# Historians and Web Archives[1]

SUSANNE BELOVARI

RÉSUMÉ Depuis les années 1990, le web comme lieu où nous menons nos activités et générons nos documents de source primaire et secondaire a pris une importance croissante. De plus en plus, de tels documents existent seulement sur le web, sans qu'aucun autre document complémentaire ou supplémentaire ne soit disponible ailleurs. Alors que les archives du web ont commencé à conserver ce patrimoine en 1991, l'histoire du web comme discipline bien établie n'a pas encore vu le jour. Une explication possible pourrait être l'inquiétude qu'ont les historiens qu'ils/elles ne pourront pas reproduire leur méthode de recherche historique en se servant des archives du web, et ne pourront pas trouver les documents essentiels et qui font autorité. La première partie de l'article propose d'imaginer une historienne en 2050 qui veut chercher l'histoire du web en se servant des archives de l'Internet telles qu'elles existaient en 2015. Elle se sert de la méthode de recherche historique habituelle à partir de laquelle les historiens choisissent un sujet et effectuent leur recherche, furètent et contextualisent leurs sources de façon itérative et poussée. L'expérience connaît un échec quand l'historienne est incapable de repérer les dépôts d'archives appropriés et les documents qui font autorité sans avoir recours au web tel qu'il existait en 2015. La deuxième partie analyse vingt-et-une archives du web en 2015 et les enjeux qui pourraient avoir un impact sur la recherche historique. La plupart des archives sur le web sont, de toute apparence, semblables aux bibliothèques de ressources d'information. Les archivistes et les historiens, cependant, ont besoin de dépôts qui contiennent et qui rendent accessibles les documents essentiels du web qui ont une valeur durable culturelle, historique et de preuve. Cet article suggère que les historiens pourraient encore une fois être indispensables pour déterminer les questions archivistiques de base liées aux documents et archives du web, tout comme ils ont aidé à façonner les politiques archivistiques d'il y a quelques siècles.

ABSTRACT Since the 1990s, the Web has increasingly become the location where we carry out our activities and generate primary and secondary records. Increasingly,

1 I thank my peer reviewers and Mark Greene, Katharina Herring, Bill Maher, Nancy McGovern, David Prochaska, and Chris Prom for reading drafts and providing detailed and invaluable suggestions. I also thank Scott Schwartz for wholeheartedly supporting my research.

such records exist only on the Web, with no complementary or supplementary records available elsewhere. While web archives began to preserve this legacy in 1991, web history has not yet emerged as a fully developed field. One explanation may be historians' concerns that they will not be able to replicate their historical research process when using web archives, and may not find essential and authoritative records. The article's first section proposes a thought experiment in which a future historian in 2050 wants to research web history using web archives as they existed in 2015. She relies on the customary historical research process through which historians choose topics and search, browse, and contextualize sources in depth and iteratively. The experiment fails when our historian is unable to locate appropriate repositories and authoritative records without resorting to the live Web of 2015. The second section then analyzes 21 eminent web archives in 2015 and issues that may have an impact on historical research. Most web archives are apparently akin to libraries of information resources. Archivists and historians, however, need web repositories to contain and make accessible essential web records of enduring cultural, historical, and evidentiary value. The article suggests that historians may once again prove invaluable in figuring out basic archival issues related to web records and archives, just as they helped shape archival policies a couple of centuries ago.

## Introduction

Over the past 20 years, the Web has become the location where we carry out more and more of our activities and increasingly generate primary and secondary records. Linked, dynamic and ephemeral, privatized and commercialized, censored and ranked via secret search algorithms, the Web we now have is public, deep, and gated.[2] Government, corporate, institutional, group, and personal interactions, activities, transactions, and records are created and/or presented on the Web. Increasingly, these only exist on the Web, with no complementary or supplementary records available elsewhere. Being continuously updated, the Web has created a constant sense of now.

Web archives attempting to preserve (portions of) the public Web have existed since 1991. However, web history, focusing on the Web as a site of lived experiences, has not yet emerged as a fully developed field. Any search of publications, courses, or portrayals of the discipline of history will confirm this.[3] After all, it was only two years ago that Ian Milligan and Peter Webster

2 While the *Web* uses Hypertext Transfer Protocol (HTTP) to access information over the Internet, the *Internet* as a network of networks uses several protocols to transfer information, including SMTP and instant messaging. Yet the two terms are used interchangeably; see, for example, Mia Consalvo and Charles Ess, eds., *The Handbook of Internet Studies* (Oxford, UK: Blackwell Publishing, 2011).

3 This statement is based on an online search for web-history publications and courses taught at major history departments in the US, UK, and Germany. For contemporary portrayals of historical research, see, for example, the *Digital History at Harvard* program, which also considers the adoption of digital approaches to be late and piecemeal, Harvard University

created the Web Archives for Historians website[4] to encourage this new field of inquiry. The large amount of data amassed in web archives entices the few scholars working in this area to pursue grants and program support to explore big data methodological approaches and technological tools.[5] Building on earlier quantitative historical methodologies such as content analysis or cliometrics, these new approaches may offer insights into the almost instantaneous spread of news, for instance, or the prevalence of certain terms across the globe. But it remains unlikely that they will replace "the traditional study of individual records."[6] How else would historians answer questions of who, what, when, why, where, and for what reason and how else would they situate phenomena within relevant contexts and considerations of power?

Given the importance of the Web over the past few decades, it is quite striking that there are so few contemporary historians studying it and our information society.[7] Several possible explanations come to mind, among them the relatively small number of contemporary history programs, the lack of "organizational homes" and developed researcher ecosystems for web historians

History Department, "Digital History at Harvard," accessed 16 July 2016, http://history.fas.harvard.edu/digital_teaching_fellows_program. See also Jennifer Rutner and Roger C. Schonfeld, "Supporting the Changing Research Practices of Historians," Ithaka S+R (7 December 2012), accessed 3 February 2016, https://doi.org/10.18665/sr.22532; Jakub Beneš, Pavlina Bobič, Klaus Richter, Kathleen Smith, and Andrea Buchner, *Report on Archival Research Practices: Work Package 4* (Collaborative European Digital Archive Infrastructure/CENDARI, 25 June 2013), accessed 2 February 2016, http://www.cendari.eu/about-us/project-deliverables; Stefan Heidenreich, "Datendichte und digitale Geschichte: 8 Thesen," *Zeitgeschichte Online* (March 2010), accessed 4 February 2016, http://www.zeitgeschichte-online.de/kommentar/datendichte-und-digitale-geschichte-8-thesen. For critical web investigations outside of history, see *Christian Fuchs: Information – Society – Technology & Media* (blog), accessed 11 March 2017, http://fuchs.uti.at.

4 See Web Archives for Historians, accessed 11 March 2017, https://webarchivehistorians.org.

5 Ian Milligan, "SSHRC Proposal: An Infinite Archive? Developing HistoryCrawler to Explore the Internet Archive as a Historical Resource," *Ian Milligan: Digital History, Web Archives, and Contemporary History* (blog), accessed 11 March 2015, https://ianmilligan.ca/sshrc-proposal.

6 Archive-It tellingly calls them "resources." See "Archive-It Research Services," https://webarchive.jira.com/wiki/display/ARS; and "Born Digital Big Data and Methods for History and the Humanities," *Born Digital Big Data* (blog), https://borndigitaldata.blogs.sas.ac.uk; both accessed 8 February 2016.

7 Jürgen Danyel, "Zeitgeschichte der Informationsgesellschaft," *Zeitgeschichtliche Forschungen/Studies in Contemporary History* 9 (2012): 186–211. One of the exceptions here is the work by Niels Brügger; see, for example, "Digital History and a Register of Websites: An Old Practice with New Implications," in *The Long History of New Media: Technology, Historiography, and Contextualizing Newness, ed*. Dave Park, Nick Jankowski, and Steve Jones (New York: Peter Lang International Academic Publishers, 2011): 283–98.

and others,[8] and historians' age, lack of training, and lack of skill sets.[9] My own historical research and conversations with academic historians point to a further factor: their worry that they may not be able to replicate their historical research process when using web archives.[10] This unease relates to three preoccupations: Do web archives preserve essential and authoritative records that have proven to be key to historical inquiries into the analog world? Do they do so in a manner that allows historians to search, browse, and contextualize? And how will historians acquire additional information related to the new aspects of a deep, gated, and dynamic Web?

This article considers these questions from several perspectives. The first section proposes a thought experiment that stipulates certain conditions: in the experiment, a historian in 2050 wants to research web history and use web archives as they existed in 2015. To do so, she relies on the traditional historical research process through which historians choose topics and search, browse, and contextualize sources in depth and iteratively. The experiment ultimately fails. The future historian is unable to locate appropriate repositories and authoritative records without resorting to the live Web of 2015 for both. In a second section, I therefore turn to 21 well-known web archives in 2015 and list issues they currently raise. I then analyze how some of these may impact historians' ability to do research in web archives.

## The Thought Experiment: Looking Back from 2050

### *Explaining the thought experiment*

Tim Hitchcock describes a web archive "as an 'object of study' created in the service of an imaginary scholar." How then would an imaginary historian from 2050 use and evaluate today's web archives?[11] This thought experiment

8 Ian Milligan, "Web Archives and Born-Digital Sources Workshop: Challenges, Future Steps, and the Field," *Ian Milligan: Digital History, Web Archives, and Contemporary History* (blog), 10 June 2016, https://ianmilligan.ca/2016/06/10/web-archives-and-born-digital-sources-workshop-challenges-future-steps-and-the-field/#more-3278.

9 This is extrapolated from a 2010 American Historical Association survey of US historians that did not yet include web history as a new area of study. Two-thirds of historians surveyed were older than 45. Half of active new media users noted lack of time and frustrations with quickly outdated software and skills as factors in not learning new programs; see Robert B. Townsend, "How Is New Media Reshaping the Work of Historians," *Perspectives on History* 48, no. 8 (November 2010), https://www.historians.org/publications-and-directories/perspectives-on-history/november-2010/how-is-new-media-reshaping-the-work-of-historians.

10 For example, the author worked as a Holocaust restitution historian and archivist for the Holocaust Victims' Information and Support Center, Jewish Community of Vienna, and is currently revising an article about Viennese culinary history before 1938.

11 Tim Hitchcock is a professor of digital history and a co-director of the Sussex Humanities

permits us to consider how today's web archives might fare in the future when – under the conditions set by the experiment – there will be no contemporary (web) sources left to provide additional information.

At various times called imaginary history, future study experiments, or what-if situations, thought experiments have long been used in the sciences and humanities – from physics and economics to philosophy and history.[12] Our experiment sets the following conditions:

(1) It is the year 2050.
(2) We freeze web archives as of August 2015. They are thus the only repositories of the web world in 2050.
(3) The highly volatile Web, as we knew it, was superseded by something different in 2017.
(4) Our imaginary historian is the first to attempt a web-based history.
(5) The imaginary historian uses the customary historical research process developed during the "analog age."

The suppositions allow us to do several things. First, we can disregard any future, as of yet hypothetical, technological solutions to issues raised below. Second, we can imagine web archives suspended in isolation. We can evaluate whether they preserve authoritative and essential records because nothing else from the web world was preserved from that time period. Third, we imagine our historian as a non-native of the Web. Without access to any web experts, she does not know more about the Web and web research than those who lived in 2015. And fourth, we envision our historian at the point where she starts her research investigation, i.e., with her first set of research questions. This type of experiment is fundamentally different in nature than traditional user studies, which ask researchers to test predetermined features within a particular web archive.[13] Instead, it starts when the researcher is just

Lab at the University of Sussex, UK. See Tim Hitchcock, "The UK Web Archive, Born-Digital Sources, and Rethinking the Future of Research," *Web Archives for Historians* (blog), 5 June 2015, https://webarchivehistorians.org/2015/06.

12 For thought experiments in historical research, see Julian Reiss, "What If? Fictions, Models, Thought Experiments in History" (presentation slides), accessed 18 February 2016, http://www.jreiss.org/Presentations/Toronto_TEs.pdf; Julian Reiss, "Counterfactuals, Thought Experiments and Singular Causal Inference in History," *Philosophy of Science* 76, no. 5 (December 2009): 712–23; and James Robert Brown, "Thought Experiments since the Scientific Revolution," *International Studies in the Philosophy of Science* 1, no. 1 (September 1986): 1–15.

13 Traditional user-study methodologies have participants start within an experimental set-up (i.e., already within a web archives) in order to test a list of predetermined features considered significant by the testing unit. As of 2016, there were only a handful of user studies, none of which included research scenarios by historians. See Julien Masanès, ed., *Web Archiving* (Berlin: Springer, 2006), esp. chap. 6; IIPC Access Working

beginning to define research areas and questions and consider where she might find relevant sources, and before she has found appropriate web archives. Traditional user studies are generally somewhat abstract, isolating aspects of what is, after all, one interconnected endeavour on the parts of users. How people search, what they search for, and for what reasons they search are interconnected questions – and this brings us back to the historical research process.

### *Explaining the historical research process*

It is not typical anymore for archivists to be historians or trained extensively in historical methods, while historians usually do not explain their research process. A search of relevant literature and of history department websites reveals a shortage of publications and classes that explicitly teach the research process. This lack of resources implies that graduate students learning the trade can simply imbibe it by reading published research. Related websites mainly offer instructions on how to do research in general and how to write general research papers. The web resource "Learning to Do Historical Research: A Primer for Environmental Historians and Others" is the exception. William Cronon, its author and a preeminent historian, was troubled by the lack of research skills among his history students. To correct this shortcoming, Cronon and his students developed the above primer in 2008, and since at least 2013, it has been "the number one Google hit if one searches for the phrase 'learning to do historical research.'"[14]

In the primer, Cronon skillfully outlines historians' iterative (i.e., continual) approach to questions, records, and the research process itself. To be able to describe and explain phenomena, historians have to define and redefine research areas, interests, and questions, moving from broad aspects to specifics and back again. In these searches, they gradually zero in on a specific aspect or source of a historical phenomenon. They then move out to contexts and comparisons of that source or aspect. They zoom in again to a specific

Group, *Use Cases for Access to Internet Archive, version 1* (International Internet Preservation Consortium, May 2006), accessed 19 February 2016, http://www.netpreserve.org/sites/default/files/resources/UseCases.pdf; Jinfang Niu, "Functionalities of Web Archives," *D-Lib Magazine* 18, no. 3/4 (March/April 2012), doi:10.1045/march2012-niu2; and Miguel Ângelo Leal da Costa, "Information Search in Web Archives" (doctoral thesis, University of Lisbon, 2014), accessed 11 March 2017, http://sobre.arquivo.pt/sobre/publicacoes-1/Documentos-acerca-do-Arquivo.pt/information-search-in-web-archives.

14 Richard White, "William Cronon Biography," from the booklet of the 2013 American Historical Association (AHA) Annual Meeting and published on the AHA's website at AHA History & Archives, "Presidential Addresses: William Cronon Biography," accessed 4 February 2016, http://www.historians.org/about-aha-and-membership/aha-history-and-archives/presidential-addresses/william-cronon-biography.

detail and its sources and then repeat this conceptual process over and over again. For that, they need to (1) browse overall, (2) search for particulars, and (3) put all of their sources in conversation with each other, to (4) thereby evaluate primary and secondary sources for their authenticity, trustworthiness, and content. It is the responsibility of the historian to view "people and processes and problems [etc.] in perspective and proportion."[15] Tellingly, Cronon does not discuss web history in his primer. Conceivably, he and others either consider the research process to be format- or medium-neutral or they have not yet contemplated how web archives might introduce novel aspects or even reconfigure the process.

A last question should be considered before starting the thought experiment: why should web archivists be concerned about the experiences and opinions of historians? There are several reasons. Their breadth of research interests and depth of analysis over time, as well as their need for authoritative and trustworthy records, make historians one of the most challenging and wide-ranging user groups any archives can have.[16] Their stories are important. At their best, their record-based narratives will not just offer decontextualized images and anecdotes; they will also become instrumental in shaping knowledge, discussions, and actions within their communities. Finally, historians' research processes and the records they require compelled them once before, in the 18th and 19th centuries, to convince archives of the need for respect des fonds, provenance, and original order.[17] Perhaps the time has come again for history scholars to help develop archival policies and principles, this time for web archives.

15 David Cannadine, "Making History Now (An Inaugural Lecture)," *History in Focus* (Autumn 2001), accessed 5 February 2016, http://www.history.ac.uk/ihr/Focus/Whatishistory/cannadine.html.

16 See RESAW (A Research Infrastructure for the Study of Archived Web Materials), accessed 20 April 2016, http://resaw.eu. RESAW is a European project for creating an infrastructure driven by researcher needs. While the first conference on web archives as scholarly sources took place in June 2015 at Aarhus University in Denmark (Web Archives as Scholarly Sources: Issues, Practices, and Perspectives), there is still almost no in-depth historical research being done using web archives, as noted in the literature. Web archives managers informally recount how scholars leave their sites quickly for Google and other corporate search engines where, as of now, they can still find some older and cached websites, regardless of all the caveats connected to them (e.g. there is no way to verify whether, when and how websites have been altered over time).

17 Primary materials were thus preserved within original webs of contexts and meanings. These historians who often also worked in archives include, for instance, Philipp Ernst Spieß, Natalis de Wailly, Samuel Muller Fz., Leopold von Ranke, Robert Fruin, and Johan Adriaan Feith. See Michel Duchlein, "The History of European Archives and the Development of the Archival Profession in Europe," *American Archivist* 55 (Winter 1992): 14–25; and Kasper Risbjerg Eskildsen, "Leopold Ranke's Archival Turn: Location and Evidence in Modern Historiography," *Modern Intellectual History* 5, no. 3 (November 2008): 425–53.

### *Starting the thought experiment*

As explained, we start our thought experiment at the point where each historian begins: with a research interest and a search for source materials. Working in 2050, the historian will look for a global web repository that preserved the former global Web. It is no trivial matter to ask *how* future historians would locate such institutions, especially if you consider endemic link rot, reference and content rot, and the online-only access points to most web archives. Furthermore, obstacles encountered in locating global web archives mirror complications encountered in locating and accessing archived sites within these archives.

From the onset, the scholar is stumped. How will she find repositories in the absence of any contemporary portal that assembled them? An elderly scholar tells her about the one truly global web archive of 2015, the Wayback Machine. As a project of the Internet Archive, it was largely created through Alexa Internet–automated web crawls. The historian is delighted by its scope when she gets to the Wayback Machine. The Wayback Machine includes more than 475 billion documents, totalling over 10PB.[18] But having found it, she cannot proceed: it is only searchable by URLs. How would she know those uniform resource locators, which were somewhat akin to analog addresses in the past? Proprietary search engines such as Google, Bing, and Yahoo, which had helped past users locate sites through full-text and keyword searches, were not preserved.[19] Undaunted, she will follow the usual procedures to dodge difficulties in finding source materials.

In a first step, she will look for directories listing historical websites. They will list the URLs needed to search the Wayback Machine and may include descriptions of websites. Unfortunately, although throughout the centuries archives and libraries had collected or even created paper directories, inventories, almanacs, catalogues, indices, registries, rosters, gazetteers, *Who's Who*, bibliographies, and so on, which listed institutions, newspapers, elites, people, holdings, businesses, articles, and even books from a particular time period, our historian cannot locate any *similar* web directories or any that are in fact searchable. When she asks retired reference librarians about such

18 For the Wayback Machine's undisclosed crawling and appraisal algorithms and how they distort holdings and therefore scholarship in unknowable ways, see Kalev Leetaru, "How Much of the Internet Does the Wayback Machine Really Archive?" *Forbes* "Tech," 16 November 2015, https://www.forbes.com/sites/kalevleetaru/2015/11/16/how-much-of-the-internet-does-the-wayback-machine-really-archive/#2f4806969446; and "A Vision of the Role and Future of Web Archives: Conclusions and the Role of Archives," *The Signal* (Library of Congress blog), 24 May 2012, http://blogs.loc.gov/digitalpreservation/2012/05/a-vision-of-the-role-and-future-of-web-archives-conclusions-and-the-role-of-archives.

19 By 2015, numerous web-related services and products had already disappeared, among them the search engines Archie, Veronica/Jughead, and AltaVista.

discovery tools, they do not understand her inquiry.[20] They had neither created nor collected web directories, and neither had they documented web developments, including domain name changes.[21] The web age had been a time of living in the now. With websites either in flux or disappearing, historical considerations had apparently moved to the digital back burner.

Our historian then locates Jefferson Bailey, a retired director of the web archiving programs of Archive-It, a service of Internet Archive. He points her to two online web directories, VLIB (1991– ) and DMOZ (1999– ), as well as "vintage" printed Internet directories dating largely from the 1990s.[22] Of course, she does not have their URLs. If she did, she would discover their scope to be too limited for in-depth queries.[23] She asks whether the Wayback Machine could implement better search functionality (e.g., full-text search or an index to text, keywords, and links embedded in source code pages). Bailey responds that while this is theoretically possible, it is not operationally practical for their small non-profit to make specific parts full-text searchable. Costs for this "would be complicated to calculate" involving "engineering time, hardware/infrastructure, machine/processing time, and maintenance."[24]

20 This scenario is based on conversations the author has had with reference librarians currently working for a large US public library whose collections include said directories. Ironically, such directories tend to be shelved in on-site-only reference sections. Since the 2000s, numerous country code top-level domains, as well as .gov and .doc domains, have allowed commercial use of their domains. Since many national archives crawl those domain names in order to capture what are considered public and/or relevant records, the commercial use of those top-level domains should now effect crawling algorithms. Otherwise, the records captured may be of quite different provenance and outside the collection mandate of the archives.

21 The Wayback Machine does not publish its master inventory of captured domains (see Leetaru, "How Much of the Internet Does the Wayback Machine Really Archive?"), which could be browsed if it did. Web directories encompass search engines, white/yellow pages, mapping sites, library catalogues, Google Books, YouTube, encyclopedias, online newspaper archives, the Directory of Open Access Journals (DOAJ), social media, Ancestry.com, etc. They are constantly updated on the Web, and for proprietary and technical reasons, they cannot be archived beyond their home pages. Archiving analog directories, in contrast, was either supported or at least not prohibited by businesses, book publishers, newspapers, etc.

22 Jefferson Bailey, email correspondence with author, 25 July 2015. Peggy Garvin, *The United States Government Internet Directory* (Lanham, MD: Bernan Press, 2013) is one of the few such print directories still published today.

23 VLIB includes a very small number of topics/themes as well as sites. As of 2016, DMOZ was vastly better. In addition to being critiqued for their commercial interests, its US-centric volunteer editors do not consistently appraise important institutions and activities across the globe; see, for example, the incomplete list of the ministries of the government of France. As of March 14, 2017 dmoz.org will no longer be available.

24 Jefferson Bailey, email correspondence with author, 25 July 2015. In late 2015, the Internet Archive announced that it had received a $1.9 million grant from the Laura and John Arnold Foundation to optimize the scope and quality of captured web pages, to improve playback of media and the user interface, and to *make some keyword searches* possible. It remains to be seen if all these changes can indeed be implemented given limited funds, the above

For our historian, there is no longer any point in trying to use the Wayback Machine. After all, she lacks basic search tools and possibilities with which to scope out and identify topics, individuals, institutions, and sources. In a second step, she therefore limits her research to smaller, presumably more accessible repositories, a strategy not unlike the one employed by historians of the medieval world when selecting projects based on surviving sources. According to Bailey, the Internet Archive had also offered the subscriber-based Archive-It, a web archives of global reach, which could be found under "Subscription Service." Its self-curated collections had also included collections of websites, some of which were full-text searchable. Of course, having 326 subscribers, 2,661 public collections, and about 9 billion URLs, Archive-It could not develop an overarching collection strategy to adequately document the Web.[25] Browsing through the collections, the historian discovers that many have only preserved digitized records, such as historical climate data or city directories. Others contain archived websites that on their own raise appraisal questions (e.g., a law faculty blog, a search engine's home page). And as was the case for the Wayback Machine, flat crawls are frequent, and captured sites regularly do not load.

Forgoing her search for a global web archive, our historian then looks for national or regional repositories in Archive-It. It is logical for her to assume that such repositories would have preserved well-defined and representative slices of web life. She gets only one promising search result: the International Internet Preservation Consortium (IIPC). As a global coalition, IIPC had worked on "tools, standards and best practices of web archiving while promoting international collaboration and the broad access and use of web archives for research and cultural heritage."[26] IIPC partner sites, however, were not preserved in Archive-It. Having learned a few tricks by now, the historian looks in the Wayback Machine for IIPC. Its web development source code pages list the URLs for its partners. With those URLs, the historian then determines that many IIPC partners were not preserved within the Wayback

response by Bailey, and previous Internet Archive announcements in this regard.

25 Its partners include libraries, archives, and educational institutions (73%), non-governmental organizations (14%), national organizations (10%), and "other" (3%); see Jefferson Bailey and Herbert Van de Sompel, *Link Rot Symposium – Panel: Strategies I*, recorded October 2014, YouTube video published by Georgetown Law Library 27 March 2015, https://youtu.be/0piT4IKu9oY.

26 "About IIPC," accessed 3 March 2016, http://netpreserve.org/about-us. Of approximately 50 partners, half were European, a fifth were from the US, four from Asia, three from Canada, two from Oceania, one from Israel, and one from Egypt (August 2015), representing 46 regional or national libraries/archives, not-for-profit foundations and libraries, as well as two service providers.

Machine.[27] If they were preserved, they could either not be searched or the search kicked the historian into the live Web of 2015. As a result, this thought experiment is put on hold.[28]

## 2015: Web Archives on the Live Web

Our thought experiment from 2050 did not get us far. Our historian could not search, browse, and contextualize sources in global web archives in any meaningful, iterative, and in-depth manner (e.g., URL search, scope, and flat crawls). She could not consistently discover and access well-curated web archives in the Wayback Machine or on Archive-It. And she could not replicate search results. In this section, I therefore look at 21 of the IIPC partner archives she could not locate (see the appendix for a list) and I analyze their online presence on the Web in 2015. Table 1 lists issues found on those sites. I then analyze how some of these issues might affect the ability of historians to do research.

### *Analysis regarding content, scope, and appraisal*

Appraisal is central to history researchers and archivists. It is therefore surprising that many web archives do not explain their appraisal policies and that available documentation is inadequate and outdated. What repositories say they have is frequently not what scholars can actually find. Institutions may be unaware of how dynamic and automatic crawling algorithms constantly change what is being selected, i.e., their inherent appraisal criteria. Yet, as a result of poor documentation and these dynamic algorithms, scholars cannot analyze holdings as to breadth, depth, significance, and representativeness. They do not know what materials they may find; what motivated a selection (e.g., in the case of capturing an event); the trustworthiness of captures; when a web archives became inactive; or whether to travel to repositories with on-site only access to their holdings. It is important to remember in

27 This was the result of searching the Internet Archive in August 2015. In October, a search on the Wayback Machine found that it included the IIPC and its 47 members, with working links to archived sites. In a November 2015 search, the link to the IIPC did not work. Searching for IIPC in Archive-It in February 2016 produced only eight theme-related collections created by the IIPC, including the Olympic Games, the European refugee crisis, international co-operation organizations, and First World War commemorations.

28 This would be akin to an archives telling researchers to go to the office of the president as of 1877 to locate the rest of the analog correspondence the archives had not transferred. Link rot was also an issue on the IIPC's site in 2015. Web archives' URLs often linked to their national library site within which the web archives could not be located.

**Table 1:** Issues found on the sites of 21 IIPC partner web archives in 2015

***Discovery***
- there is no truly functioning "web archives portal" pooling web archives
- links to their own web archives sites not provided/not working (e.g., within IIPC institutions)

***Content, Scope, and Appraisal***
- fragmented coverage of web life
  - limited scope
  - no/insufficient information regarding appraisal, scope, and trustworthiness of captures
  - outdated or unrealistic scope/content descriptions
- no online information once web archives goes inactive

***Searching, Browsing, Contextualizing***
- limited search options (e.g., URL)
- no extensive (or findable) directories of websites
- unreliability of complex search options and website filtering
- search results cannot be replicated
- dead links
- lists of presumably captured sites are out-of-sync with actual captures
- metadata inconsistencies/errors (e.g., dates, authority control, descriptions, etc.)

***Assistance***
- national web archives require on-site access but offer no (functioning) catalogue or directory for researchers to know whether to travel to the repository
- no assistance in navigating discovery limitations
- no information about how to evaluate historical websites
- no information about FOIA-type rights to government and other websites
- no information regarding historical web developments

***Provenance and Original Order***
- redirecting to pages on the live Web or to pages of different historical dates
- lack of clear visual or textual alerts regarding redirects
- dead links, flat crawls (regardless of significance of site), and no URLs for pages not crawled

***Web Native Features***
- no information about native features (essential to document the history of the Web)
- no information about operational issues (e.g., unreadable attached medial files or virus alerts)

*Note: In order to be included in table 1, each issue had to occur twice, at a minimum. For example, at least four national web archives as well as the Internet Archive (Wayback Machine and many Archive-It subscribers) offer only a URL search option.*[29]

29 Another repository does not state what search options are available to on-site researchers, while several more offer searches by URLs and subjects.

this context that in contrast to the analog era, there is less and less available "outside" of the Web with which to compare holdings and substitute source materials.

Without available appraisal documentation, a collection of websites may raise serious questions as to what web archives are/were doing when, for example, commercial advertisements are preserved but not the local government pages; when club sites are included but none of the national or regional newspapers; or when significant newspaper, organizational, or government sites were crawled only once, years ago, and possibly only the site's first page. All of this may be more evocative of old-fashioned scrapbooking than archiving. My findings here are corroborated in part by Leetaru, who evaluated the Wayback Machine holdings by using large-scale archival mining techniques.[30] As it turned out, the Wayback Machine's automatic crawling and re-crawling did not necessarily capture representative or significant sections of the Web, about whose algorithms and profound changes we are further left in the dark. Yet what Leetaru suggests as alternative algorithmic appraisal criteria – random selection and prioritizing crawls by rate of change or popularity of a site[31] – misses much of what historians and archivists identified as significant appraisal factors in the preceding 150 years.[32] Perhaps Leetaru's criteria reflect his background as co-founder and manager of the global news monitoring project Global Database of Events, Language, and Tone. A political or social historian would probably not suggest change or popularity as valid crawling criteria for documenting stable institutions or minority issues, for instance.

As an alternative perspective, consider historian Niels Brügger's take on web appraisal. Studying the history of the Internet and media theory, he describes the issue as archiving "what is useful for the scholar." This is a rather idiosyncratic way of stating what is at stake here: preserving both the evidence of web history and an understanding of the scope, limitations, and biases of this evidence. Of course, individual scholars tend to privilege source materials that are central to their particular subject expertise; this being the case, we need experienced researchers from a variety of fields collaborating

30 Leetaru, "How Much of the Internet Does the Wayback Machine Really Archive?"

31 See, for example, Silvia Shenkolewski-Kroll and Assaf Tractinsky, *Using Web Analytics in Appraisal of Records on the Foreign Ministry of Israel Website* (EU25), 2015–16, and *Research of Retention and Disposition Processes in an Internet Website of the Government of Israel: The Ministry of Foreign Affairs as a Case Study* (EU01), InterPARES Research Studies, both accessed 18 February 2016, https://interparestrust.org/trust/about_research/studies. The authors propose the use of web analytics for the appraisal of websites.

32 See note 17 for literature on historians in this regard. Among archivists, appraisal factors and approaches may focus on value, significance, functional, content, or object analysis, the broader documentation strategy, macroappraisal, and reductive models as well as cost-benefit analysis, for instance.

with archivists to ensure adequate appraisal of the Web. They may help us solve some of the current issues regarding content and scope as well as help analyze the Web's native features and how these may relate to archival principles for web records, as discussed further below.[33]

### *Analysis regarding searching, browsing, contextualizing*

Several repositories are searchable only by URL. This presupposes two untenable assumptions: (1) access to a live Web that provides the URLs in order to retrieve archived websites; and (2) directories for sites that no longer exist or whose URL has changed.

Paraphrasing Zittrain, Albert, and Lessig, the mismatch between transient online and permanently archived web materials creates peculiar discovery and access issues.[34] To gather web directories in an online archive may be a solution to the lack of access to a type of public Google Index Database archives, to the lack of guarantees that representative sections of the web were archived, or to issues of link and content rot as well as inadequate discovery tools.[35] However, such *online* directories indexing people, institutions, locations, units, as well as primary and secondary sources, are constantly updated on the Web. They are not preserved as time-stamped records. Robots.txt keep web archives from capturing these proprietary directories; search options are blocked on their archived home pages.[36] Two examples may help illustrate what this means for historians. In contrast to print directories, historians cannot search Whitepages.com – on the live Web or in web archives – to verify whether Jane Doe lived at a certain address in 2005, 2009, or 2012 for instance. Nor

33 We can assume that national web archives appraise thoughtfully, but appraisal documentation is often not available online.

34 See Jonathan Zittrain, Kendra Albert, Lawrence Lessig, "Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations," *Legal Information Management* 14, no. 2 (June 2014): 88–99, http://dx.doi.org/10.1017/S1472669614000255. The authors discuss link and reference rot, noting that 70% of the URLs in three leading Harvard legal journals (1996–2012) and 50% of URLs within United States Supreme Court opinions (1996–2013) no longer work.

35 The Google Index Database crawls about 16% of the Web (about 100 million gigabytes). It is unlikely to ever be available for public archiving. What we need is a publicly funded web, successful public browsers, search engines, social media outlets, and public entertainment platforms. If considering public funding for a full-text searchable Wayback Machine, one would need to address automated, limited, and flat crawls and automated appraisal decisions.

36 For an interesting twist, consider Google Books. A search inside Archive-It brings up only the "Google Books Settlement Web Archive" (as of January 2016). Nothing indicates that Google Books may in fact be archived in the Internet Archive. And yet a search in the live Web reveals that individual users have uploaded over a million Google books to the Internet Archive; see https://archive.org/details/googlebooks.

can they get a historical overview of types and numbers of trade schools or newspapers in a particular city when searching in Whitepages.com or any of its equivalents internationally (for example, www.herold.at for Austria) for these years. Without this information, how will historians make an informed choice regarding which five relevant and significant schools and journals to study for any particular year?[37] When negotiating with parent companies of web directories, repositories should try to persuade these companies to archive their directories at thoughtfully selected time intervals. If need be, temporary access restrictions to archived instances might need to be accepted.

For title and keyword search functionality, good metadata and authority control are critical in order to find particular entities. An example: researchers will reasonably assume that national repositories preserve the websites of the executive, judiciary, and legislative government offices and that they will offer good authority control to find them.[38] However, a comparison of national government offices (listed on today's Web) with holdings in two national web archives over the preceding 10 years reveals that lack of consistent authority control is a real problem. In the absence of such control, captured government sites are not discoverable through keyword and title searches. Moreover, some national web archives crawled government sites only once or did not crawl them at all. This is probably because many were being archived in the Wayback Machine. But these national web archives do not inform researchers about this fact, and without having the requisite URLs, how can scholars locate government sites in the Wayback Machine? If they are in the Wayback Machine, will its generally flat crawls be able to preserve the essential, authoritative records beyond the top layers of a site?

If there is a lack of good metadata, of good search functionality, and of local resource guides; if rampant link rot and content drift continue; and if researchers cannot get access to time-stamped and relevant online directories, where will they find something as basic as a government's structure and entities from a certain year? How will they drill down to an office that will prove germane to their research project? How will they assess accuracy, historical relevance, and bias of the office and its website? Moreover, there also appears to be a qualitatively new aspect to web repositories. Barring huge disasters,

37 The Library of Congress states that "there are no current city directories for many major U.S. cities – among them Washington, D.C., New York City, Chicago, and Los Angeles. Therefore, some editions … may be many years out-of-date"; see Library of Congress, Local History & Genealogy Reference Services, "Telephone and City Directories in the Library of Congress: Current Directories," accessed 22 February 2016, https://www.loc.gov/rr/genealogy/bib_guid/telephon.html#usci.

38 Almost all national web archives are organized by national libraries. Do libraries' web archiving policies differ from those of archives given that archivists have appraisal experience and follow archival principles? Or are policies largely shaped by IT staff?

historians could expect an analog archival collection or item to be available in a physical archives on any given day. Scholars searching web archives, conversely, may get search results that cannot be replicated weeks or even hours later.[39] It is an instability in web archives that curiously mirrors the instability of the Web, albeit for different reasons.

***Analysis regarding assistance***

Due to copyright and privacy legislation, many national web archives provide only on-site access to their holdings. As of now, researchers will have to accept this, even though a major mandate of these archives is to preserve sites and records of public institutions published on an interconnected and instantly available Web. If such restrictions exist, how will historians decide whether to travel long distances to conduct research when repositories usually do not offer (functioning) online catalogues of their web records?

More broadly, there is little assistance for scholars trying to navigate discovery limitations. For example, given the realities of URL-only search and flat crawls, researchers will face dead links to potentially essential sites (URLs are usually not provided). Why not point out that the source codes of recently captured pages (e.g., path in Mozilla: tools-web developer option-source code/page source) include the URLs of pages not crawled as well as keywords, descriptions, titles, and tags? Most web repositories also do not provide resources about (1) how to evaluate historical websites, (2) how to assess the authority and validity of web content, or (3) how historical web development will affect authority, provenance, and trustworthiness of sites (e.g., domain naming conventions).

***Analysis regarding provenance and original order, and Web native features***

Researchers clicking on an archived URL are often redirected to the *live* website without being told (they have to notice the changed URL) or they receive a redirect alert that is all but invisible. Even more problematic, clicking links of a sub-page of a certain date may redirect a researcher to a sub-page from a completely different date. In both cases, the rendered information may pre- or postdate the original search date by months or even years. However, there is never any documentation indicating that these temporal drifts reflect specific crawling criteria (e.g., "significant" site changes). In my mind, it remains unclear whether temporal drifts can ever be a valid archival undertaking. If redirects are here to stay, however, temporal drifts of sub-pages and linking to the live Web should be barred. Furthermore, based on a proposal by

39 I learned to take screen shots of my search results.

Miguel Ângelo Leal da Costa, archives need to develop better user interfaces to highlight the temporal drift of search results.[40]

Redirects, flat crawls (regardless of significance of site), and dead ends destroy original order and provenance, which have been at the very foundation of valid historical research and archival policies for centuries.[41] Not knowing yet what these concepts may mean in a web environment, let me take a first stab by modifying Niu's definition of original order as the external structure of an archived web object.[42] For my purposes, we first have the internal context or hierarchical structure of a website. I would argue that this is akin to how a record creator's papers are arranged in folders and represents the original order. Second, we have its linked external contexts, i.e., its explicitly significant link relationships. This may be akin to references in analog texts that list significant external contexts (including other sites of the same record creator), verification, or corroboration. Original order, contexts, and provenance cannot necessarily be maintained when sites are not crawled in their entirety (relevant content), subsections are crawled at different times (relevant time frame), and links to outside pages dead-end and lack link URLs.[43] For example, consider how crawling only the home page of a complex site will make it nearly impossible for historians to infer actual provenance and significance, let alone content. The analogy here is to a scholar who can locate only the first page or fragment of a physical letter, newspaper, or extensive business or government records in an analog archives. Ultimately, these scholars can only analyze such materials outside their actual historical contexts.[44]

40 There are, of course, other challenges, such as *temporal inconsistency*, i.e., the fact that sites change while being archived; see Niels Brügger, "Web History and the Web as a Historical Source," *Zeithistoriche Forschungen/Studies in Contemporary History* 9, no. 2 (2012), http://www.zeithistorische-forschungen.de/2-2012/id=4426. See also Costa, "Information Search in Web Archives."

41 While respect des fonds is different from provenance (the latter referring to the record creator as well as the original ordering principle), I here consider provenance (record creator or origin) and original order to be distinct from each other. This reflects, for example, the Society of American Archivists' *Glossary of Archival and Records Terminology*, certain metadata schemas (e.g., PROV family of specification by W3C), and popular understanding of the terms.

42 Jinfang Niu, "An Overview of Web Archiving," *D-Lib Magazine* 18, no. 3/4 (March/April 2012), http://www.dlib.org/dlib/march12/niu/03niu1.html. This publication is also available at http://scholarcommons.usf.edu/si_facpub/308.

43 Imagine using a volume containing a print newspaper from May 1809 in which page 4 dates from 6 April 1754 and nothing or only a tiny header (i.e., URL) informs readers about the inserted page. Imagine studying the historical development of the State Department in 1877 by looking through its paper records and suddenly encountering records dating from 1932.

44 Although Niels Brügger and Peter Webster have begun to discuss websites as new types of primary sources, we need a broader discussion within the field. For example, historians' experience in analyzing undated manuscript drafts and various daily newspaper editions may point to ways of analyzing ephemeral websites as "digital drafts." See Brügger, "Web History and the Web as a Historical Source"; and Peter Webster, "Web Archives: A New

Web archives usually do not provide any documentation about changes in domain names, web developments (which will impinge on provenance evaluations), or the trustworthiness and authenticity of web captures and chain of custody.[45] Maintaining administrative and secondary documentation is not a new archival task in itself. Analog archives have long collected pertinent materials and pointed users to sources elsewhere. Today, though, the need for this is more urgent, given a global, interconnected, and currently inherently ephemeral network where less and less exists beyond its reach. Documentation may also include such issues as functionality (browsers and algorithms); changing types of prevalent sources materials (e.g., the proliferation of personal narratives and popular entertainment); historical web developments (in the gated and deep Web, cloud services, and social media, for instance); statistics regarding link rot, content drift, and users; micro-targeting and micro-delivery by browsers, commercial sites, and social media outlets (e.g., how to replicate searches and what contemporaries saw, how to document the reach of news or advertising outlets); among many others.[46]

## Final Observations

Within a few decades the web has become a constant presence in our lives. Dominated as it is by corporate interests, it is part of information(al) or digital capitalism. In this context, running not-for-profit web archives has become a daunting undertaking. Resources for archiving are minuscule when facing a

Class of Primary Source for Historians?" (presentation slides), accessed 8 February 2016, http://www.slideshare.net/historyspot/aadda-dhist-11-jun-2013. See also James Baker, "A Page, but Not as We Know It," *Analytical Access to the Domain Dark Archive* (blog), 13 June 2013, http://domaindarkarchive.blogspot.com/2013/06/a-page-but-not-as-we-know-it_13.html.

45 Current developments include the W3C provenance (PROV) family of specification (not yet widely adopted) and InterPARES Trust and its focus on forensics, authenticity, and trustworthiness. See, for example, Luciana Duranti and Corinne Rogers, "Memory Forensics: Integrating Digital Forensics with Archival Science for Trusting Records and Data," *eForensics Magazine* 2, no. 15 (10 November 2013): 96–111; Olaf Hartig, "Provenance Information in the Web of Data," in *Proceedings of the WWW2009 Workshop on Linked Data on the Web, Madrid, Spain, 20 April 2009*, ed. Christian Bizer, Tom Heath, Tim Berners-Lee, and Kingsley Idehen (CEUR Workshop Proceedings 538), accessed 11 March 2017, http://ceur-ws.org/Vol-538/ldow2009_paper18.pdf.

46 For many, the unrest in Ferguson, MO, in August 2014 originally had not "happened" because Facebook algorithms apparently filtered out relevant newsfeeds. How will Facebook's project of directly hosting news content affect the production of news now that it will be shaped by Facebook's corporate interests, algorithmic filters, and micro-targeting? How does the Web change our lives when the lines traditionally separating consumers, workers, and employers are disappearing or becoming invisible? For example, online users not only produce massive data (about users and their contacts), which are sold to the highest bidder, but also create online content and represent a captive advertising audience.

global, ephemeral, distributed, and increasingly proprietary and gated Web. Partial and technology-driven solutions may thus appear to be the only feasible ones for documenting those portions of the public Web that are most relevant to the mandate of each repository.[47] Figuring these out can be so overwhelming that basic questions are pushed to the back burner; they can be solved later, presumably.

Focusing on how historians research and what records they need has allowed us to underscore some of the basic issues at stake here. The interests of historians and archives overlap as they relate to appraisal, scope, discovery, access, and provenance. Historians conduct wide-ranging contextual inquiries for historical, representative, authoritative, and accurate records; archivists, in accordance with their professional principles and legal mandates, identify and preserve representative, authoritative, authentic, and accurate records of enduring value and make those accessible. This is a very different kind of research than the exact search for one URL, one photograph, or one news story. This is a very different kind of archiving than the creation of a Wunderkammer or its everyday counterpart, a scrapbook; of library subject collections; personal photo albums; or Facebook entries. Today, most web archives are more akin to contemporary libraries of web information resources, with users moving back and forth between repositories and the live Web to cobble together needed information, records, and sources.[48] Historians and archivists, however, want or need repositories to be historical archives that contain and make accessible essential web records of enduring cultural, historical, and evidentiary value for all of us in the future.

Of course, one can argue that both web archives and historians will change in the coming years. This is likely true, but ultimately unknowable. Understanding some of the issues facing web archives today may, however, be a first step in addressing them. Just as historians helped shape archival policies a couple of centuries ago, they might once again prove invaluable, this time in helping figure out basic issues regarding the Web and web archiving.

47 It is therefore paramount to have a portal that allows users to search across repositories. The Time Travel Service (http://timetravel.mementoweb.org) relies primarily on Wayback Machine holdings and is still plagued by major user functionality issues.

48 The Internet Archive describes itself consistently as an "Internet library;" see "About the Internet Archive," accessed 20 April 2016, http://archive.org/about. Link and content rot make the contemporary Web an unreliable source for web history; see Alex Wellerstein, "The Year of the Disappearing Websites," *Restricted Data: The Nuclear Secrecy Blog*, 27 December 2013, http://blog.nuclearsecrecy.com/2013/12/27/year-disappearing-websites. For estimates about website duration, see Francine Barone, David Zeitlyn, and Viktor Mayer-Schönberger, "Learning from Failure: The Case of the Disappearing Web Site," *First Monday* 20, no. 5 (4 May 2015), http://dx.doi.org/10.5210/fm.v20i5.5852.

***Susanne Belovari*** *is the archivist for faculty papers and an assistant professor at the University Archives, University of Illinois, Urbana-Champaign. Her previous positions include working as Holocaust restitution historian and archivist for the Jewish Community of Vienna, Austria, rebuilding its archives, which had been closed by the National Socialists in 1938. Her recent research and publications have concentrated on international archival topics and history, practical approaches to digital issues, as well as culinary history, Holocaust restitution, and the history of anthropology. In her current historical research, she is investigating how migrant and refugee objects and practices help rediscover culinary and other histories "back home." Other recent research has focused on developing simple and quick approaches to appraising and processing digital records and collections.*

**Appendix**

The 21 IIPC Partner Web archives analyzed in 2015 were Webarchiv Österreich (Austrian National Library), Patrimoni Digital de Catalunya (Digital Heritage of Catalonia, Spain), Archivo de la Web Chilena (Biblioteca Nacional de Chile/National Library of Chile), Archiv ceskeho webu (National Library of the Czech Republic), Netarkivet.dk (Danish Royal Library et al.), Bibliothèque nationale de France (National Library of France), Deutsche Nationalbibliothek (German National Library), Landsbókasafn Íslands – Háskólabókasafn (National and University Library of Iceland), Israeli Internet Archive, Webarchief van Nederland (Web Archive of the Netherlands), Nasjonalbiblioteket (National Library of Norway), Arquivo.pt (Portuguese Web Archive at FCCN-FCT), Bibliothèque et archives nationales du Québec (BaNQ) Web Archiving (Quebec), Archivo de la Web Española (Spanish Web Archive), Kulturarw3 (National Library of Sweden), Webarchiv Schweiz (Web Archive Switzerland), British Library Web Archiving (UK), Harvard's Web Archiving Collection (Harvard University, US), Library of Congress Web Archives (US), Stanford Web Archive Portal (Stanford University, US), and Internet Archive.

P.O. Box 2596, Station D, Ottawa, ON K1P 5W6 phone 613 234-6977 fax 613 234-8500
email aca@archivists.ca www.archivists.ca

Association of
Canadian Archivists
Association canadienne
des archivistes

# Archives, Disrupted.

**June 8 - 10, 2017**
**Ottawa, Ontario, Canada**

The Association of Canadian Archivists invites you to the 42nd Annual Conference "Archives, disrupted.", June 8-10, 2017 in Ottawa, Ontario. Just a few short weeks before the 150th anniversary of Canadian Confederation, archivists from around the world will explore how the archival endeavor can emerge strengthened and changed from the effects of disruptive forces, and define a new path that is relevant and valuable to its practitioners, institutions, users and stakeholders. The Association extends a warm welcome to all archivists, from home and away.

Conference sessions and activities are based in and around the Marriott hotel in the heart of downtown Ottawa, a short walking distance from the Parliamentary precinct, national museums, and Library and Archives Canada's downtown facility. Join us as we celebrate the 150th anniversary of Confederation right in the heart of it all! Highlights include:

- Experiencing Ottawa's red-hot food and drink scene at a great selection of restaurants and pubs, including some of the city's best microbreweries
- Behind-the-scenes guided tours of some of the capital's world-class archives and cultural organizations, including Library and Archives Canada's Gatineau Preservation Centre and the Library of Parliament
- Walking tours highlighting the diverse history of our capital city
- Many of our national museums, including the National Gallery of Canada, the Canadian War Museum, and the Canadian Museum of History
- The beautiful natural setting of Ottawa-Gatineau, including the UNESCO world heritage site, the Rideau Canal, and beautiful Gatineau Park

Ottawa extends a warm welcome to all archivists, national and international. Comments from previous international delegates indicate the experience is second-to-none:

*"It was a privilege and pleasure to be able to attend ACA 2014 and I enjoyed it immensely. It was also highly thought-provoking and has given me some great ideas for framing my research. I really hope to be able to attend future conferences and follow up on the interesting projects in play."*

*"I enjoyed the conference very much. I've always found ACA conferences so much more stimulating than professional conferences in other countries!"*

Plan to extend your stay in Canada's capital! Visit Ottawa Tourism or Tourism Ontario to see the many things there are to do.

Conference information and registration is available at the ACA's 2017 Conference website, http://archivists.ca/content/annual-conference.