

Tasbire Saiyera, Brenda Reyes Ayala, Qiufeng Du
University of Alberta, Edmonton, Alberta, Canada

Assessing the loss of Western Canadian digital heritage (Poster)

Abstract:

In 2003, UNESCO recognized the volatility of our heritage on the web. In response, many Canadian institutions rose to the challenge to preserve our digital heritage. This study examines web archives created by the University of Alberta Libraries relevant to Western Canadian heritage. We examine these collections in order to (1) assess their degree of link rot (which occurs when a website is no longer online) and (2) to determine how extensively these websites have been preserved.

1. Introduction

In 2003, the United Nations Educational, Scientific and Cultural Organization (UNESCO) established its Charter on the Preservation of the Digital Heritage, acknowledging the volatility of the world's digital heritage (UNESCO, 2003). The practice of capturing websites using web archives to preserve them as legal, historical, or informational records has been in effect since 2009 for the University of Alberta (University of Alberta, 2020). Using Archive-It (Archive-It, 2014), the university libraries classified websites as relevant to Western Canadian history if they resonated with the themes of their 19 collections (themes including western Canadian art, important events in Alberta history, and Canadian indigenous movements). This study addresses the following research questions:

1. Of the websites classified as being relevant to Western Canadian cultural heritage, how many of them have disappeared from the live web?
2. How extensively are these websites preserved/archived by institutions (including the University of Alberta Libraries)?

2. Literature Review

The problem link rot was first studied in depth by (Koehler, 2002), who monitored the status of a random set of URIs over four years, showing that approximately 67% of URIs became inaccessible after the four-year period. In their research, McNally, Wakaruk, and Davoodi (2015)

examined the extensive removal of Canadian government web content and its impact on researchers, concluding that the web archiving programs were crucial to maintaining their role as stewards of government information. For this reason, it is also worth examining how much of the University of Alberta’s Western Canadian digital history collections are currently preserved.

3. Methodology and Discussion

In order to assess how much of the University of Alberta’s archived Western Canadian historical websites were still on the live web, we created a Python program that checked to see if a specific URL is still available by checking its HTTP status code. We categorized "lost" websites as those that returned an HTTP status code other than 200. Table 1 lists some of the collections and their link rot percentages.

Collection Name	Age of collection in years	Description	No. of sites	% of collection unavailable
Idle No More	8	A Canadian political movement encompassing environmental concerns and the rights of indigenous communities	196	46.43 (91)
Western Canadian Arts	6	Born digital resources created by filmmakers in Western Canada	101	13.87 (14)
Fort McMurray Wildfire 2016	5	Websites related to the Fort McMurray Wildfire of 2016	52	25 (13)

Table 1: Web archive collections and their link rot percentages

Furthermore, we deployed MemGator to determine if copies of the websites were also present in web archives around the world. MemGator is able to search the web archives of many institutions and return a list of archived copies of a website, the time and date they were captured, and the institution that captured them (Alam & Nelson, 2016). Table 2 lists the number of captures held by institutions explored by MemGator.

Collection Name	No. of Archive-It captures (mostly UofA)	No. of Internet Archive captures	Total No. of captures
Idle No More	7005 (32.3%)	14049 (64.8%)	21664
Western Canadian Arts	289 (8.6%)	2930 (87.7%)	3342
Fort McMurray Wildfire 2016	4693 (34.3%)	8706 (63.7%)	13677

Table 2: Web archive collections and their preservation status

Although 13% - 46% of the websites from the Western Canadian heritage collections disappeared from the live web, it is promising to see the thousands of captures by the University of Alberta (using Archive-It) and the Internet Archive. We found that the Library and Archives of Canada did not preserve any of the websites from the three collections. Nevertheless, since Archive-It belongs to the Internet Archive, potential corruption of the Internet Archive’s servers pose the threat of Single Point of Failure (SPOF), where the failure of a single component leads directly to the failure of a preservation system (Rosenthal et. al, 2005). Our digital heritage should not be compromisable; further studies using collections from different institutions need to be conducted in order to assess the extensiveness of the loss of our digital Canadian heritage.

Reference List:

- Alam, S. & Nelson, M. (2016). MemGator - A portable concurrent memento aggregator: Cross-Platform CLI and server Binaries in go. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, JCDL 2016* (pp. 243-244). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2910896.2925452
- Archive-It. (2014). Archive-It: University of Alberta. Retrieved from <https://archive-it.org/organizations/401>
- Koehler, W. (2002). Web page change and persistence—a four-year longitudinal study. *Journal of the American Society for Information Science and Technology*, 53(2), 162-171. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.10018>
doi:10.1002/asi.10018
- McNally, M.B, Wakaruk, A., & Davoodi, D. (2015). Rotten by design: Shortened expiry dates for government of Canada web content. *Proceedings of the Annual Conference of CAIS, Canada*. doi: <https://doi.org/10.29173/cais909>
- Rosenthal, D.H, Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005). Requirements for digital preservation systems: A bottom-up approach, *D-Lib Magazine*, 11(11). Retrieved from: <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
- United Nations Educational Scientific and Cultural Organization. (2003). Charter on the preservation of the digital heritage. Retrieved from http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html
- University of Alberta. (2020). Digital Preservation Services. Retrieved from <https://www.library.ualberta.ca/digital-initiatives/preservation>