WebArchiving@UNT

Web Archiving Bibliography 2013

Prepared By

Brenda Reyes Ayala
Brenda.Reyes@unt.edu

June 28, 2013
<table>
<thead>
<tr>
<th><strong>Document Information</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
</tr>
<tr>
<td><strong>Original Creation Date</strong></td>
</tr>
<tr>
<td><strong>Version</strong></td>
</tr>
<tr>
<td><strong>Date of Current Version</strong></td>
</tr>
<tr>
<td><strong>Revised By</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Rights Information</strong></td>
</tr>
</tbody>
</table>

**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 ........................................................................................................................................... 1  
Towards a Definition of Web Archiving .......................................................................................... 1  
A Brief History of Web Archiving ................................................................................................... 1  
Literature About Web Archiving ..................................................................................................... 2  
How this Bibliography is Organized .................................................................................................. 2  
Bibliography ..................................................................................................................................... 3  
Introduction to Web Archiving ........................................................................................................ 3  
Institutional Web Archiving ................................................................................................................ 4  
Personal Web Archiving .................................................................................................................... 5  
Challenges in Web Archiving ............................................................................................................. 6  
Legal Issues in Web Archiving .......................................................................................................... 7  
Practices, Standards, and Tools ......................................................................................................... 7  
Web Archiving and Digital Libraries .................................................................................................. 9  
Quality in Web Archiving .................................................................................................................. 10  
Research on Web Archives ............................................................................................................... 10
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


**Institutional Web Archiving**


**Personal Web Archiving**


Challenges in Web Archiving


Legal Issues in Web Archiving


Practices, Standards, and Tools


### Web Archiving and Digital Libraries


Quality in Web Archiving


Research on Web Archives


