CAIS Paper: Rotten by Design: Shortened Expiry Dates for Government of Canada Web Content

Authors: Michael B. McNally, Amanda Wakaruk, Danoosh Davoodi

Affiliations:

Michael B. McNally – Assistant Professor, School of Library and Information Studies, University of Alberta

Amanda Wakaruk – Acting Digital Repository Services, Government Information Librarian, University of Alberta Libraries, University of Alberta

Danoosh Davoodi – Masters of Library and Information Studies Candidate, School of Library and Information Studies, University of Alberta

Abstract: This paper examines the Government of Canada's Web Renewal Initiative and impact on the provision of web based government information. A content analysis of documents received through Access to Information requests reveals the government's limited timelines for providing access to web content.

1. Introduction

In the same way the passage of time will cause ripened food to spoil and rot, government information on the internet may become redundant or outdated and URLs may become inactive - a phenomenon known as link rot. On Nov. 6, 2012, Tony Clement, President of the Treasury Board of Canada Secretariat (TBS), announced that the Government of Canada (GoC) would consolidate the number of government websites from over 1500 to six or less as part of a Web Renewal Initiative (WRI) (Clement, 2012). While the change is supposed to ensure government information is easier for Canadians to locate, two significant outcomes will result. Not only will the WRI cause numerous URLs to become broken (link rot), the government actively plans on eliminating redundant, outdated and trivial content (or what the GoC itself calls 'ROT') (TBS 2012b). These changes will have a significant impact on Canadians' ability to access government information. The impact will be particularly acute for researchers who will no longer have access to historical GoC web content essential for scrutinizing government policy and activities. The WRI has already begun, and changes in URLs and the elimination of content is underway (TBS 2012a).

This paper examines the GoC's own guidelines for eliminating ROT through a content analysis of guidelines provided by the TBS and two federal departments (Industry Canada (IC) and Citizenship and Immigration Canada (CIC)) obtained through Access to Information requests (ATI). The paper is part of a larger study aimed at identifying and assessing the removal of government web content (though the analysis in this paper focuses exclusively on internal government guidelines for identifying and removing ROT). The analysis of departmental guidelines obtained through ATI is particularly important as this information is not published by the government.

2. Literature Review

Concerns over link rot on the Internet are not new (Fichter, 1999; Tyler & McNeil, 2003; Koehler, 2004), and the subject has been of particular concern to American legal scholars (Rhodes, 2010; Torres, 2012; Liebler and Liebert, 2013; Zittrain et al. 2013). While the findings of U.S. legal scholars does not address GoC link rot, it does highlight the importance and dangers of link rot. For example Zittrain et al. found that more than 70% of the URLs within a sample of legal journals were subject to link rot (2013, p. 167). Liebler and Liebert found that 29% of URLs within U.S. Supreme Court decisions suffered from link rot (2013, p. 278).

It is important to note that link rot is not a phenomenon unique to the United States or legal scholarship, and has been observed elsewhere. A study found that in a sample of New Zealand academic journals that 30% of electronic citations for articles from 2002-2005 were inaccessible by 2006 (Parker 2007). Carnevale and Aronsky found that 30% of URLs in a sample of biomedical informatics journals were subject to link rot (2007, p. 271). However, while there are studies noting problems in government information in Canada (see especially Wakaruk, 2014) no studies to date have examined GoC link rot, which is particularly timely given the WRI.

3. Methodology

The paper employs an inductive content analysis (Crano et al. 2015) of three sets of government documents produced through ATI Requests. ATI requests to IC, CIC and Health Canada[i] were made,[ii] along with a fourth ATI request to TBS[iii] as TBS has the responsibility for providing guidance on the WRI to government departments. The results of the ATI requests varied considerably in size (for example CIC provided only three pages, IC's response was 25 pages, and TBS produced 134 pages). Each response was thematically coded using an inductive coding system of manifest content (Crano et al. 2015, Rubin and Babbie, 2010) to identify themes in departmental guidance for identifying and removing ROT content, timelines for the removal of ROT, and any discrepancies in approaches between departments. The results provide an insightful view of how the GoC is restructuring and removing its web content.

4. Preliminary Findings and Discussion

The analysis of departmental guidance on removing ROT provides several important findings. First, there is considerable difference between CIC and IC with regards to how much information is provided to departmental employees with regard to identifying and removing ROT. CIC provides only a brief description of the problem of ROT and a set of timelines for when various types of content should be removed (e.g. news items and yearly/quarterly publications are to be removed two years from the date of publishing) (CIC n.d.). In contrast, Industry Canada provides much greater detail on the process of identifying and removing ROT to its employees including web content lifecycle flowcharts, a content assessment questionnaire, checklists and timelines. (IC n.d.). While both departments provide timelines, in some cases the timelines for removal

of content can be quite short. For example, IC suggests content about a retired or cancelled program be removed after one year (IC n.d.). Such short timelines mean that historic program information may be unavailable to researchers who seek to assess the effectiveness of government programs. Also, there is little consistency between departments as IC suggests removing newsletters when no longer of interest to clients (based on the use of web metrics with no threshold specified) and event specific items after six months (IC n.d.), while CIC suggests that such content be removed after two years (CIC n.d.).

The documents also indicate the importance of web metrics in identifying ROT content. IC's content assessment questionnaire suggests employees use web metrics to determine if removing content "will either cause harm or great inconvenience to the public OR decrease IC's efficiency or effectiveness to provide a service" (IC n.d., p. 12). CIC also employs web metrics in its timelines for document removals with several classes of web content (including audits, evaluations, research and statistics) to be deleted after two years unless the document receives 200 visits per month. While the use of web metrics may be useful in identifying seldom used web content, CIC's 200 visits per month metric is arbitrary. Furthermore, web visits per month may vary over the course of a year from month to month, and most importantly, a document can still be an invaluable information source for a few groups or individuals and never receive 200 visits per month.

The TBS guidance on link rot reveals some political considerations in the ROT removal process. For example TBS advises departments not to remove content when the removal may be "politically embarrassing" (TBS n.d., p. 75). TBS also highlights removal of large quantities of government information as success stories. For example, Transport Canada's removal of 40% of its web content and Canada Revenue Agency's Pacific Region removal of 9,949 pages (of a total of 10,671 pages) are identified as successes under the marker "Mission Possible !!" (TBS n.d., p. 85 and 93). The TBS documents suggest a government that is actively encouraging the removal of voluminous amounts of web content, though mindful of how such removals may affect the political fortunes of the GoC.

Finally one key finding from the IC documents, is that the ROT assessment process began in the spring of 2011 if not earlier (IC n.d.), well before Minister Clement's November 2012 announcement, and publication of the ROT information on the TBS website (TBS 2012). This preliminary analysis will be supplemented with further analysis of Health Canada departmental guidance on link rot when (of if) the ATI request is completed, and the discussion of additional themes (which space here does not permit).

5. Conclusion

The analysis of GoC departmental guidance around identifying and removing ROTten web content highlights the dangers of government information on the internet. With the TBS aggressively pushing to remove web content, and highlighting "success stories" where huge swaths of web content are removed, invaluable government information is at risk. The WRI initiative underscores the importance of third-party web-archiving. In this regard Canadian academic libraries, which already engage in GoC web

archiving initiatives in partnership with the Internet Archive are preforming a crucial role in maintaining their role as stewards of government information. Without such archives, the public, corporations, non-governmental organizations, journalists and academics may lose access to government information and be unable to hold governments accountable. This concern is perhaps best echoed by the quote (from Stephen Harper while leader of the Official Opposition):

> Information is the lifeblood of a democracy. Without adequate access to key information about government policies and programs, citizens and parliamentarians cannot make informed decisions, and incompetent or corrupt governance can be hidden under a cloak of secrecy. (Harper 2005).

The study does have some limitations. Documents were coded only a single individual. Further the study did not submit ATI requests to all federal departments, and within the sample of departments (IC, CIC and Health Canada), one department (Health Canada) has still not produced any documents as a result of the request.

References

Canada - Citizenship and Immigration Canada (CIC). N.d. "Archiving Strategy." [Document released through Access to Information, on record with the author].

Canada - Industry Canada (IC). N.d. "ICWeb: Internet Presence Policy." [Document released through Access to Information, on record with the author].

Canada - Treasury Board of Canada Secretariat (TBS). 2012a. *Action Plan for the Renewal of the GC Web Presence: Presentation for Senior Department Officials for the Web, Nov 28th, 2012*. British Columbia Freedom of Information and Privacy Association. http://fipa.bc.ca/library/Government%20Documents/GoC_web_plan_Part1.pdf

Canada - TBS. 2012b. "Reduce Redundant, Outdated and Trivial Content." https://www.tbs-sct.gc.ca/ws-nw/wu-fe/rot-rid/index-eng.asp

Canada – TBS. N.d. "Web Archiving." [Document released through Access to Information, on record with the author].

Carnevale, Randy J. and Dominik Aronsky. 2007. "The Life and Death of URLs in Five Biomedical Informatics Journals." *International Journal of Medical Informatics*, 76: 269-273.

Clement, Tony. 2012. "Using Technology to Challenge the Status Quo in Government Operations." *Treasury Board Secretariat of Canada*. Nov 6, 2012: http://www.tbs-sct.gc.ca/media/ps-dp/2012/1106a-eng.asp

Crano, William D., Marilynn B. Brewer, and Andrew Lac. 2015. Principles and Methods of Social Research. 3rd Ed. New York: Routledge.

Fichter, Darlene. 1999. "Do I Look Like a Maid? Strategies for Preventing Link Rot." *Online*, Sept/Oct 1999: p. 77-79.

Harper, Stephen. 2005. "Cleaning up the Mess in Ottawa: Transparency is Key to Preventing Scandal." *Montreal Gazette*. 7 Jun. 2005, p. A21.

Koehler, Wallace. 2004. "A Longitudinal Study of Web Pages Continued: A Consideration of Document Persistence." *Information Research*, 9(2): http://www.informationr.net/ir/9-2/paper174.html

Liebler, Raizel, and June Liebert. 2013. "Something Rotten in the State of Legal Citation: The Life Span of a United States Supreme Court Citation Containing an Internet Link (1996-2010)." Yale Journal of Law and Technology, 15(2): p. 273-311.

Parker, Ailsa. 2007. "Link Rot: How the Inaccessibility of Electronic Citations Affects the Quality of New Zealand Scholarly Literature." *Library Journal Articles*: http://repository.digitalnz.org/system/uploads/record/attachment/215/link_rot__how_the_inaccessibility_of_electronic_citations_affects_the_quality_of_new_zealand_scholarly_literature.pdf

Rhodes, Sarah. 2010. "Breaking down Link Rot: The Chesapeake Project Legal Information Archive's Examination of URL Stability." *Law Library Journal*, 102(4): p. 581-597.

Rubin, Allen, and Earl Babbie. 2010. Essential Research Methods for Social Work. 2nd Ed. Belmont, CA: Brooks/Cole.

Torres, Arturo. 2012. "Is Link Rot Destroying Stare Decisis as We Know It? The Internet-Citation Practice of the Texas Appellate Courts" *Journal of Appellate Practice and Process*, 13(2): 269-299.

Tyler, David, and Beth McNeil. 2003. "Librarians and Link Rot: A Comparative Analysis with Some Methodological Considerations." *Portal: Libraries and the Academy*, 3(4): 615-632.

Wakaruk, Amanda. 2014. "What the Heck is Happening up North? Canadian Government Information Circa 2014." *Documents to the People*, 42(1): 15-20.

Zittrain, Jonathan, Kendra Albert, and Lawrence Lessig. 2013. "Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations." *Harvard Public Law Working Paper*, No. 13-42: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2329161

---

[i] While a request was made to Health Canada on July 25, 2014, the department responded by granting itself a 150 day extension beyond the statutory requirement that request be filled in 30 days. As of January 2015, the request has still not been completed.

[ii] The text of the Access to Information Request (for each department) was: "Departmental guidelines, practices, frameworks and/or rubrics for assessing redundant, outdated or trivial web content developed in relation to the Treasury Board Secretariat's 'Reduce Redundant, Outdated and Trivial Content' (http://www.tbs-sct.gc.ca/ws-nw/wu-fe/rot-rid/index-eng.asp) web usability plan. Limit to documents in English."

iii The text of the Access to Information Request was: "All documents related to the 'Reduce Redundant, Outdated, and Trivial Content' (http://www.tbs-sct.gc.ca/ws-nw/wu-fe/rot-rid/index-eng.asp) web usability plan including any reports, guidelines, practices, frameworks and/or rubrics provided to departments on how to assess and identify redundant, outdated and trivial content. Limit to documents in English."