WebArchiving@UNT

Web Archiving Bibliography 2013

Prepared By

Brenda Reyes Ayala
Brenda.Reyes@unt.edu

June 28, 2013

UNT Libraries
THE POWER OF IDEAS STARTS HERE

WebArchiving@UNT

Document Information	
Title	Web Archiving Bibliography
Author(s)	Reyes Ayala, Brenda
Original Creation Date	2013-06-28
Version	1.0
Date of Current Version	2013-06-28
Revised By	Reyes Ayala, Brenda
Description	The following is a bibliography of the field of web archiving. It includes a preface as well as list of bibliographical resources.
Rights Information	Copyright 2012 Web Archiving Project
Licensing Information	This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/>. This license allows you to copy, distribute, transmit, or adapt the content of this work to derivative work. The license requires that you credit the original creators of the work (attribution). This work may not be used for commercial purposes. _____ This document derives some of its content from the following works licensed under Creative Commons. "Example Work Title," Copyright 2010 by John Doe, used under a Creative Commons Attribution license: http://creativecommons.org/licenses/by-nc/3.0/.

Revision History	
Previous Version	none
Changes	

Table of Contents

Preface

Towards a Definition of Web Archiving

Any attempt at a bibliography is by necessity an attempt at naming, defining, and categorizing. As web archiving is still a nascent field, clear definitions and concepts are sometimes difficult to find. For example, the term "web archiving" is often used interchangeably with "web preservation" in the literature, without any clarification or distinction between the two. I argue that web preservation is a subset of digital preservation, that is, the process of maintaining Internet resources in a condition suitable for use. Web archiving is the action of storing Internet resources to preserve them as a historical, informational, legal, or evidential record. A web archive is a system which contains such records.

Web archiving entails several technical processes and usually occurs in the following order:

1. Selection: During this phase, web archivists select the Internet resources they are most interested in preserving.
2. Acquisition/Capture/Harvest: A piece of software known as a "crawler" visits every resource to be captured, makes a copy of it, and stores it in a database.
3. Access: The institution provides access to the captured content.

A Brief History of Web Archiving

Web archiving as a field is only a few years old and can be said to have begun in 1996. In that year, Brewster Kahle founded the Internet Archive with the mission of creating a universally accessible digital library. As its website states, the Internet Archive works "to prevent the Internet – a new medium with major historical significance – and other 'born-digital' materials from disappearing into the past" (Internet Archive, n.d.). Due to its size and scope, the Internet Archive has remained the largest and most ambitious web archiving project to date.

Also in 1996, the National Library of Australia inaugurated the first-ever web archiving program by a national library, an effort to capture portions of Australia's national domain (.au) as well as other Internet resources deemed significant to Australia's history and culture. In 2000, the Library of Congress began its Minerva Project (now the Library of Congress Web Archives), a web archiving effort aimed at creating topical collections of materials relevant to American history and culture. Many national libraries soon followed suit, for example the British Library, which began its UK Web Archive in 2004, and the National Library of France, which began to archive the French national domain in 2006. Over the years, many universities have also begun web archiving projects that supplement their existing physical and digital collections.

In 2003, the International Internet Preservation Consortium (IIPC) was founded with the mission of "improving the 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" (International Internet Preservation Consortium, n.d.). It has remained the only organization addressing the needs of the web archiving field.

Literature About Web Archiving

Substantive documentation about the field of web archiving did not emerge until the turn of the century. The field is still dominated by gray literature, that is, relevant materials that lie outside of the standard scholarly publication sphere of peer-reviewed journal articles and conference papers. Examples of this gray literature include technical reports, workshop presentations, committee working papers, and technical documentation. Scholarly articles about web archiving are still a relatively small percentage of the whole.

In this bibliography, I have chosen to include grey literature, but I have also included, whenever possible, scholarly, peer-reviewed works. I have excluded websites, and blogs, due to the constantly-changing nature of these sources.

How this Bibliography is Organized

The articles under "Introduction to Web Archiving" serve as an overview of the field for those who are unfamiliar with it, or seek a more historical perspective of the field's development.

Institutional web archiving is tailored to address the needs of a particular institution and takes place in the context of large national or academic libraries. It often involves archiving material that is of scholarly, cultural, historical, or political importance. The corresponding section recounts the numerous efforts undertaken by libraries and other institutions in order to preserve our digital heritage.

Unlike institutional web archiving, *personal web archiving* is tailored to the needs of a specific person and involves archiving very diverse material that is of interest to an individual. The section on "Personal Web Archiving" highlights a field that, though small, has been the subject of some very high-quality academic research.

"Challenges in Web Archiving" contains articles that describe the challenges currently facing the field, including issues of data quality and the persistence of Internet resources.

Web archiving practices are not universal, but are often deeply rooted in the laws and regulations of a country. Some countries have established legal frameworks in order to harvest their national domains, while others have not. "Legal Issues in Web Archiving" includes articles that discuss the legal and ethical issues that come with the practice of web archiving.

"Practices, Standards, and Tools" covers the many processes, approaches, and software tools that have been developed by different institutions to address their web archiving needs.

The section on "Web Archiving and Digital Libraries" features articles that address the issue of incorporating web archiving into the current practices of digital libraries.

"Quality in Web Archives" covers the process of ensuring that archived material is of sufficient quality. It highlights the work that various institutions have done to create processes and standards for their materials.

The section on "Research on Web Archives" describes research that has been conducted utilizing the many types of information found in web archives.

Bibliography

Introduction to Web Archiving

Ball, A. (2010). *Web archiving* (DCC State of the Art Report). Retrieved from Digital Curation Centre website: http://www.dcc.ac.uk/sites/default/files/documents/reports/sarwa-v1.1.pdf

Brown, A. (2006). *Archiving websites : a practical guide for information management professionals*. London, United Kingdom: Facet Publishing.

Brügger, N. (2005). *Archiving websites: general considerations and strategies.* Retrieved from Centre for Internet Research website: http://cfi.au.dk/fileadmin/www.cfi.au.dk/publikationer/archiving_underside/archiving.pdf

Brügger, N. (2011). Web Archiving - Between Past, Present, and Future. In M. Consalvo & C. Ess (Eds.), *The handbook of internet studies* (pp. 24–42). Malden, MA: Wiley-Blackwell.

Brügger, N. (Ed.). (2010). *Web history*. New York: Peter Lang.

Day, M. (2003a). *Collecting and preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust* (p. 91). Bath, United Kingdom: Joint Information Systems Committee.

Day, M. (2003b). Preserving the fabric of our lives: A survey of web preservation initiatives. In T.Koch & I.T. Solvberg (Eds.), *Lecture Notes in Computer Science: Vol 2769. Research and advanced technology for digital Libraries* (pp. 461–472). doi: 10.1007/978-3-540-45175-4_42

Day, M., & Pennock, M. (2006, September 28). *Archiving the Web*. Workshop presented at the Driving the Long-Term Preservation of Electronic Records, London.

Glenn, V. D. (2007). Preserving government and political information: The Web-at-Risk project. *First Monday*, *12*(7). doi:10.5210/fm.v12i7.1917

Gomes, D., Miranda, J., & Costa, M. (2011). A survey on web archiving initiatives. *Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries: Research and Advanced technology for Digital Libraries*, 408–420. Berlin, Germany: Springer-Verlag.

Hockx-Yu, H. (2011). The past issue of the web. *Proceedings of the ACM WebSci'11*. Koblenz, Germany: ACM.

International Internet Preservation Consortium Access Tools Working Group. (2006). *Use cases for access to internet archives*. Retrieved from International Internet Preservation Consortium website: http://www.netpreserve.org/resources/use-cases-access-internet-archives

Joint Information Systems Committee. (2008). *PoWR: The preservation of web resources handbook*. Retrieved from JISC PoWR website: http://jiscpowr.jiscinvolve.org/wp/guide/

Masanès, J. (2006). *Web archiving*. Berlin, Germany: Springer.

Niu, J. (2012). An overview of web archiving. *D-Lib Magazine Magazine*, *18*(3/4). doi:10.1045/march2012-niu1

Pennock, M. (2013). *Web-Archiving* (DPC Technology Watch Report). Retrieved from Digital Preservation Coalition website: http://www.dpconline.org/component/docman/doc_download/865-dpctw13-01pdf

Rauber, A., & Hunter, J. (2007). Introduction to the Special Issue on Web Archiving. *New Review of Hypermedia and Multimedia*, *13*(1), 3–5. doi:10.1080/13614560701520052

Toyoda, M., & Kitsuregawa, M. (2012). The history of web archiving. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1441 –1443. doi:10.1109/JPROC.2012.2189920

WebArchivists. (n.d.). The Web-Archives Timeline, by WebArchivists. Retrieved February 21, 2013, from http://timeline.webarchivists.org/

Institutional Web Archiving

Ardvidson, A. (2001, September 8). *Kulturarw3: The Swedish WWW-archive.* Workshop presented at the 5th European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, Germany. Retrieved from http://bibnum.bnf.fr/ecdl/2001/index.html

Aubry, S. (2010). Introducing web archives as a new library service: the experience of the National Library of France. *Liber Quarterly*, *20*(2), 179–199. Retrieved from: http://liber.library.uu.nl/index.php/lq/article/view/7987

Bailey, S., & Thompson, D. (2006). UKWAC: Building the UK's first public web archive. *D-Lib Magazine*, *12*(1). doi:10.1045/january2006-thompson

Fellows, G., Harvey, R., Lloyd, A., Pymm, B., & Wallis, J. (2008). Separating the wheat from the chaff: Identifying key elements in the NLA .au domain harvest. *Australian Academic and Research Libraries*, *39*(3), 137–148. Retrieved from: http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=8217&local_base=GEN01-CSU01

Grotke, A., & Jones, G. (2010). DigiBoard: A tool to streamline complex web archiving activities at the Library of Congress. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 17–23). Vienna, Austria.

Gupta, A. (2001). *Preserving presidential library websites: A case study with the Franklin D. Roosevelt Library, Museum and Digital Archives* (Technical Report No. SDSC TR-2001-3. Retrieved from San Diego Supercomputer Center website: http://legacy.sdsc.edu/TR/TR-2001-03.pdf

Hoffman, S. (2008). Preserving access to government websites: Development and practice in the Cybercemetery. Paper presented at the World Library and Information Congress: 74th IFLA General Conference and Council, Québec, Canada. Retrieved from http://digital.library.unt.edu/ark:/67531/metadc67623/

Kenney, A. R., McGovern, N. Y., Botticelli, P., Entlich, R., Lagoze, C., & Payette, S. (2002). Preservation risk management for web resources. *D-Lib Magazine*, *8*(1). doi:10.1045/january2002-kenney

Lasfargues, F., Oury, C., & Wendland, B. (2008, September). Legal deposit of the French web: Harvesting strategies for a national domain. Paper presented at the International Web Archiving Workshop (IWAW), Aaarhus, Denmark.

Martin, K. E., & Eubank, K. (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*(1), 7–26. doi:10.1080/13614560701423638

Masanès, J. (2002). Archiving the deep web. Workshop presented at the 2nd ECDL Workshop on Web Archiving, Rome, Italy. Retrieved from http://bibnum.bnf.fr/ecdl/2002/BnF/BnF.html

Masanès, J. (2002). Towards continuous web archiving. *D-Lib Magazine*, *8*(12). doi:10.1045/december2002-masanes

McGovern, N. Y., Kenney, A. R., Entlich, R., Kehoe, W. R., & Buckley, E. (2004). Virtual Remote Control: Building a preservation risk management toolbox for web resources. *D-Lib Magazine*, *10*(4). doi:10.1045/april2004-mcgovern

Pearce-Moses, R., & Kaczmarek, J. (2005). An Arizona model for preservation and access of web documents. *Documents to the People (DTTP)*, *33*(1), 17–24.

Schostag, S., & FØnss-JØrgensen, E. (2012). Webarchiving: Legal deposit of internet in Denmark. A curatorial perspective. *Microform & Digitization Review*, *41*(3/4), 110–120.

Žabička, Petr. (2003). Archiving the Czech web: issues and challenges. Workshop presented at the 3rd ECDL Workshop on Web Archives, Trondheim, Norway. Retrieved from http://bibnum.bnf.fr/ecdl/2003/

Žabička, P., & Matejka, L. (2007). Czech web archive analysis. *New Review of Hypermedia and Multimedia*, *13*(1), 27–37. doi:10.1080/13614560701450383

Personal Web Archiving

John, J. L. (with Rowlands, I., Williams, P., & Dean, K.) (2010). *Digital Lives: Personal digital archives for the 21st century: An initial synthesis* (p. 259). Retrieved from http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis01a.pdf

Marshall, C. C. (2008). From writing and analysis to the repository: Taking the scholars' perspective on scholarly archiving. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 251–260). Pittsburgh, PA, USA: ACM.

Marshall, C. C. (2011). Challenges and opportunities for personal digital archiving. In C.Lee (Ed.), *I, digital: Personal collections in the digital era* (pp. 90–114). Chicago, IL: Society of American Archivists.

Marshall, C. C., McCown, F., & Nelson, M. L. (2007). Evaluating personal archiving strategies for internet-based information. In *Proceedings of Archiving 2007* (pp.151-156). Springfield, VA, USA: Society for Imaging Science and Technology.

McCown, F., Marshall, C. C., & Nelson, M. L. (2009). Why web sites are lost (and how they're sometimes found). *Communications of the ACM*, *52*(11), 141–145. doi:10.1145/1592761.1592794

Thomas, S., & Martin, J. (2006). Using the papers of contemporary British politicians as a test bed for the preservation of digital personal archives. *Journal of the Society of Archivists*, *27*(1), 29–56.

Challenges in Web Archiving

Benczúr, A. A., Siklósi, D., Szabó, J., Bíró, I., Fekete, Z., Kurucz, M., Szabó, A. (2008). Web Spam: a survey with vision for the archivist. Paper presented at the International Web Archiving Workshop (IWAW), Aaarhus, Denmark. Retrieved from http://www.iwaw.net/

Cafarella, M., Chang, E., Fikes, A., Halevy, A., Hsieh, W., Lerner, A., … , J., Muthukrishnan, S. (2008). Data management projects at Google. *SIGMOD Record*, 37(1), 34-38. Retrieved from http://www.sigmod.org/publications/sigmod-record/

Charlesworth, A. (2003). *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: A study undertaken for the JISC and Wellcome Trust*. Retrieved from JISC website: http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf

Dasgupta, A., Ghosh, A., Kumar, R., Olston, C., Pandey, S., & Tomkins, A. (2007). The discoverability of the web. *Proceedings of the 16th International Conference on World Wide Web*, 421–430. New York, NY, USA: ACM. doi:10.1145/1242572.1242630

Hank, C., Sheble, L., & Choemprayong, S. (2009). *Considerations for the preservation of blogs*. Retrieved from http://www.digitalpreservationeurope.eu/publications/briefs/

Jacobsen, G. (2007, August). *Webarchiving internationally: Interoperability in the future?* Paper presented at the World Library and Information Congress: 73rd IFLA General Conference and Council, Durban, South Africa.

Jacobsen, G. (2008). Web Archiving: Issues and problems in collection building and access. *Liber Quarterly*, *18*(3/4). Retrieved from http://liber.library.uu.nl/index.php/lq/article/view/7936/8202

Kenney, A. R., McGovern, N. Y., Botticelli, P., Entlich, R., Lagoze, C., & Payette, S. (2002). Preservation risk management for web resources. *D-Lib Magazine*, *8*(1). doi:10.1045/january2002-kenney

McCown, F., Chan, S., Nelson, M. L., & Bollen, J. (2005). The availability and persistence of web references in D-Lib Magazine Magazine. Paper presented at the 5th International Web Archiving Workshop (IWAW05), Vienna, Austria. Retrieved from http://iwaw.europarchive.org/05/

McCown, F., & Nelson, M. L. (2009). What happens when facebook is gone? *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (pp. 251–254).

Meyer, E. T., Thomas, A., & Schroeder, R. (2011). *Web archives: The future(s)*. Retrieved from International Internet Preservation Consortium website: http://netpreserve.org/resources/web-archives-futures. doi:10.2139/ssrn.1830025

Rosenthal, D. S. H., Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005). Requirements for digital preservation systems. *D-Lib Magazine*, *11*(11). doi:10.1045/november2005-rosenthal

Spaniol, M., Mazeika, A., Denev, D., & Weikum, G. (2009). "Catch me if you can: Visual analysis of coherence defects in web archiving. In *Proceedings of the 9th International Web Archiving Workshop (IWAW), Corfu, Greece, September 30 - October 1, 2009* (pp. 27‐37).

Weikum, G., Ntarmos, N., Spaniol, M., Triantafillou, P., Benczúr, A., Kirkpatrick, S., Williamson, M. (2011). Longitudinal analytics on web archive data: it's about time! *CIDR 2011 5th biennial conference on innovative data systems research* (pp. 199–202).

Legal Issues in Web Archiving

Charlesworth, A. (2003). *Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia: A study undertaken for the JISC and Wellcome Trust*. Retrieved from JISC website: http://www.jisc.ac.uk/uploaded_documents/archiving_legal.pdf

Rauber, A., Kaiser, M., & Wachter, B. (2008). Ethical issues in web archive creation and usage – towards a research agenda. Paper presented at the 8th International Web Archiving Workshop, Aaarhus, Denmark. Retrieved from http://iwaw.europarchive.org/08/

Stirling, P., Illien, G., Sanz, P., & Sepetjan, S. (2012). The state of e-legal deposit in France: Looking back at five years of putting new legislation into practice and envisioning the future. *IFLA Journal*, *38*(1), 5 –24. doi:10.1177/0340035211435323

Practices, Standards, and Tools

Columbia University Libraries and Information Services. *Making the ephemeral endure: Collecting the web in research libraries*. (2011, April). Presented at the Association of College Research Libraries Annual Conference, Philadelphia, PA.

Curtis, J., Koerbin, P., Raftos, P., Berriman, D., & Hunter, J. (2007). AONS—An obsolescence detection and notification service for web archives and digital repositories. *New Review of Hypermedia and Multimedia*, *13*(1), 39–53. doi:10.1080/13614560701423711

Donovan, L., Fallon, T., & Harder, G. (2011, November). *What is worth saving? Selection and curation in web archiving*. Paper presented at the iPres 2011 conference, Singapore.

Fitch, K. (2003). *Web site archiving – an approach to recording every materially different response produced by a website*. Paper presented at the 9th Australasian World Wide Web Conference, Sanctuary Cove, Queensland, Australia. Retrieved from http://ausweb.scu.edu.au/aw03/papers/fitch/paper.html

Hockx-Yu, H., Crawford, L., Coram, R., & Johnson, S. (2010). Capturing and replaying streaming media in a web archive – a British Library case study. *Proceedings of the 7th International Conference on Preservation of Digital Objects iPres* (pp.339-343). Vienna, Austria: iPres 2010.

Gomes, D., Santos, A. L., & Silva, M. J. (2006). Managing duplicates in a web archive. *Proceedings of the 2006 ACM Symposium on Applied Computing* (pp. 818–825). New York, NY, USA: ACM. doi:10.1145/1141277.1141465

Grotke, A. (2008). *International Internet Preservation Consortium 2008 member profile survey results*. Retrieved from International Internet Preservation Consortium website: http://netpreserve.org/sites/default/files/resources/Membersurvey.pdf

ISO Technical Committee ISO/TC 46. (2009). *Information and documentation - WARC file format* (No. ISO 28500:2009). Geneva, Switzerland: International Organization for Standardization. Retrieved from International Organization for Standardization website: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44717

Jaja, J., & Song, S. (2009). Robust tools and services for long-term preservation of digital information. *Library Trends*, *57*(3), 580–594. doi: 10.1353/lib.0.0053

Mazeika, A., Denev, D., Spaniol, M., & Weikum, G. (2010). The SOLAR System for sharp web archiving. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 24–30). Vienna, Austria.

Minor, D., Kozbial, A., Burek, M., & Smorul, M. (2009). *Chronopolis: Preserving our digital heritage*. Retrieved from San Diego Supercomputer Center website: http://chronopolis.sdsc.edu/publications/assets/docs/ipres_chronopolis.pdf

Minor, D., Moore, R., Zhu, B., & Cowart, C. (2007). Archiving, indexing and accessing web materials: Solutions for large amounts of data. Paper presented at the 7th International Web Archiving Workshop, Vancouver, Canada. Retrieved from http://www.iwaw.net/

Oita, M., & Senellart, P. (2010). Archiving data objects using web feeds. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 31–41). Vienna, Austria.

Oury, C., Steinke, T., & Jones, G. (2012, December). Ensuring long-term access to the memory of the web: Preservation Working Group of the International Internet Preservation Consortium. *International Preservation News*, *58*(December 2012), 34–37. Retrieved from: http://www.ifla.org/news/december-issue-of-international-preservation-news-just-released

Oury, C., & Peyrard, Sébastien. (n.d.). From the World Wide Web to digital library stacks: Preserving the French web archives. Paper presented at the 8th International Conference on Preservation of Digital Objects iPRES 2011, Singapore.

Pop, R., Vasile, G., & Masanes, J. (2010). Archiving web video. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 42–47). Vienna, Austria.

Pope, J., & Beresford, P. (2007). IIPC web archiving toolset performance testing at The British Library. *Ariadne*, (52). Retrieved from http://www.ariadne.ac.uk/issue52/pope-beresford/

Rosenthal, D. S. H., Lipkis, T., Robertson, T. S., & Morabito, S. (2005). Transparent format migration of preserved web content. *D-Lib Magazine*, *11*(01). doi:10.1045/january2005-rosenthal

Rumianek, M. (2013). Archiving and recovering database-driven websites. *D-Lib Magazine*, *19*(1/2). doi:10.1045/january2013-rumianek

Song, S., & JaJa, J. (2008). *Archiving temporal web information: Organization of web contents for fast access and compact storage* (Technical Report UMIACS-TR-2008-08).Retrieved from Digital Repository at the University of Maryland website: http://hdl.handle.net/1903/7569

Spencer, A., Sheridan, J., Thomas, D., & Pullinger, D. (2009). UK government web continuity: Persisting access through aligning infrastructures. *International Journal of Digital Curation*, *4*(1). doi:10.2218/ijdc.v4i1.82

Smith, J. A., & Nelson, M. L. (2008). Creating preservation-ready web resources. *D-Lib Magazine*, *14*(1/2). doi:10.1045/january2008-smith

Stawowczyk Long, A. (2009). *Long-term preservation of web archives — experimenting with emulation and migration methodologies* (IIPC project to evaluate emulation and migration as long-term preservation solutions for web archives). Retrieved from International Internet Preservation Consortium website: http://www.netpreserve.org/resources/long-term-preservation-web-archives-%E2%80%93-experimenting-emulation-and-migration-methodologies

Tahmasebi, N., Zenz, G., Risse, T., & Iofciu, T. (2010). Terminology evolution module for web archives in the LiWA context. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 55–62). Vienna, Austria.

Wu, P. H. J., Heok, A. K. H., & Tamsir, I. P. (2007). Annotating web archives—structure, provenance, and context through archival cataloguing. *New Review of Hypermedia and Multimedia*, *13*(1), 55‑75. doi:10.1080/13614560701423620

Web Archiving and Digital Libraries

Bermes, E., & Illien, G. (2009). Metrics and strategies for web heritage management and preservation. Paper presented at the World Library and Information Congress: 75th IFLA General Conference and Council, Milan, Italy.

Blue Ribbon Task Force on Sustainable Digital Preservation and Access Preservation. (2010). *Sustainable economics for a digital planet: Ensuring long-term access to digital information* (Final Report). Retrieved from http://brtf.sdsc.edu/

Center for Research Libraries, & OCLC Online Computer Library Center. (2007). *Trustworthy repositories audit & certification: Criteria and checklist.* Retrieved from Center for Research Libraries website: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Gregory, L. (2009). *The practice and perception of web archiving in academic libraries and archives*. Student paper, School of Information and Library Science, University of North Carolina at Chapel Hill. Retrieved from http://dc.lib.unc.edu/cdm/ref/collection/s_papers/id/1771

Guenther, R., & Myrick, L. (2007). Archiving web sites for preservation and access: MODS, METS and MINERVA. *Journal of Archival Organization*, *4*(1/2), 141–166. doi: 10.1300/J201v04n01_08

Lyman, Peter. (2002). Archiving the World Wide Web. In *Building a national strategy for digital preservation : issues in digital media archiving.* Washington, D.C.: Council on Library and Information Resources and Library of Congress.

Pitschmann, L. A. (2001). *Building sustainable collections of free third-party web resources.* Retrieved from Council on Library and Information Resources website: http://www.clir.org/pubs/reports/pub98/contents.html

Sharpe, R. (2010). Active preservation of web sites. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 48–64). Vienna, Austria.

Steinke, T. (2008). Harvester results in a digital preservation system. Paper presented at the iPres 2008 conference, London, United Kingdom. Retrieved from: http://www.bl.uk/ipres2008/presentations_day1/25_Steinke.pdf

Web, C., Pearson, D., & Koerben, P. (2013). Oh, you wanted us to preserve that?! statements of preservation intent for the National Library of Australia's digital collections. *D-Lib Magazine*, *19*(1/2). doi:10.1045/january2013-webb

Wright, R., Miller, A., & Addis, M. (2009). The significance of storage in the "cost of risk" of digital preservation. *The International Journal of Digital Curation*, *4*(3), 104–122.

Quality in Web Archiving

Denev, D., Mazeika, A., Spaniol, M., & Weikum, G. (2011). The SHARC framework for data quality in web archiving. *The VLDB Journal*, *20*(2), 183–207. doi:10.1007/s00778-011-0219-9

Gray, G., & Martin, S. (2013). Choosing a sustainable web archiving method: A comparison of capture quality. *D-Lib Magazine*, *19*(5/6). doi:10.1045/may2013-gray

Illien, G. (2006). *Sketching and checking quality for web archives : a first stage report from BnF*. Working draft, International Internet Preservation Consortium internal review. Retrieved from Bibliothèque Nationale de France website: http://bibnum.bnf.fr/conservation/bnf-qualityforwebarchives-feb06.pdf

Internet Archive. (2012). *QA checklist*. Retrieved from https://webarchive.jira.com/wiki/display/ARIH/QA+Checklist

Shallcross, M., & Deromedi, N. (2011). *Quality assurance for Bentley Historical Library Web Archives: Guidelines and procedures*. Retrieved from http://hdl.handle.net/2027.42/94162

Spaniol, M., Denev, D., Mazeika, A., Weikum, G., & Senellart, P. (2009). Data quality in web archiving. *Proceedings of the 3rd workshop on Information Credibility on the Web* (pp. 19–26). New York, NY. doi:10.1145/1526993.1526999

Research on Web Archives

Alnoamany, Y., Weigle, M. C., & Nelson, M. L. (2013). Access patterns for robots and humans in web archives. *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*. Indianapolis, Indiana.

Brügger, N. (2012). Historical network analysis of the web. *Social Science Computer Review*, *31*(3), 306–321. doi:10.1177/0894439312454267

Brügger, N. (2008). The archived website and website philology: A new type of historical document? *Nordicom Review*, *29*(2), 155–175.

Brügger, N. (2009). Website history and the website as an object of study. *New Media & Society*, *11*(1-2), 115–132. doi:10.1177/1461444808099574

Costa, M., & Silva, M. (2010). Understanding the information needs of web archive users. *Proceedings of the International Web Archiving Workshop IWAW 2010* (pp. 9–16). Vienna, Austria.

Dougherty, M., Meyer, E. T., Madsen, C. M., Heuvel, C. V. D., Thomas, A., & Wyatt, S. (2010). Researcher engagement with web archives: State of the art. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1715000

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web science: an interdisciplinary approach to understanding the web. *Communications of the ACM*, *51*(7), 60–69.

Jackson, A. N. (2012). *Formats over time: Exploring UK web History*. [Conference paper]. Computer Science.Digital Libraries [arXiv:1210.1714]. Computing Research Repository, Cornell University Library.

Leung, S. A., Perl, S. E., Stata, R., & Wiener, J. L. (2001). Towards Web-scale Web Archaeology (SRC Research Report No. 174). Retrieved from HP Labs website: http://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-174.pdf

Murray, K. R., & Hartman, C. N. (2012). Classifying the End-of-Term Archive. *Archiving 2012 Final Program and Proceedings* (pp. 84–87). Retrieved from http://digital.library.unt.edu/ark:/67531/metadc93305/

Paudel, B., Anand, A., & Berberich, K. (2013). User-Defined redundancy in web archives. *10th International Workshop on Large-Scale and Distributed Systems for Information Retrieval* , 23–27. Retrieved from http://www.lsdsir.org/

Phillips, M.E., & Murray, K.R. (2013). Improving access to web archives through innovative analysis of PDF content. Paper presented at the IS & T – the Society for Imaging Science and Technology Archiving Conference, Washington, DC, United States. Retrieved from http://digital.library.unt.edu/ark:/67531/metadc155622/

Rauber, A., Bruckner, R. M., Aschenbrenner, A., Witvoet, O., & Kaiser, M. (2002). Uncovering information hidden in web archives. *D-Lib Magazine, 8*(12). doi:10.1045/december2002-rauber

Rauber, A., & Masanès, J. (2008). Report on the 8th International Workshop on Web Archiving - IWAW 2008. *D-Lib Magazine, 14*(11/12). doi:10.1045/november2008-rauber

Riberic, T. (2009). Automatic preparation of ETD material from the Internet Archive for the DSpace repository platform. *Code4Lib Journal, (8)*. Retrieved from http://journal.code4lib.org/articles/2152

Sanderson, R., Phillips, M. E., & Van de Sompel, H. (2011). Analyzing the persistence of referenced web resources with Memento. Paper presented at the Sixth Annual International Conference on Open Repositories, Austin, TX, USA. Retrieved from http://digital.library.unt.edu/ark:/67531/metadc39318/

Spaniol, M., & Weikum, G. (2012). Tracking entities in web archives: the LAWA project. *Proceedings of the 21st International Conference Companion on World Wide Web* (pp. 287–290). doi:10.1145/2187980.2188030

Stirling, P., Chevallier, P., & Illien, G. (2012). Web archives for researchers: Representations, expectations and potential uses. *D-Lib Magazine*, *18*(3/4). doi:10.1045/march2012-stirling

Thomas, A., Meyer, E. T., Dougherty, M., Heuvel, C. V. D., Madsen, C. M., & Wyatt, S. (2010). Researcher engagement with web archives: Challenges and opportunities for investment. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1715000

Weltevrede, E. (2009). *Thinking nationally with the web: A medium-specific approach to the national turn in web archiving*. (Master's thesis). Retrieved from https://wiki.digitalmethods.net/pub/Dmi/SummerSchool2009/weltevrede_national_webs.pdf

Weltevrede, E., & Helmond, A. (2012). Where do bloggers blog? Platform transitions within the historical Dutch blogosphere. *First Monday*, *17*(2). doi:10.5210/fm.v17i2.3775

Žabička, P., & Matejka, L. (2007). Czech web archive analysis. *New Review of Hypermedia and Multimedia, 13*(1), 27–37. doi:10.1080/13614560701450383