be and to have been created by the agency with which it is identified (NARA, 2005). To demonstrate the authenticity of a web site, agencies should implement and document policies and procedures that control the creation, transmission, receipt, and maintenance of web site records to ensure that records creators are authorized and identified and that records are protected against unauthorized addition, deletion, and alteration. The integrity of a web content record refers to it being complete and unaltered. The agency's Web management policies and procedures for updating and modifying their web sites help ensure integri-

ty. Another aspect of this concept is the structural integrity of a Web site's content-related records. The structure of a Web site, that is, its physical and logical format and the relationships between the pages and content elements composing the site, should remain intact. Failure to maintain the Web site's structural integrity may impair its reliability and authenticity. A usable Web site is one that can be located, retrieved, presented, and interpreted (NARA, 2005). In retrieval and use, one should be able to directly connect the Web site to the business activity or transaction that produced it. Both the site and its content within the context of broader business activities and functions should be identifiable. The links between content, contextual, and structural Web site-related records that document agency Web site activities should be maintained. These contextual relations should provide an understanding of the transactions that created and used them. Application of records management principles to Web site maintenance ensures that records published on the site are controlled and can meet the requirements for organizational accountability.

2.2 Literature Review

A few researchers and institutions have already recognized the importance of Web resources.

Barry specifically identifies the Web as a recordkeeping system containing important business contents as the result of business operations (Barry, 2004). The Smithsonian Institution Archives commissioned a study to conduct a high-level assessment of requirements for preserving their Web sites and HTML pages and develop a strategy, guidelines, and best practices to facilitate access to usable and trustworthy Web sites and HTML pages (Dollar, 2001). McClure and Sprehe conducted broad research to develop better records management and preservation strategies for electronic information available on government agency Web sites (McClure & Sprehe, 1998). In addition, several studies approach Web sites from a preservation perspective (Martin & Eubank, 2007; Hodge, et al., 2004; Kenney, et al., 2002). Considering the importance of archiving all Web resources, many Web archiving projects use a harvesting approach by crawling Web contents and adding metadata about resources to the sites. For example, the Internet Archive project attempts to collect and archive the entire Internet (Pennock & Kelly, 2006).

However, there are not many studies that consider Web records from a records management perspective, which would provide a systematic approach to the organization and management of the Web as records.

3. Methodology and Findings

The objective of this study is to explore the current management of Web sites as records and recordkeeping systems. This study will take a close look at the Web management policies and guidelines of five Canadian government agencies, including Department of Justice Canada, Library and Archives Canada (LAC), Health Canada, Public Works and Government Services Canada, and Treasury Board of Canada. This study will conduct content analysis to investigate how these Canadian government agencies currently handle Web documents and how they implement guidelines. Specifically this study will: 1) examine the definition of Web sites as records; and 2) examine the current policies and practices of managing Web records in Canadian government agencies.

3.1 Overall Guides of the Government of Canada

Currently, the majority of organizations, including government agencies, have set up their Web sites to provide information for users and conduct official business activities. As the number of Web sites increases, so does the importance of accountability, evidence and reliability whenever creating recorded information. Government agencies need to be aware that, by using the Web for business opera-

tions, they are creating records. Just as they would manage other non-Web-based resources that they create in the conduct of their business, they have an obligation to manage the Web-based information as well.

In 1999, the Government of Canada published *An Approach to Managing Internet and Intranet Information for Long-term Access and Accountability*, which was developed by the Internet and Intranet Working Group (IIWG) (Government of Canada, 1999). This policy provided government-wide guidance on managing records and publications on the Internet and on departmental intranets and extranets. The guidelines have evolved out of the recognition that all employees in the Government of Canada are accountable for the information they produce and disseminate. It is the principal means of showing how they have fulfilled their obligations and thus holds them accountable for their actions. As for records, this ensures that records which are posted to the Web are captured and/or described as evidence in a corporate record keeping system, and managed from creation through disposition in keeping with the intent of the National Archives of Canada Act (1987) and the provisions of the Treasury Board Policy on the Management of Government Information Holdings (Government of Canada, 1999).

Library and Archives Canada also mentions that Web-related files and documents are the same as any other organizational activity that creates and maintains records. Therefore, Web-related files and documentation are often official records and administration of these records must be incorporated into overall organizational records management activities. The first step to assert is to place Web-related files and documents within the agency.

3.2 Appraisal and Retention of Web Resources

According to the Implementation Guide created by the Treasury Board of Canada, a Web schedule should cover Web content records that document the information on the site itself (Treasury Board of Canada, 2002). A Web schedule should include Web site management and operations records, which provide the site's context and structure. For example, Web content records include the following: the content pages that compose the site, inclusive of the HTML markup; records generated when a user interacts with a site; and, if the agency chooses to document its site this way, lists of the URLs referenced by the site's hyperlinks. Web management and operations records that provide context to the site include: Web site design records, records that specify an agency's Web policies and proce-

dures by addressing such matters as how records are selected for the site and when and how they may be removed, records documenting the use of copyrighted material on a site, records relating to the software applications used to operate the site, and records that document user access and when pages are placed on the site, updated, and/or removed. Finally, Web management and operations records that provide structure related to the site include: site maps that show the directory structure into which content pages are organized and software configuration files used to operate the site and establish its look and feel, including server environment configuration specifications.

A Web site is often a singular collection that can provide important evidential and informational value. With its metadata tags and links, a Web version of a document differs significantly from hard copy and other versions. Most sites do not provide for secure filing and cannot guarantee that all information presented is complete or accurate. The starting point in appraising Web records is that not all data needs to be retained: the basic premise of records management is that it is neither necessary nor desirable to retain everything.

Then, the first step in implementing proper records retention schedules for Web-based records

is to identify which records should be preserved and for how long against that which should be destroyed. Retention periods specify the length of time information must be retained. Different categories of information have differing retention periods. Records management guidelines and retention periods must be applied to Web-based data to ensure that legal obligations are met for records posted on both the Internet and Intranets as well.

The *Approach to Managing Internet and Intranet Information for Long-term Access and Accountability* reminds institutions that all records posted to a departmental Web site are subject to the requirements of the National Archives of Canada Act and should therefore be managed in accordance with retention guidelines issued by the institution and approved disposition authorities issued by the National Archives (Government of Canada, 1999).

3.3 Capture and Preservation of Web Resources

None of the Canadian government agencies mentioned if an electronic records management system was used to manage their Web-based records. Nevertheless, the Government of Canada decided to develop an integrated system called Records, Document and Information Management System (RDIMS; Public Works and Government Services Canada

[PWGSC], 2006). RDIMS grew out of the Treasury Board Shared Systems Initiative in the mid-1990s in collaboration with Library and Archives Canada in order to support streamlined information technology and information management for all government administration and information services. Through RDIMS, all kinds of documents and records can be, and are supposed to be, created, managed, searched, preserved, and accessed, including Web documents, e-mail messages, and all other kinds of documents in electronic form (Hummingbird Ltd., 2006). We assume that Canadian government agencies use RDIMS to manage their Web resources.

The Treasury Board of Canada mentions that the agency's action plan for preserving Internet and electronic information should include plans for preserving its origin, authenticity, functionality, context, presentation and content (Treasury Board of Canada, 2002). Library and Archives Canada, on the other hand, mentioned that the Internet Archive, a non-profit organization that was founded to build an Internet library with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format, maintains an archive of publicly accessible Web pages from around the world, including Canadian Web sites (Library and Archives Canada, 2006).

The Treasury Board of Canada has developed an automated system that crawls the Web every few months and archives new Web sites or those that have changed since the previous crawl (Treasury Board of Canada, 2002). They also offer a service for users to browse through the archive, called the Wayback Machine. LAC also archives some Canadian Web content on a selective basis. Despite these initiatives, no perfect solution has yet been devised to address the technical challenges of collecting and preserving Web site content (Library and Archives Canada, 2006).

4. Conclusion

This investigation provides an overview summary of current Web site management policies at five Canadian government agencies from a records management perspective. These government agencies are working hard to implement Web management policies to solve problems related to Web resources. However, although the importance of Web resources has been recognized widely, we assume that current management is still at an initial stage and needs to develop further. In addition, the findings of this pilot study demonstrate the need for further investigation involving a greater number of government agencies, as well as an in-depth analysis. Through a

consideration of Web resources as a kind of electronic record, we see that this is one of the most challenging issues of the digital age to government agencies, and will increase in importance to information professionals in coming years.

References and bibliography

- Barry, Rick. 2004. "Web Sites as Recordkeeping & Recordmaking Systems". *The Information Management Journal*. November/December, 38, no 6: 26-32.
- Dollar, Charles. 2001. *Archival Preservation of Smithsonian Web Resources: Strategies, Principles, and Best Practices*. <<http://www.si.edu/archives/archives/dollar%20report.html>>.
- Government of Canada. 1999. *An Approach to Managing Internet and Intranet Information for Long Term Access and Accountability*. <http://www.imforumgi.gc.ca/documents/1999/approachapproche_e.asp>.
- Hodge, Gail, Nikkia Anderson, Daniel Smith Jr., Alessandro Senserini & Robert B. Allen. 2004. "Archiving and Accessing Web Pages. The Goddard Library Web Capture Project". *D-Lib Magazine*, 10, no 11, November 2004. <<http://www.dlib.org/dlib/november04/hodge/11hodge.html>>.
- International Standards Organization. 2001. *ISO 15489-1, Information and Documentation – Records Management – Part 1: General*. Geneva, Switzerland: ISO, 2001.
- Kenny, Anne R., Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze & Sandra Payette. 2002. "Preservation Risk Management for Web Resources". *D-Lib Magazine*, 8, no 1, January 2002. <<http://www.dlib.org/dlib/january02/kenney/01kenney.html>>.
- Library and Archives Canada. 2006. *Canadian Digital*

Information Strategy. Mapping the Current Situation in Canada.

<<http://www.collectionscanada.gc.ca/cdis/012033-700-e.html>>.

Lyman, Peter. 2002. Archiving the World Wide Web. In *Building a National strategy for digital preservation: issues in digital media archiving*. Washington: Library of Congress, p. 38-51. <<http://www.clir.org/pubs/reports/pub106/pub106.pdf>>.

Martin, Kristin E. & Kelly Eubank. 2007. "The North Carolina State Government Website Archives: A case study of an American government Web archiving project". *New Review of Hypermedia and Multimedia*, 13, no 1: 7-26.

McClure, Charles R. & J. Timothy Sprehe. 1998. *Guidelines for electronic records management on state federal agency Websites*. <http://www.ii.fsu.edu/~cmclure/nhprc/nhprc_toc.html>.

National Archives and Records Administration (NARA). 2005. *NARA Guidance on Managing Web Records*. <<http://www.archives.gov/records-mgmt/policy/managing-web-records-index.html>>.

Pennock, Maureen & Brian Kelly. 2006. "Archiving Web Site Resources: A Records Management View". *Poster presented at WWW2006, 15th International World Wide Web Conference*. <<http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/www2006/>>.

Public Works and Government Services Canada. 2006. *About RDIMS*. <<http://www.pwgsc.gc.ca/rdims/text/about-e.html>>.

Treasury Board of Canada Secretariat. 2002. *Government of Canada Internet Guide*. <http://www.tbs-sct.gc.ca/ig-gi/gi-gs-cd/pur-obj/pur-obj_e.asp>.

Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British Columbia, Vancouver, June 5-7, 2008