

Marilyn E. Rackley. Navigating the Web Archives: A Study of Users' Understanding of Context. A Master's Paper for the M.S. in L.S degree. July, 2007. 95 pages. Advisor: Christopher Lee

Web archiving is a rapidly growing area of electronic records archiving, with third-party service providers developing comprehensive Web archiving solutions. Currently in the United States, there are two major Web archiving services being used, the Internet Archive's Archive-It and OCLC's Web Archives Workbench. These two services are based, respectively on the "technocentric" and "archival" approaches to archiving Web sites, underlying which are specific assumptions about the nature of capturing, managing, and providing access to this type of archival material.

Focusing on how these approaches affect the access and presentation methods supported by Archive-It and the Web Archives Workbench, this paper describes a study conducted at the North Carolina State Archives and Library that tested the effect of the two methods on users' understanding of contextual information. Study participants' responses indicate that the "archival" model may provide users with a better understanding of a records context, but that generally users are confident about their ability to understand the records regardless of access methods. Nevertheless, despite these high levels of confidence, participants in this study did not necessarily have a good understanding of the nature of materials captured and archived directly from the Internet. The results of this study also demonstrate that users would like to have contextual information built into document displays, whether in the form of a header containing appropriate metadata or in the documents themselves.

Headings:

Archives/Cataloging

Metadata

Electronic Data Archives/Conservation and Restoration

Internet/United States

NAVIGATING THE WEB ARCHIVES: A STUDY OF USERS' UNDERSTANDING
OF CONTEXT

by
Marilyn E. Rackley

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

July 2007

Approved by

Christopher Lee

CONTENTS

ACKNOWLEDGEMENTS.....	3
INTRODUCTION	3
ARCHIVING ELECTRONIC RECORDS	6
DESCRIPTION IN AN ELECTRONIC AGE.....	8
MACHINE-READABLE DESCRIPTION OF PAPER RECORDS	10
DESCRIPTION OF ELECTRONIC RECORDS.....	13
ARCHIVING WEB SITES.....	16
TOOLS FOR WEB ARCHIVING: ARCHIVE-IT AND WEB ARCHIVES WORKBENCH.....	19
WEB ARCHIVING WORKFLOW MODELS	22
ACCESSING WEB ARCHIVES	24
THEORIES BEHIND WEB ARCHIVING.....	29
THEORIES PUT INTO PRACTICE: WAW AND ARCHIVE-IT	29
CRITIQUES.....	30
USER STUDIES IN ARCHIVES	35
A USER STUDY IN VIRTUAL ARCHIVES.....	41
METHODOLOGY	42
FINDINGS.....	52
DISCUSSION.....	59
LIMITATIONS.....	62
CONCLUSION	66
BIBLIOGRAPHY	69
APPENDIX A PARTICIPANT INSTRUCTIONS (WEB SITE ARCHIVES).....	73
APPENDIX B PARTICIPANT INSTRUCTIONS (STATE LIBRARY CATALOG).....	77
APPENDIX C MARS RECORDS FOR ABC COMMISSION RECORD SERIES.....	80
APPENDIX D STATE LIBRARY CATALOG RECORDS FOR ABC COMMISSION RECORD SERIES	81
APPENDIX E NORTH CAROLINA STATE GOVERNMENT ARCHIVES PORTAL AND WAYBACK MACHINE PAGES FOR THE ABC COMMISSION WEB SITE.....	82
APPENDIX F STUDY CONSENT LETTER	85

APPENDIX G STUDY RECRUITMENT SCRIPT..... 87

APPENDIX H CODING OF RESPONSES TO QUESTIONS IN SECTIONS A THROUGH D..... 88

APPENDIX I SUMMARY OF REPOSSES TO QUESTIONS IN SECTIONS E AND F FOR EACH
DOCUMENT 90

APPENDIX J SUMMARY OF COMMENTS FROM SECTION G 92

ACKNOWLEDGEMENTS

I would like to thank the many people at the North Carolina Department of Cultural Resources who helped make this study possible, especially, Chris Black, Kelly Eubank, and Dick Lankford at the State Archives and Eve Neville, Cheryl McKean, and Jennifer Ricker at the State Library; and all the study participants for their generous contributions of their time. I would also like to thank my advisor, Cal Lee, and my friends and family, Paul Pettengill, Catherine Rackley, Honor Rackley, Mary Rackley, Robert Rackley, Harry McKown, and Havivah Schwartz, who helped test the survey and made comments on this paper.

INTRODUCTION

The digital age has seen numerous changes in the way people work and live. People now have huge amounts of information readily available from their computers and they can use that information more quickly and efficiently than ever before for work and play. Archivists have experienced similar changes in their own personal and professional lives, but archivists also have a different set of challenges and opportunities to deal with in the digital age. Archivists, as intermediaries between records creators and records users, must respond to changes in the way people document business and personal activity now that networked computers are ubiquitous in the developed world.

Technological progress has resulted in changes in the nature of information materials created and preserved. It has also changed who has access to these materials and how they access them. Not only do more people have access, different people have access, broadening the audience for archival material. This change in audience is the result not just of technological changes, but societal changes as well (Forde 2005, Menne-Haritz 2001).

Research into the presentation of these resources to future users has been a low priority when the ability to ensure their survival into even the very near future has been doubtful. However, the number of records being created in the future is unlikely to decline, and archives are equally unlikely to receive a significant increase in resources to deal with the records. Moreover researchers now want and expect almost immediate remote access to archival materials in all formats and from all time periods. Consequently, it is important that archivists not put off discussing these issues indefinitely.

This study looks at two methods of approaching the processing and description of one kind of electronic records: records captured from the Web. It compares how each method affects the ability of researchers to understand the content and context of these records, both in the environment of their creation and the environment of their preservation. Because, as Thornsteinn Hallgrímsson mentions three times in his chapter on access and finding aids for Web archives, “[e]very current access tool needs to be improved and new tools must be developed” (Hallgrímsson 2006, 133, 141, 143), this study also looks more generally at the issue of describing electronic records from the user’s in order to provide suggestions for improvement. Taking two Web capture

services, Archive-It and the Web Archives Workbench, as representative of two distinct approaches to Web archiving, this study looks at their presentation to the end-user of state government records captured from the Web.

The findings described in this paper will be valuable to archivists trying to understand how this new type of record will be processed and presented to users. Continuing growth in the number of electronic records being created presents a challenge to archivists trying to efficiently process these records, and it is important to evaluate how new processing methods affect the users ability to interpret the records.

Because they will have less control over arrangement and description in their own institutions than they have had in the past, archivists will need user-centered research to convince outside parties of the need for changes in their products. Documentation concerning the use of archived Web sites for research and analysis through tools like the Wayback machine will allow archivists to influence third-party providers in the development of user interfaces for accessing these historic materials and to help them present their unique perspective on preserving and providing access to records of continuing and evidential value.

In contrast to some Web archiving activity and research, which focuses on the Internet as a mine of data that shows macro-cultural and technological trends (Rauber, Aschenbrenner and Witvoet 2002; Jatowt, et al. 2006), this study looks at Web archiving as an activity that preserves and protects electronic records. While many Web sites document the ever-changing nature and interests of society, an important subset of sites contain static records whose authenticity and reliability must be guaranteed in order for them to function as evidence of individual and organizational activity. This study is

about those records and how to make them available to researchers in ways that allow the researchers to evaluate the records properly.

ARCHIVING ELECTRONIC RECORDS

Electronic records present archivists with a new challenge because they cannot simply continue to manage records in the relatively passive way they have for the last several decades, when they could afford to wait for records to come into their institutions. Instead they must now be more actively involved in their preservation, and they must get involved earlier in the records lifecycle. Not so long ago, many archivists would not have considered electronic records as worth preserving because they were only temporary representations of more permanent paper records (Bailey 1989, 180), but archivists must now deal differently with electronic records that will never have paper counterparts, such as Web sites, for instance.

Archivists are beginning to address this challenge and to rethink some aspects of their profession. Thus, it is now generally accepted that to preserve electronic records, archivists will need to intervene much earlier in the records lifecycle than they have traditionally done, and preferably before the lifecycle even begins. They will need to involve themselves in the creation of recordkeeping systems in order to ensure that the records created within these systems will be authentic and reliable in the future (Hedstrom 1997, Hedstrom 1998). Even more importantly, perhaps, their participation in the design of systems and processes will help ensure that these fragile electronic records exist and are accessible in the future. The process of transferring the records to an archival institution will also need to satisfy the users' need for evidence of trustworthy

custodial practices, especially when third-parties are involved in the process and complicated technologies are being developed to address this need for trustworthiness in a new and unfamiliar recordkeeping environment (Hedstrom 1997, 62).

On the other hand, the problem of how archivists can best provide access to electronic records in the archives has not been as thoroughly addressed (Dryden 1995).¹ With paper records, archivists generally perform two basic tasks to gain intellectual and physical control over the records they collect. The first task is to accession the records. Accessioning involves creating a preliminary inventory and documenting basic information about the donor and the context of the original location of the records and their transfer to the archives. The second task, the more complicated one, is to arrange and describe the records. Arrangement and description involves the creation of a more detailed inventory of the records and the creation of an access aid (or aids), normally a finding aid, which allows the archives to manage and locate the records and allows users to discover and understand the records. It is unclear how these activities will be performed in the area of electronic records archiving. The growing backlog of unprocessed paper records, as well as the huge body of unaccessioned digital records that will hopefully find their way into an archival repository at some point, would, it seems, discourage archivists from trying to manage records in the same ways.

As the practices of records creators have changed, so too is it likely that the practices of records users and those who facilitate records use will change. So far, no

¹ One interesting project currently underway at the Library of Virginia is the development of an application to help archivists efficiently process and describe born-digital materials. The archivists in Virginia are testing the process on the papers of former governor Mark Warner (Chalfant and Jordan 2007).

particular methods for archiving electronic records -- covering records creation through secondary use -- have been developed to take the place of traditional processing methods. However, in the area of Web archiving, there are two specific approaches that have been gaining widespread acceptance as models for archiving that type of electronic record. These approaches to Web archiving will be examined in more detail below.

DESCRIPTION IN AN ELECTRONIC AGE

Description is one of the major archival functions. Usually closely related to arrangement, description “involves developing a summary ‘representation’ or access tool that includes information on the context in which the materials were created, their physical characteristics, and their informational content” (Roe 2005, 7). Description as an archival function is inseparable from other important functions, such as records management, preservation, and reference (Roe 2005, 8-10). Therefore, archivists must consider the impact of their descriptive choices on all areas of their profession.

Especially important in the digital era are description’s relationships with other areas of processing activity and reference activity. Because of the nature of access in this era, many archives patrons will use description as a replacement for reference contact with an archivist; and because of the nature of records in this era, archivists will need to describe records in a way that makes processing less time-consuming and more efficient (Greene and Meissner 2005). Any attempt to evaluate current or potential descriptive practices for all types of archival materials must take into account these changes in the nature of processing and reference.

Although description has not always had such a prominent role in controlling archives as it does today, Luciana Duranti explains that “[t]he primary purpose of writing about the records has gradually become in Europe (and is beginning to become in North America) that of illuminating provenancial and contextual relationships” (Duranti 1993, 51). In Duranti’s view of description, it does not merely serve to allow for discovery of records; it is essential for understanding those records. This function of description takes on new importance in an age when users may come to the archives with a variety of backgrounds and experience with archival materials. The unique purpose of archival description in contrast to other forms of bibliographic representation requires a unique approach to cataloging:

Archival description focuses on groups of materials rather than discrete items. The need to explain context of creation and use dictates collective description methods, and the voluminous quantities of modern records and manuscripts reaffirm this requirement. It would require countless years of staff-time to item-catalog most modern manuscript collections (Jimerson 2002).

In recent years, the most commonly-discussed form of description within the archival profession has been the finding aid, but there are many other ways to describe archival collections and items, including library catalog records, homegrown databases, repository guides, and national union catalogs. The finding aid is generally considered by archivists to be superior to other descriptive products, though, because of its ability to provide the contextual and provenancial information Duranti describes, in addition to basic control and discoverability.

Archivists need to respect the nature and purpose of archival description as they provide access to new records formats, but they can and should make changes in their descriptive practices to accommodate and take advantage of new technologies.

MACHINE-READABLE DESCRIPTION OF PAPER RECORDS

Thus far, archivists have primarily responded to the increased usage of computers for business and personal matters by making their traditional paper-based finding aids machine readable and available on a networked server so that remote users can access them. As the archivists working to develop the “next generation” finding aid point out, “[d]espite the transition from paper to electronic form, online finding aids retain much of the look and functionality of their paper counterparts and make only minimal use of available technologies, usually for browsing and searching” (Yakel, Shaw and Reynolds 2007). Recent developments in archival access aids have achieved the relative standardization of the finding aid format and content in order to promote usability and federated searching across finding aids and across repositories. This standardization work has produced Encoded Archival Description (EAD).

EAD is the subject of a great deal of the recent scholarship on archival description and access, including the contents of two whole issues of *The American Archivist* in 1997 and an issue of the *Journal of Internet Cataloging* in 2001. An EAD finding aid is not significantly different in structure or content from its paper counterpart, and many encoded finding aids are based on already existing finding aids (Hostetter 2004), although there has been some discussion about improving their labels and organization for EAD implementation (Meissner 1997). The benefit of the encoding is that it improves the ability of the computer to process and present information to users.

Encoded finding aids are touted for their flexibility in allowing users to customize their archival access aids and their ability to provide the user with as much or as little information as she needs to find and understand records and documents relevant to her

research. While “[t]he paper finding aid serves as a mirror...of the arrangement of the archival materials within collections, and insofar as the physical arrangement and intellectual inter-relationships of those materials remain stable, so too does the arrangement of the finding aid” (Gilliland-Swetland 2001, 230), the encoded finding aid can be manipulated in a variety of ways to enhance the discoverability and the accessibility of archival materials. Gilliland-Swetland offers suggestions on how to improve discovery and retrieval through EAD finding aids and suggests that perhaps the role of the finding aid for “finding” materials has been (too often) secondary to its role of explaining materials.

Users generally do not recognize and make use of the added benefits of EAD finding aids over other types of finding aids available on the Internet in formats such as Portable Document Format (PDF) (Roth 2001, 231). EAD implementation requires a significant amount of time and resources on the part of the implementing institution. Because of the current state of technologies and practices for using EAD have not developed enough to allow these institutions and their patrons to take full advantage of the benefits of XML-encoded finding aids, EAD has not significantly changed the reference practices of archivists or the access methods of users (Roth 2001).

What makes EAD different from other types of description/cataloging is its ability to present hierarchical relationships, but EAD alone is probably not enough to make archival materials significantly more discoverable. Putting finding aids and the materials themselves online may also require catalog records and other more widely understood access tools to complement the finding aids and will almost certainly require the creation of more metadata than paper records have (Hensen 2001).

The introduction of technologies designed to provide better access to archival collections across institutions has further complicated the problem of how to describe them. Christopher Prom describes the results of an experiment in providing access to EAD finding aids through the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH). He concludes that the stripped-down access aids that can be exposed to harvesters do not adequately convey contextual information to users and fail to describe the hierarchical nature of the underlying records they are supposed to represent (Prom 2003). One of the most important conclusions archivists can draw from his article is that making archival materials intellectually accessible in the digital era requires rethinking traditional means of representing primary sources in the digital era and not just making these representations machine readable so that they can be sorted and resorted in various ways intended to make them more discoverable.

Unfortunately, the significant amount of attention paid to Encoded Archival Description (EAD) and other machine-readable access aids comes at the expense of a more nuanced examination of the finding aid structure and its content. The focus on the encoded finding aid as an enhanced discovery tool leaves little room for considering its potential as intellectual access aid that plays a contextualizing role vis-à-vis the records being described. For instance, few archivists have considered how the information on the finding aid can be displayed with the records themselves (when they are available electronically, of course) on the same screen display to guide researchers as they view the records. It might also be useful to consider what parts of the finding aid researchers make the most use of and for what purposes in order to create better displays of records and access aids. Moreover, users really want both item-level description, from which

they can work backward to find contextual information, and collection-level description, from which they can work forward to individual items (Sweet and Thomas 2000). With archival access aids and documents available on the Internet, archivists could support both approaches, but they might have to explore less traditional gateways to born digital and born-again digital records.

As for the “next generation” finding aid, it took a group of people working for about a year to create an access aid that took full advantage of the capabilities of XML and Web 2.0, and they were working with primarily pre-existing metadata, including EAD finding aids (Yakel, Shaw and Reynolds, *Creating the Next Generation of Archival Finding Aids* 2007). While the resulting finding aid may provide an unprecedented level of intellectual access to one particular collection or even group of collections, as in this case with the Polar Bear Expedition collection, this model is clearly not one that every repository could or even would want to follow.

Even with the growing deployment of EAD and other machine-readable access aids, archivists are still unsure that the access aids they are creating are effective in helping users discover and understand archival materials, making it unclear how practical it is to expend resources on creating these descriptive products, unless the process becomes more automated.

DESCRIPTION OF ELECTRONIC RECORDS

Nor has the increase in the number of electronic records being created and archived changed the nature of access aids. Although archives are beginning to accession electronic records into their collections, most of the processing activity in these

institutions still centers around the processing of paper records, and when it is undertaken, the description of electronic records is often administratively separated from description of paper records (Dryden 1995, 103). Consequently, archivists generally have not been able to experiment with how well traditional archival access aids provide access to these records and what other types of access aids might be appropriate to represent collections of them.

Archivists attempting to provide access to electronic records have two major challenges to face: the lack of models for describing and presenting these records and the large quantities of records which they must make accessible.

Some archivists have argued that traditional methods of processing and describing records will not adequately translate into the digital era. They argue that the only way to effectively deal with the massive amount of digital records being created is by designing systems that have archival quality metadata built in (Wallace 1995, Hedstrom 1998, Dryden 1995). They argue that the requirements for long-term preservation and access should not be satisfied after the records have served their primary purpose as the result of additional activity by the archivist, but that satisfying these requirements should be a regular part of the business activity of records creators and managers. A well-designed documentation system could, in fact, provide organization-level metadata, as well as item-level metadata, to facilitate access to individual digital objects, but also adequately represent their context and provenance.

Although Hedstrom, Wallace, and others in this camp do not necessarily argue that old methods of describing records should be completely abandoned, the automated, metadata-centric approach they advocate is in opposition to the manual, carefully-written

narrative of the finding aid. While their recommendation is an important one, just having large amounts of metadata of any quality will not necessarily be enough to meet the access needs of users. The authors that advocate strongly for the need to include archival-quality metadata at the time of record creation do not address in detail the issue of how to make these records discoverable and comprehensible for secondary users. For Hedstrom, it seems that archivists will be needed to define requirements for archival-quality metadata, but it will be up to the primary creators and users to fulfill those requirements, freeing up the archivists' time to focus on context-based description rather than content-based description (Hedstrom 1993), while Wallace seems to think that systems could indeed capture content and context.

Responding to these ideas, Heather MacNeil posits that archivists theorizing about access to electronic records may be privileging content over context in their descriptive practices (MacNeil 1995, 24). The result will be archival description that focuses more on individual records and less on the overall environment in which the records were created and that makes access for secondary usage dependent on the access mechanisms designed to facilitate primary usage. Ultimately, for MacNeil, good metadata should contribute to effective records management during the active portion of the record lifecycle and support higher-level description once that part of the lifecycle is over.

Following Wallace and MacNeil's debate in the pages of *Archivaria* about metadata and description, Wendy Duff, who is a clear advocate of user perspectives informing archival practices, calls for more research to answer the questions "Do users want broad descriptions or would they prefer item level access without many levels of

description? Do they need the broad context provided by fonds and series level description, or will metadata systems that link content, context, and structure obviate the need for higher level description?” (Duff 1995, 36). To conclude this debate, Duff calls for research that compares the “retrieval performance of the two types of systems” (Duff 1995, 37). However, while discovery is an important aspect of electronic records accessibility, it is not the only aspect. Research should also compare the performance of the two types of system with respect to intellectual accessibility.

Because individual text-based electronic records are machine-readable, there is the potential to combine item-level description and collection-level description in a way that is manageable for archivists and helpful to users. Archivists will need to determine how they can combine different levels of access efficiently and effectively and will need to consider the nature of the description that will be associated with each level.

ARCHIVING WEB SITES

Although Web sites only represent a part of the electronic records universe, they are extremely important records for archivists to preserve. The Internet has radically changed communication practices among individuals and organizations of all types. It also has the potential to change long-term documentation and archiving practices. In an effort to better serve their users, state agencies are presenting a variety of information on the Internet. Records and documents that libraries and archives have worked so hard over the years to collect in physical copies are now more readily available for them to collect and use in the present and, at least theoretically, to preserve for long-term use (Eubank and Martin 2007). If records managers and archivists are able to capture and

preserve records directly from the Internet, they have the potential to implement retention schedules and appraisal decisions with greater precision and thoroughness than ever before. Organizations with legal requirements to collect and preserve institutional records are already responding to this opportunity and experimenting with harvesting the records and documents they find on the Internet. State archives, records centers, and libraries, for instance, see much potential in Web capture to fulfill a mandate to preserve government publications and other evidence of government activity – a task which has always been difficult when records creators were required to physically transfer records to these departments. In addition, the Web sites themselves are frequently defined as public records subject to records retention laws, making Web archiving an essential part of public records management activity.

Current Web capture applications allow archivists to pro-actively acquire records and documents remotely with very little assistance from records creators and managers.² Although some institutions have chosen to ask an organization's permission before harvesting its site, as the Library of Congress and the National Archives of the United Kingdom have, acquiring permission is not necessary for Web sites that are in the public domain (Brown 2006, Grotke and Ruth 2007). In North Carolina's case, state agency documents and records are not copyrighted, and while it may be necessary to enlist the help of agency Webmasters to ensure successful capture, the Archives and Library may harvest any Web sites produced by state government without going through the process of acquiring permission (Eubank and Martin 2007).

² This ability is a mixed blessing since archivists and libraries do not want to encourage other state employees to ignore their responsibility to cooperate with information management professionals to make government information available to citizens now and in the future.

In addition to the opportunities Web harvesting presents for archivists trying to collect records in compliance with a legal requirement, remote Web capture allows them to respond quickly to the ephemeral nature of Web documents. The first people to begin archiving Web sites were not, in fact, archivists with a duty to preserve organizational or personal records. They were technologists concerned with the rapid disappearance of large amounts of content on the Internet (Kimpton and Ubois 2006). The idea that Web content is especially at risk of being lost has inspired many of the Web archiving projects to date, including the Web-at-Risk project by the California Digital Library, New York University, and the University of North Texas, and the MINERVA Project at the Library of Congress.³

Nevertheless, even if this information can actually be captured through tools like Web crawlers, the work of preserving it for the future is far from straight forward and will require archivists and libraries to rethink many of their long-established practices. Because many archival institutions do not have the technological support they would need to create and maintain a Web archiving program in-house, they will, at least for the foreseeable future, need to use a third-party service provider to help them capture, preserve, and provide access to selected Web sites.

Currently in the United States, there are two major third-party providers of remote Web capture services. The Online Computer Library Center (OCLC) has created the

³ California Digital Library, "The Web at Risk: Preserving Our Nation's Cultural Heritage," <http://www.cdlib.org/inside/projects/preservation/Webatrisk/>. Date accessed: June 14, 2007. Library of Congress, "MINERVA Web Archiving & Preservation Project," <http://lcWeb2.loc.gov/cocoon/minerva/html/minerva-home.html>. Date accessed: June 14, 2007. The goal of the Web-at-Risk project is to create a Web archiving application, while the goal of the MINERVA project is to create event-based Web site archives, but both projects explicitly state that they are responding to the problem created by the short lifespan of most Web sites.

Web Archives Workbench (WAW) for Web harvesting, and the Internet Archive has created Archive-It.⁴ Both services offer Web-based applications for creating and managing collections of materials archived remotely from the Web, with the actual crawlers and captured files being maintained by the service providers. Despite these similarities, the WAW and Archive-It services are based on two very different models for Web archiving. Consequently, their features, workflows, and access and discovery methods are markedly dissimilar. These differences have important consequences for all stages of the Web archiving process, from collection development to use by researchers. Although the purpose of this study is compare how well each service provides intellectual access to Web archives, it is important to be familiar with the basic features of WAW and Archive-It, in order to understand the types of access and presentation methods they support.

TOOLS FOR WEB ARCHIVING: ARCHIVE-IT AND WEB ARCHIVES WORKBENCH

ARCHIVE-IT

The Internet Archive created Archive-It specifically “for institutions that have been mandated to preserve content from the public Web but do not have the IT infrastructure or technical staff necessary to meet that mandate at the current time.”⁵ For a fee, partner institutions can use the Archive-It interface to initiate and document crawls of Web sites they deem to be within the scope of their collecting missions. They are also

⁴ There are a number of open source Web capture tools, including many of the tools used by both the Internet Archive and OCLC, but with few exceptions, most archives and libraries do not have the in-house technical support to configure, manage, and maintain these open source tools.

⁵ Archive-It, “Archive-It Questions,” <http://www.archive-it.org/public/faq#507>. Date accessed: February 27, 2007.

able to attach Dublin Core metadata elements to each Uniform Resource Locator (URL) - or seed, in Web archiving terminology -- used as a crawl starting point and to collections of seeds.

With Archive-It, archives and libraries are able to use the Internet Archive's technical infrastructure, but maintain some control over the content of their Web archives collections. The Internet Archive maintains its own collection of archived Web sites, donated by Alexa Internet, but this collection tends to be a broad and shallow snapshot of the Web as a whole. It often does not include the lower levels of a site where large amounts of important content are available. Crawls initiated in Archive-It by subscribers travel further into a Web site and attempt to capture every document (as the individual files making up the site are called) in the site. Archive-It crawls are also more focused and generally do not travel outside the domains specified in the seeds. The Internet Archive uses the open-source crawler Heritrix to capture Web sites. It controls the spider settings and maintains the crawler.

It also stores and provides access to the archived Web sites. Mike Burner and Brewster Kahle, at the Internet Archive, created the ARC file format in 1996 to efficiently store Web files harvested remotely. Archive-It uses this file format to store the files from partner-initiated crawls on the Internet Archive servers. Partners are able to purchase copies of their archived documents, but must install their own access application locally in order to properly access and display the files from their networks.

WEB ARCHIVES WORKBENCH

OCLC developed the Web Archives Workbench in collaboration with the University of Illinois and the Library of Congress primarily to collect electronic

publications. The basic functionality of WAW is very similar to that of Archive-It. Like Archive-It, WAW uses the Heritrix Web crawler to remotely harvest files available on the Internet. It also relies on user-specified URLs as the starting point for its crawls and attempts to harvest those entire sites, avoiding links outside of the user-specified domains.

While Archive-It was developed to capture entire Web sites and to render them exactly as they were on the live Web (or at least as closely as possible), WAW was initially developed to harvest individual publications and documents. Because it uses the Heritrix crawler, WAW can also capture and render entire Web sites, as well as more traditional documents available on the Internet.

WAW has many features to assist librarians and archivists in developing and managing their Web collections, including discovery tools and series-creation functionality.⁶ Sites harvested in WAW must be manually ingested into a repository before they can become publicly accessible. The archivist has control over which materials are ingested into her repository at the site, directory, or file level. Once the archivist selects the materials to include, they are ingested into OCLC's digital archive or another compatible repository. In contrast to Web sites captured using Archive-It, which are made available through the Wayback Machine as soon as the crawl is finished, Web sites captured with WAW not only need to be manually ingested, they must be cataloged before they become available to the public.

⁶ A detailed description of the discovery and analysis modules in WAW is beyond the scope of this paper, but more information can be found in the WAW User Guide (OCLC 2007).

Cataloging these sites with WAW involves creating a Dublin Core metadata record for the site and importing that record into a library catalog based on Machine-Readable Cataloging (MARC) format.⁷ Users then discover and access the archived Web sites through the library catalog, using keyword or structured searching to search the metadata of the catalog record. It is important to note that very little of this metadata is created automatically during the harvesting, either from the harvesting process or the Web site itself.

WEB ARCHIVING WORKFLOW MODELS

Archive-It and WAW support two very different models for Web archiving workflows and policies. Archive-It's approach to Web archiving emphasizes automation and the management of aggregates, i.e. collections of sites, over discrete objects, i.e. pages or files. WAW's approach depends heavily on labor-intensive, manual management and administration of collections of files. These applications respectively provide examples of the "technocentric" and "archival" approaches to Web archiving described by Richard Pearce-Moses and Joanne Kaczmarek (Pearce-Moses and Kaczmarek 2005).

With Archive-It, the archivist may set up collections of seeds and then set crawls to run automatically as often as necessary. Once the crawls have run, the archived sites are available immediately through the Wayback Machine. The archivist may choose to manually download the reports that summarize the results of an entire crawl, but the

⁷ Although there are other methods for making these materials accessible, WAW works most efficiently if this method is the one used. For more about other methods of providing access to Web sites in the Digital Archive, see the Limitations section below.

successful implementation of Archive-It does not require this step of the process. These reports do not even provide site- or page-specific capture results, only aggregated results. While in reality, managing a Web archiving program does require occasional manual intervention (for instance, to ensure that all seed URLs are still valid and within the scope of the collection), theoretically, an archiving program that uses Archive-It would be largely automated after the initial set-up.

With WAW, more flexibility and more features mean more work for archivists and librarians. Information professionals using WAW have a suite of tools that allow them to discover the documents within Web sites that fit their collecting scope. They also have a number of ways to capture and package this material. They may choose to capture entire sites, specific directories, or specific files. Each item or group of items from a Web site must be manually catalogued and relationships with other items in the collection established. This work takes place prior to ingestion into a digital repository.

Post ingest, the archivist must create catalog records, finding aids, or other access aids to provide public access to the materials. During the ingest process, WAW packages the selected files in an XML wrapper with their associated metadata, but it is up to the Archivist to make the XML encoded metadata available to the user in a manner that is understandable. The OCLC Digital Archives allows the archivist to create MARC Catalog records with relative ease, but these catalog records do not make use of the complex, hierarchical metadata that can be created in WAW.

Table 1 presents an overview of the key aspects of the workflow models supported by each application.

	Archive-It “Technocentric”	WAW “Archival”
Collection Development	Emphasis on breadth; capture everything approach	Emphasis on selectivity
Processing	Aggregate collection processing; mostly administrative work required	Individual site, series, or file processing,
Access	Immediate	Dependent on manual cataloging

TABLE 1 COMPARISON OF ARCHIVE-IT AND WAW WORKFLOW MODELS

ACCESSING WEB ARCHIVES

With Archive-It, users access the archived Web sites from the Internet Archive servers through the Wayback Machine. The Wayback Machine is an open-source, publicly-accessible online interface developed by the Internet Archive. The Internet Archive uses this interface for its own collection of archived Web sites and for the collections its Archive-It partners create. The main way (and the most effective way) to search for archived versions of a Web site is to enter the URL into the Wayback Machine. The Wayback Machine then returns a list of dates on which that particular URL was captured and archived (a screenshot of the Wayback Machine page for the Metropolitan Museum of Art is shown in Figure 1). The user may then select the version she is most interested in, and with the click of the mouse, she is able to see how that Web site looked and what content it contained on the date that she selected. With the exception of a header at the top containing the name of the collection created by the partner institution, the Archive-It Wayback Machine page for any given URL looks the same as in the Internet Archive’s general collection (See Figure 13 in Appendix E for a screenshot of a Wayback Machine page with the Archive-It partner heading). Even

though a seed URL may have metadata attached to it through Archive-It, this metadata is not displayed at all in the Wayback Machine.

INTERNET ARCHIVE
Wayback Machine

Enter Web Address: All Adv. Search Compare Archive Pages

Searched for <http://www.metmuseum.org> 1289 Results

Note some duplicates are not shown. [See all](#)
* denotes when site was updated.

Search Results for Jan 01, 1996 - Jun 22, 2007											
1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
2 pages	7 pages	3 pages	9 pages	19 pages	157 pages	8 pages	28 pages	192 pages	463 pages	77 pages	0 pages
Nov 11, 1996 *	Feb 17, 1997 *	Jan 18, 1998 *	Jan 17, 1999 *	Mar 01, 2000 *	Jan 18, 2001 *	Aug 10, 2002 *	Jan 17, 2003 *	Mar 29, 2004 *	Jan 01, 2005 *	Jan 01, 2006 *	
Dec 19, 1996 *	Apr 16, 1997 *	Dec 01, 1998 *	Jan 25, 1999 *	Mar 03, 2000 *	Feb 02, 2001 *	Sep 01, 2002 *	Jan 22, 2003 *	Mar 30, 2004 *	Jan 02, 2005 *	Jan 01, 2006 *	
	May 11, 1997 *	Dec 05, 1998 *	Feb 02, 1999 *	Mar 03, 2000 *	Feb 24, 2001 *	Sep 05, 2002 *	Feb 08, 2003 *	Jun 02, 2004 *	Jan 03, 2005 *	Jan 03, 2006 *	
	Jun 19, 1997 *		Feb 08, 1999 *	Mar 03, 2000 *	Feb 26, 2001 *	Sep 30, 2002 *	Feb 11, 2003 *	Jun 05, 2004 *	Jan 04, 2005 *	Jan 04, 2006 *	
	Jun 29, 1997 *		Feb 18, 1999 *	Mar 04, 2000 *	Mar 01, 2001 *	Oct 07, 2002 *	Feb 13, 2003 *	Jun 06, 2004 *	Jan 05, 2005 *	Jan 05, 2006 *	
	Oct 24, 1997 *		Feb 20, 1999 *	Mar 04, 2000 *	Mar 01, 2001 *	Nov 03, 2002 *	Mar 23, 2003 *	Jun 09, 2004 *	Jan 06, 2005 *	Jan 06, 2006 *	
	Dec 10, 1997 *		Apr 23, 1999 *	Apr 08, 2000 *	Mar 01, 2001 *	Nov 17, 2002 *	Mar 29, 2003 *	Jun 10, 2004 *	Jan 07, 2005 *	Jan 10, 2006 *	
			Apr 27, 1999 *	May 10, 2000 *	Mar 01, 2001 *	Nov 26, 2002 *	Apr 06, 2003 *	Jun 11, 2004 *	Jan 08, 2005 *	Jan 10, 2006 *	
			Apr 29, 1999 *	May 10, 2000 *	Mar 01, 2001 *		Apr 19, 2003 *	Jun 12, 2004 *	Jan 09, 2005 *	Jan 10, 2006 *	
				May 11, 2000 *	Mar 01, 2001 *		Apr 23, 2003 *	Jun 13, 2004 *	Jan 11, 2005 *	Jan 11, 2006 *	
				May 20, 2000 *	Mar 01, 2001 *		Apr 25, 2003 *	Jun 14, 2004 *	Jan 11, 2005 *	Jan 11, 2006 *	
				Jun 12, 2000 *	Mar 02, 2001 *		Jun 08, 2003 *	Jun 15, 2004 *	Jan 12, 2005 *	Jan 12, 2006 *	
				Jun 21, 2000 *	Mar 02, 2001 *		Jun 10, 2003 *	Jun 16, 2004 *	Jan 12, 2005 *	Jan 13, 2006 *	
				Aug 15, 2000 *	Mar 02, 2001 *		Jun 17, 2003 *	Jun 16, 2004 *	Jan 12, 2005 *	Jan 13, 2006 *	
				Aug 18, 2000 *	Mar 02, 2001 *		Jun 17, 2003 *	Jun 16, 2004 *	Jan 12, 2005 *	Jan 13, 2006 *	
				Oct 18, 2000 *	Mar 09, 2001 *		Jun 20, 2003 *	Jun 18, 2004 *	Jan 15, 2005 *	Jan 15, 2006 *	
				Nov 16, 2000 *	Mar 30, 2001 *		Jun 22, 2003 *	Jun 19, 2004 *	Jan 16, 2005 *	Jan 16, 2006 *	
				Dec 01, 2000 *	Apr 05, 2001 *		Jul 21, 2003 *	Jun 22, 2004 *	Jan 17, 2005 *	Jan 17, 2006 *	
				Dec 06, 2000 *	Apr 18, 2001 *		Jul 25, 2003 *	Jun 23, 2004 *	Jan 18, 2005 *	Jan 18, 2006 *	
					Apr 28, 2001 *		Aug 05, 2003 *	Jun 24, 2004 *	Jan 19, 2005 *	Jan 20, 2006 *	
					May 03, 2001 *		Aug 07, 2003 *	Jun 24, 2004 *	Jan 20, 2005 *	Jan 25, 2006 *	
					May 06, 2001 *		Oct 11, 2003 *	Jun 25, 2004 *	Jan 21, 2005 *	Jan 26, 2006 *	
					May 08, 2001 *		Oct 17, 2003 *	Jun 26, 2004 *	Jan 22, 2005 *	Jan 27, 2006 *	
					May 15, 2001 *		Oct 28, 2003 *	Jun 27, 2004 *	Jan 23, 2005 *	Jan 28, 2006 *	
					May 16, 2001 *		Oct 29, 2003 *	Jun 27, 2004 *	Jan 24, 2005 *	Jan 29, 2006 *	
					May 17, 2001 *		Nov 20, 2003 *	Jun 28, 2004 *	Jan 25, 2005 *	Jan 30, 2006 *	
					May 18, 2001 *		Dec 04, 2003 *	Jun 29, 2004 *	Jan 26, 2005 *	Jan 31, 2006 *	
					May 19, 2001 *		Dec 16, 2003 *	Jun 29, 2004 *	Jan 27, 2005 *	Feb 02, 2006 *	
					May 20, 2001 *		Dec 18, 2003 *	Jun 29, 2004 *	Jan 29, 2005 *	Feb 02, 2006 *	
					May 22, 2001 *			Jun 30, 2004 *	Jan 30, 2005 *	Feb 02, 2006 *	
					May 23, 2001 *			Jul 01, 2004 *	Jan 31, 2005 *	Feb 03, 2006 *	
					May 29, 2001 *			Jul 02, 2004 *	Feb 03, 2005 *	Feb 03, 2006 *	
					May 30, 2001 *			Jul 03, 2004 *	Feb 04, 2005 *	Feb 04, 2006 *	

FIGURE 1 WAYBACK MACHINE PAGE FOR THE METROPOLITAN MUSEUM OF ART

While entering the URL into the Wayback Machine is currently the only way to access materials in the Internet Archives' general collection, Archive-It does offer a search feature and an overlay for the Wayback Machine in the form of a collection- or institution-specific portal. The ability to create such a portal allows partner institutions to provide a bit more information about their collection and its contents. The portal for the North Carolina State Government Web Site Archives, for instance, is made up of several WebPages describing the collection's background, providing answers to frequently asked questions, and listing the agencies whose Web sites the institution is capturing (screenshots of this portal are in Appendix E). Instead of requiring the user to know the URL for the state agency Web site she is interested in viewing, the Web Site Archives portal allows the user to click on the name of the agency or division within the agency to

get to the results page for the Wayback Machine. A custom collection portal gives archivists some flexibility in how they present the archived Web sites to users, but the process of creating it requires technical and design expertise and time. These portals are not a feature of the Archive-It product itself. If archivists choose to provide users with the kind of description included in traditional finding aids or other metadata besides Web site capture date, they must add these features outside of the Archive-It and Wayback Machine environment.

The portal also offers a full-text search engine that only searches within a particular collection. However, so far, this search engine does not successfully find relevant results.⁸ The Internet Archive is working to improve this search functionality, but if a researcher chooses a document found by the search engine, she is taken directly to that document, bypassing any contextualizing information, and even the Wayback Machine interface.

The Wayback Machine was designed specifically to access archived Web sites. As mentioned above, WAW's access mechanisms are based on traditional access aids. To provide access to materials in the Digital Archive, librarians and archivists can add hyperlinks to records in their OPAC, in-house databases, or finding aids. In this way, they have the ability to provide as much metadata or description as they choose, as their systems allow, and as their users need.

⁸ For example, the top result for a search using the keywords: Alcoholic Beverage Control Minutes, performed on March 7, 2007 was a document containing minutes for the Board of the Department of Transportation in October, 2003. The second result was a document containing regulations to obtain permits to test blood alcohol levels.

There is also a variation on the Wayback Machine, used primarily in Europe, that deserves to be mentioned here. “WERA (Web ARchive Access) is a freely available archive viewer application that gives an Internet Archive Wayback Machine-like access to web archive collections as well as the possibility to do full text search and easy navigation between different versions of a web page.”⁹ Figure 2 shows the main access page for WERA which allows users to begin their search either with a URL or keywords. Also included is a date range option that allows users to limit their search to a specific year or years. Figure 3 shows a Web page that was accessed through WERA. The banner at the top of the screen shows the version date as well as the original URL of the page and any other search terms used to find it.

⁹ “WERA Manual,” <http://archive-access.sourceforge.net/projects/wera/articles/manual.html#N1004C>.
Date accessed: June 20, 2007.

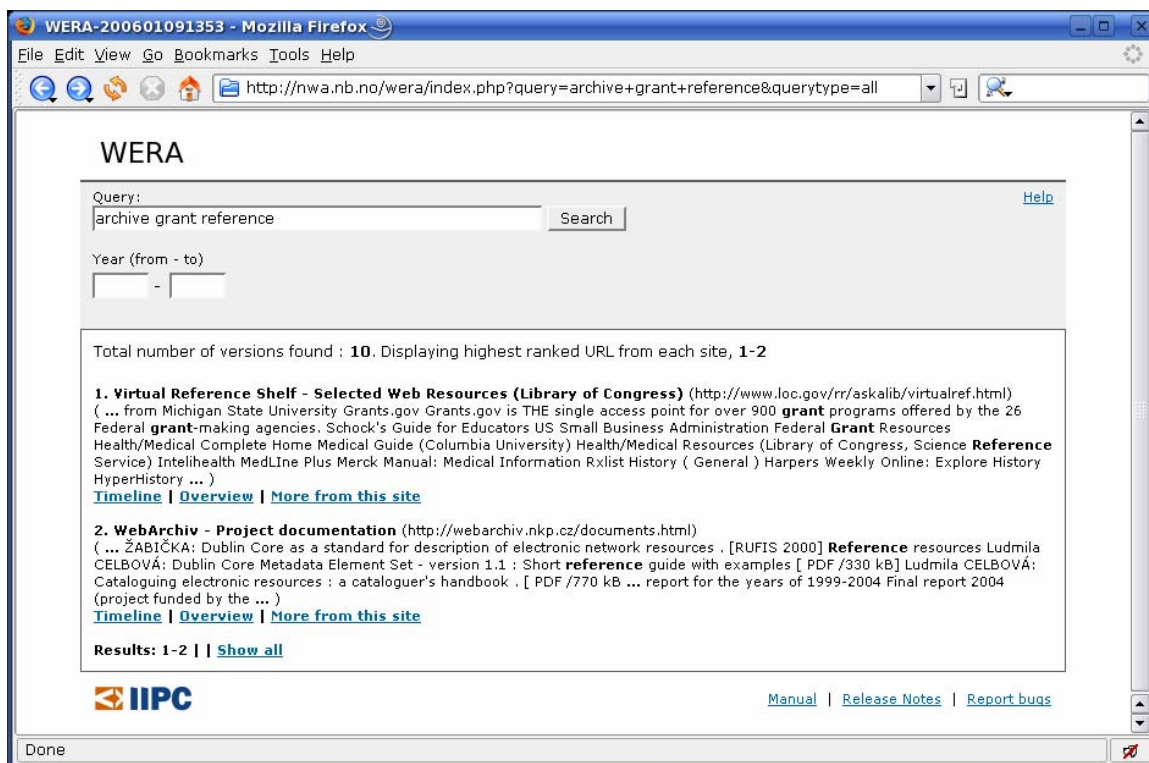


FIGURE 2 WERA SEARCH PAGE (FROM THE WERA MANUAL)

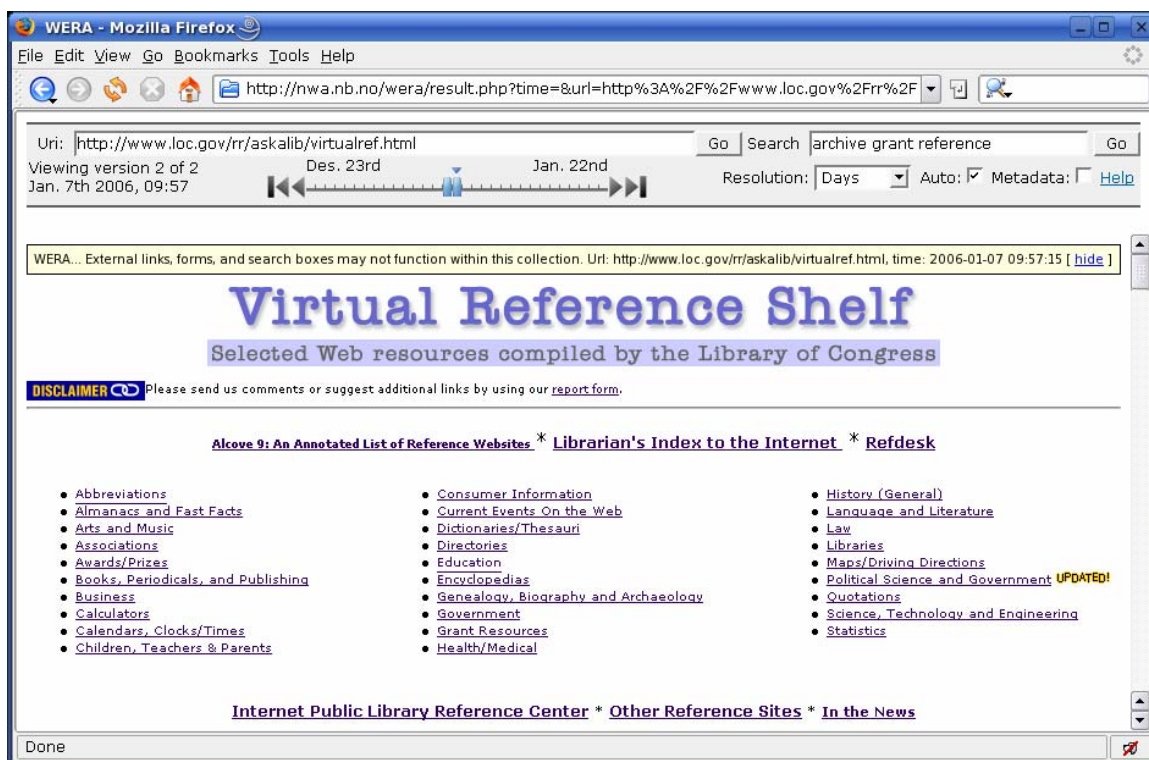


FIGURE 3 ARCHIVED WEB PAGE IN WERA (FROM THE WERA MANUAL)

WERA is not currently being used by any repositories in the United States, but because it does not require archivists to manually describe the captured Web sites and makes more of the metadata automatically captured during the archiving process easily accessible, it presents an attractive alternative to the Wayback Machine. However, since the Wayback Machine is still the more common of the two access tools, even in Europe, this study uses the Wayback Machine as the primary access tool for the “technocentric” approach to Web archiving.

THEORIES BEHIND WEB ARCHIVING

Web archiving is a young practice, and information professionals are still working to develop appropriate tools and methods, especially for preservation and access. Nevertheless, archivists and technologists have begun to work out systematic approaches to Web archiving built around comprehensive philosophies embracing every step of the process from capture to end-user access. When evaluating Web archiving access methods, it is important to understand these philosophies and how they affect access provision.

THEORIES PUT INTO PRACTICE: WAW AND ARCHIVE-IT

Dictating the (optimal) workflow models and access methods behind WAW and Archive-It are assumptions about the very nature of providing access to Web archives that have important implications for archivists and librarians. Brewster Kahle, Director of the Internet Archive, claims that “the Web is a self-documenting, self-cataloging machine” (Kahle 2002). Therefore, an approach to Web archiving that captures everything would essentially require no processing or description on the part of archivists

in order to make the material discoverable and intellectually accessible to users. Kahle's Internet Archive takes this approach. Its Archive-It service takes a similar approach, but applies it to smaller, institutionally-defined collections of Web sites, rather than the entire Web. The Internet Archive and Archive-It collections are both available through the Wayback Machine with minimal descriptive metadata. However, the idea that the whole Web is self-documenting and self-cataloging in a way that will be understandable to future users has not been tested, and it is unclear whether Kahle's description would also characterize discrete subsets of the Web taken out of the context of the whole.

Some archivists, on the other hand, argue that the records on the Internet need the kind of contextual explanations traditionally found in archival finding aids and other access aids, like bibliographic records. The Arizona Model for Preservation and Access of Web Documents advocates thinking about electronic records on the Internet in terms of series and describing them at the series level in hierarchically organized finding aids with scope and contents notes giving users background information to introduce them to the records (Pearce-Moses and Kaczmarek 2005).¹⁰ This model, an example of a modified "bibliocentric" model for Web archiving, draws on traditional library and archives practices.

CRITIQUES

¹⁰ The "archival" approach recommended as part of the Arizona Model is not the same as a "bibliocentric" model, but the two are similar in their emphasis on adapting traditional techniques and access aids for Web archiving. The Arizona recommendations are intended to be a medium between the "technocentric" and "bibliocentric" approaches. The archival approach allows Web archives curators to take advantage of some of the automation Web capture technologies support without complete abdicating their role as information organizers. However, WAW does not currently support the creation of complex, hierarchical access aids, such as finding aids, but it does support the creation of bibliographic records.

Both positions have merit at a time when librarians, archivists, and other information professionals are trying to fit information from the Internet into previously established categories of information resources. This type of information may fit into the category of publications, which are generally self-describing and more-or-less self-cataloging (in that the information needed to catalog them is usually found within the publication itself). It may also fit into the category of records, which are generally not considered to be self-describing and often require knowledge that is external to the documents themselves to make them intellectually accessible and to preserve their evidential character.

The debate over the “technocentric” access mechanisms and the “bibliocentric” or “archival” mechanisms is very similar to the debate captured in the exchange between David Wallace, Heather MacNeil, and Wendy Duff in the pages of *Archivaria*. On the one side are people like Kahle and Wallace who essentially believe that the records are best contextualized by their creators, while on the other are people like Pearce-Moses and MacNeil, who believe that records cannot (or will not) be described by their creators in a manner that will ensure future usability. However it is important to note a difference between the types of records the Wallace discusses and the types Kahle discusses, because the contextual information created by the interlinked nature of all Web resources is very different from the discrete metadata elements in Wallace’s ideal recordkeeping system. General approaches to electronic records archiving may not be appropriate for Web archiving. For instance, Bailey asserts that electronic records have their own organization that resides in the index that the computer system uses to access the files

(Bailey 1989, 187), but Web sites have an organizational structure that does not necessarily reflect the underlying file directory structure.

In general, traditional cataloging has been the subject of a great deal of criticism in the last few years for failing to keep up with changes in how people discover and assess resources. Critics and users alike identified subject-based access as an area that needed significant improvement. It may even be true that in the 21st century, “[a] complex metadata surrogate describing resources in detail is unneeded when the actual item can be viewed within a few seconds and with little effort on the part of the user” (Coyle and Hillmann 2007).¹¹

Julien Masanès critiques using the traditional cataloging approach for Web archives because, as he asserts, libraries and archives have only a small role to play in the global effort to edit and provide access to the Web. This effort constructively involves everyone using the Web, and especially those creating new content. Librarians and archivists who attempt to continue in their roles as “information organizers” will only serve “to freeze and preserve their own sample of a larger living culture artifact” (Masanès 2006, 20). In Masanès view, this result is not only undesirable, it is one that cannot be successfully sustained as the amount of material to be cataloged grows exponentially.

Hallgrímsson also argues that cataloging is not appropriate for Web sites archives, although for different reasons. “The experience from the Pandora and the Minerva projects shows that except for a very small Web archive traditional cataloging is not an

¹¹ Intentionally or not, these assumptions underpin the approach of the Internet Archive.

option” (Hallgrímsson 2006, 134), and certainly, those who have experience in Web archiving know that selecting, capturing, and cataloging Web sites is time-consuming and expensive (Murray and Phillips 2007). Hallgrímsson also makes the point that the metadata automatically harvested and the metadata automatically added during harvesting will not be enough for effectively providing access in the context of Web archiving. His primary focus is on discovery, and he discusses search engines in depth as a solution to the problems of Web archives accessibility.

A group of researchers developing a specialized browser specifically for viewing Web archives has suggested that “an interface similar to those for the live Web seems the most convenient for Web archives. The interface should support fast and easy information retrieval and browsing,” with the caveat that “current interfaces to Web archives are still in the development phase and do not necessarily provide such access” (Jatowt, et al. 2006, 136).

Other researchers have questioned the assumption that users of archived Web sites are best served by the same tools as users of the live Web. Because archival materials have traditionally required collection-level description to explain them, direct item-level access may cause confusion or misunderstanding among users. “When indexing is done at the page level only, the inevitable result is that queries will return perhaps thousands of URLs from various sites, without any clear indication where to find the entry point of the site that contains that page, and thus contextualize it” (Guenther and Myrick 2006, 145). Especially as the amount of information in the Archives grows and the number of versions of each page grows, full-text searching becomes extremely cumbersome (Hallgrímsson 2006, 136).

Currently, Internet search engines do a poor job of determining relevance based on the date of the materials being indexed (Bar-Ilan 2006), but the temporal coverage of archival materials is extremely important to researchers using them (Duff and Johnson 2003, 87). Allowing users to limit their search results by capture date may offer a decent solution to this problem, but even within a given year, there may be many archived versions of a particular page. A good search engine alone will not be sufficient to provide effective access to Web archives.

Archivists at the Library of Congress advocate a variation of traditional access aids for Web archiving. They suggest using a METS (Metadata Encoding Transmission Standard) package to house descriptive metadata, technical metadata, and administrative metadata. The use of METS will allow for the effective presentation of the complex “tangle of relationships that is an archived Web site” (Guenther and Myrick 2006, 161). However, they do not describe how this metadata will be made available to researchers for discovery and access purposes.

Jatowt et al. present an alternative to the Wayback Machine, in the form of an interesting experimental browser that would allow users to rapidly assess and analyze changes in Web pages over time, but they focus primarily on change rather than on static documents or records. “[Their] approach is based on the belief that browsing of the past Web should be based on change management” (Jatowt, et al. 2006, 138), and their browser would not necessarily be appropriate for accessing and interpreting static records captured from the Web.

This study tests, with real researchers, the theories behind the ways in which Archive-It and WAW provide access to archived Web sites, starting from the premise

that any presentation of Internet-based records must first and foremost be able to convey both their evidential value and their informational value. Users of these records must be able to understand who created them, when, why, and who is responsible for their safekeeping, in addition to being able to understand the content of the record.

The experiment described below tests the applicability of Kahle's proposition to a subset of the Web and the applicability of traditional cataloging to Web archiving in general by studying users' understanding of records on the Internet that have been archived and made accessible through the Wayback Machine and through the State Library's catalog. The Wayback Machine attempts to contextualize information by preserving its original presentation on the Internet, while archival description attempts to recreate the original context through an archivist-created access aid.

USER STUDIES IN ARCHIVES

In the archival community, there are frequent calls for studies of archival materials users. Archivists do not have a full understanding of their users' needs, making it difficult for them to meet those needs.¹² More work needs to be done to determine how researchers of all types find, evaluate, and use primary sources, both digital and analog; and it is becoming increasingly important to study user needs and behavior in the digital age, when many researchers can use archival access aids and archival materials remotely and may never have any direct contact with an archivist. These remote users will be the

¹² See, for instance, "Does AMC Mean "Archives Made Confusing"? Patron Understanding of USMARC AMC Catalog Records" on the archival community's recognition of the need for more user studies (Spindler and Pearce-Moses 1993, 331-332).

most difficult to serve and current practices are unlikely to meet the needs of these “invisible researchers” (Hill 2004).

User studies were one of the three broad research areas identified in the Burlington Agenda as essential for “providing intellectual access to electronically published historical documents” (Dow, et al. 2001, 301).¹³ More specifically, of the eight research issues outlined in the Agenda, half of them involve ensuring that contextual information is effectively presented during resource discovery and use. Knowing how users interact with primary sources online has implications not just for descriptive practices, but it also has implications for collection development. In discussing issues that curators need to address in their collection planning and management, Hur-li Lee includes issues of access in an online environment; “[o]ther questions are of particular relevance in the online environment. How do users use electronic collections? Do they prefer navigating freely in the cyberspace by themselves or starting with a collection as they do in a traditional library?” (Lee 2000, 1112).

The user studies that have been conducted generally explore one of two issues. Either they attempt to determine how people find primary sources relevant to their research¹⁴ or they explore how researchers interact with traditional descriptive products, such as the finding aid or the catalog record, with a view toward assessing their “usability.” With a few notable exceptions, these studies do not examine how (and how

¹³ The Burlington Agenda was the result of a three-day meeting in 2000 of experts in documentary editing, electronic publishing, computer science, and library and information science to discuss intellectual access to historical documents available electronically. The other two broad research areas identified were publication management studies and studies of access to information.

¹⁴ A review of what archivists see as the major benefit of online finding aids concludes that they think the most important effect of putting access aids online is a general increase in visibility of archives and their collections (Hostetter 2004), as opposed to a better understanding of the documents.

well) people make sense of the records themselves, either based on their interactions with the records or interactions with their bibliographic representations.

For instance, a study of the finding aid usability for the POLARIS Project at the Florida State University Claude Pepper Library generally found that users appreciated having online access to the finding aid, but that with access aids readily available to remote users, they also expect to have access to item-level discovery tools and even to the items themselves (Altman and Nemmers 2001). This study did not investigate the issue of whether the participants were better able to understand the documents if they used the online finding aid rather than the paper finding aid to access them. A study by Wendy Duff and Penka Stoyanova examined the usability of a number of archival display systems and compared them with an archival display based on specific design guidelines, concluding that users preferred the standards-based display (Duff and Stoyanova 1998). Here again, the focus was on responses to document representations separated from what they were representing.

Even good usability studies are not sufficient to ensure that records are intellectually accessible. Access aids that are easy to use do not necessarily provide researchers with the information they need to evaluate and understand the records. Advocates of user-centered design admit that there are limitations to usability testing. In the words of one author, usability tests “should be used to help create intuitive, usable sites, not to test the need for or usefulness of a new service. It is certainly possible to create an extremely usable site that is not useful” (Lack 2006, 76).

It is questionable whether current archival access aids are even useful, much less usable. A number of studies show that researchers looking for relevant archival

collections do not make effective use of online resources to discover collections (Hamburger 2004; Anderson 2004; Duff, Craig and Cherry 2004; Tibbo 2003), even though archivists are keen to make the majority of their collections discoverable online.¹⁵ Finding aids, as one study concludes, are not well understood by researchers, and EAD does not make them any easier to understand (Yakel 2004). EAD finding aids were designed for historians and do not necessarily support the type of searching that genealogists do (Duff and Johnson 2003).

Moreover, the use of online finding aids presents its own problems. An experiment conducted by Tim Hutchinson demonstrated that full-text searching in finding aids yields greater recall but less precision in result sets than searching through short, more precise collection representations, such as the introductory portions of the finding aid (Hutchinson 1997). While Hutchinson concludes that EAD will facilitate more precise retrieval when it can be used for structured, field-specific searching, his study has important implications for the retrieval of documents themselves, as well as their descriptions. The results of searching for either documents and/or access aids through an Internet search engine are likely to have the same low precision, high recall characteristics as the results of the experiment with retrieval through full-text finding aids. Increased recall and decreased precision will require archivists to think carefully about how they help remote users decide if a given result is what they are seeking.

¹⁵ One very interesting study, though it did not have real users as participants, simulated users' attempts to discover archival finding aids through Internet search engines. The authors determined that, at the time of the study, search engines rarely successfully found even known items, so that even if researchers were searching for finding aids online, their success rates would likely be low (Tibbo and Meho 2001).

Many of the user studies that scholars in the archives community have conducted focus on researchers in academic environments, primarily professional historians and students, paying less attention to the non-academic professionals and recreational researchers that make up a large portion of archives' user population. For example, the Gladys Kriebel Delmas Foundation funded three large-scale studies of historians in the US, UK, and Canada (Anderson 2004; Duff, Craig and Cherry 2004; Tibbo 2003). A recent article purporting to explain the basics of user studies was similarly limited in the scope of its audience (Proffitt 2006).

Historians generally use print or informal sources of information to find primary source materials (Anderson 2004; Duff, Craig and Cherry 2004; Tibbo 2003). One interesting study of another user group, genealogists, by Wendy Duff and Catherine Johnson found that they also generally preferred informal sources of information rather than formal sources, such as finding aids and collection guides (Duff and Johnson 2003, 94-95). It is unclear whether and how these discovery methods apply to Web archives and other electronic records.

The user studies conducted to date in the archival community have been either exploratory in nature or very specifically tied to one particular situation or descriptive product. Three very different studies have delved into the problem of providing access to electronic records in extremely interesting and useful ways. Margaret Hedstrom and others working on the CAMiLEON project at the University of Michigan tested electronic records preservation from the user's perspective in an attempt to discover what properties of the original digital objects were important to users. The experiments described in the resulting article tested a computer game and textual documents. While

the results indicated that contextual information was, indeed, important for using and understanding archived digital objects, the authors also noted that the study participants did not make use of the metadata associated with the textual documents to interpret them (Hedstrom, Lee, et al. 2006, 184, 187).

In a study dealing specifically with access to Web site archives, a group of researchers in Japan developed an interactive browser that would allow researchers to access Web sites archived in multiple repositories and to view the various versions chronologically in a slideshow-like manner (Jatowt, et al. 2006). They designed the browser, the “past Web browser,” specifically to highlight the changes in the site over time, rather than to provide a static record of a site at a particular time.¹⁶ The researchers then conducted an experiment in which they had users complete a set of tasks using their browser and using the Wayback Machine. The results suggest that the past Web browser allowed for more efficient and effective completion of the study tasks. However, the tasks in the experiment concentrated on identifying and evaluating changes in a Web page over time and the results may not be indicative of how the two access tools would compare if the tasks evaluating records at a specific point in time.

A study of user understanding of MARC AMC records conducted by Robert Spindler and Richard Pearce-Moses presents findings that, like Prom’s article (Prom 2003), show how difficult it is to recreate the contextual information contained in more complicated archival access aids in the type of access aids found in library catalog

¹⁶ A much simpler viewer called “Web Tour” was used by Brian Kelly for a survey he conducted in 2002 of the availability of historic Web sites in the Wayback Machine. The Web tour is essentially an automatic slide show of older Web sites, with some user control of the pace, but without the change analysis features of the past Web browser (Kelly 2003).

(Spindler and Pearce-Moses 1993). Both of these articles show that it is not easy to replicate the kind of contextual information needed to understand and evaluate archival records in a way that is easier to integrate into the larger world of information available in many formats and in many types of repositories and in a form that is actually less time-consuming for archivists to create and manage.

User studies in the 21st century should still address the issue of resource discovery, but they also need to address aspects of user understanding, as these last three studies do. Moreover, they should include a broader definition of archival users, so that archivists will be effectively serving all types of users, some of whom may be remotely using their collections without their knowledge.

A USER STUDY IN VIRTUAL ARCHIVES

This study investigates the issue of users' understanding of the context of records in Web archives through a scenario-based test, which "involves presenting representative end-users with scenarios, or specific tasks, designed to cover the major functionality of the software system and to simulate expected real-life usage patterns" (Levi and Conrad 2002). It compares users' experiences of the access methods associated with the "technocentric" and "archival" approaches to Web archiving (as implemented in Archive-It and WAW, respectively), in order to test whether both methods of accessing archived Web sites are equally effective in presenting the contextual information users need to evaluate, understand, and trust the archived records. The "users" in this study are actual researchers who use archival materials, but they do not represent any particular subgroup of users.

METHODOLOGY

The participants who completed the study were researchers at the North Carolina State Archives and State Library. Although the participants did not provide any personal information as a part of this study, the user population for these two collections comprises students; academics; professionals, including other state employees; and recreational researchers. Most of the users live in the state of North Carolina, but the Archives and the Library also serve patrons from other states and other countries. Researchers range in age from college students to retirees and generally have a varying degree of experience with computers and primary source materials.¹⁷

For the study, each participant was asked to imagine that she was a researcher working in the year 2150 to follow very specific step-by-step instructions to find two different records, using either the State Library Catalog or the State Government Web Site Archives Portal and the Wayback Machine as the starting point for the search. The participant then completed a questionnaire that required her to explain the content and context of the records and to answer questions about how well she understood and trusted the records and how easy it was to locate contextual information about the records. The final section of the questionnaire asked the participant to reflect on the process of finding the information needed to explain the records' content and context. Each participant was exposed to only one of the two archival environments and had explicit instructions not to leave that environment for the live Web. All participants examined the same two records,

¹⁷ Students in grades K-12 also use the Archives and the Library, but because of the nature of the consent process, no one under the age of 18 participated in this study.

which were minutes of a meeting of the North Carolina Alcoholic Beverage Control Commission and a revenue report from the Commission.

There were four versions of the survey instrument. Two versions gave the participant instructions to use the State Government Web Site Archives to retrieve the records, but the order of retrieval was varied between those two versions. The other two versions directed the volunteer to use the State Library Catalog to retrieve the records, and again the order of retrieval was varied between those two sets. Thus one group of participants used the Web Site Archives to examine first the meeting minutes and then the revenue report, one group of participants used the Web Site Archives to examine first the revenue report and then the meeting minutes, one group used the library catalog to examine first the meeting minutes and then the revenue report, and one group used the library catalog to examine first the revenue report and then the meeting minutes.

Access Condition		
First Document	Wayback Machine (n=15)	State Library Catalog (n=8)
Meeting Minutes	Version 1 (n=9)	Version 2 (n=5)
Revenue Report	Version 3 (n=6)	Version 4 (n=3)

TABLE 2 QUESTIONNAIRE VERSIONS

Setting up this study required four main steps: selecting two sample records created by the state of North Carolina available on the Internet, capturing those two records using the Web Archives Workbench and Archive-It and adding the metadata allowed in each tool, developing the participant instructions and questionnaire, and recruiting participants.

SETTING UP THE STUDY

The first phase of the study entailed setting up the study environment, including choosing the records, archiving the records with WAW and Archive-It, describing them, and writing the participant instructions and questionnaire. I used the following five criteria to choose the records:

1. The records belong to a series with an existing description in the State Archives' online catalog MARS;
2. The records have continuing value;
3. The organizational structure of the agency creating the records had not significantly changed since the description was written;
4. The records would not be too familiar to the average researcher;
5. The records are short enough for researchers to read and analyze both within twenty minutes.

I chose records on the basis of whether a description of the corresponding record series already existed in the Archives' online catalog MARS and whether the records had continuing value. Unfortunately, many state records have not yet been processed and described, limiting the types of records that could be used for this study, because it was not feasible to create new descriptions from scratch in a reasonable time frame. I considered any record that fit into a record series with an existing description in the MARS catalog to have continuing value and therefore, appropriate for this study. Because the MARS records were used to create the library catalog records for this study, an additional criteria for inclusion in the study was that the record series had to continue up to the present day under the same organizational structure as described in the MARS record, so that the catalog record would accurately reflect the context of the records' creation. Because these records would have had no time to "age," it was especially important to choose records with which users would not be already too familiar, in order to recreate the relative foreignness of historical records.

To find the records I browsed the descriptions of State records' series in MARS. Many of the series with descriptions in the catalog are closed and consequently, inappropriate for this study. A number of other descriptions reflected historical agency organizational structures that would no longer apply to more recent records. The only series that fit all the selection criteria were in the records group for the Alcoholic Beverage Control Commission. From these series I chose the Commission's detailed meeting minutes series and the public revenue from alcoholic beverage series because records fitting into these series were currently available on the Commission's Web site and because they are short enough that participants could analyze them quickly during the study. From these series, I chose the meeting minutes for December 2006 and the Annual Revenue by ABC Board Report for July 1, 2005 – June 30, 2006, which can be seen in Figure 4 and Figure 5 below.

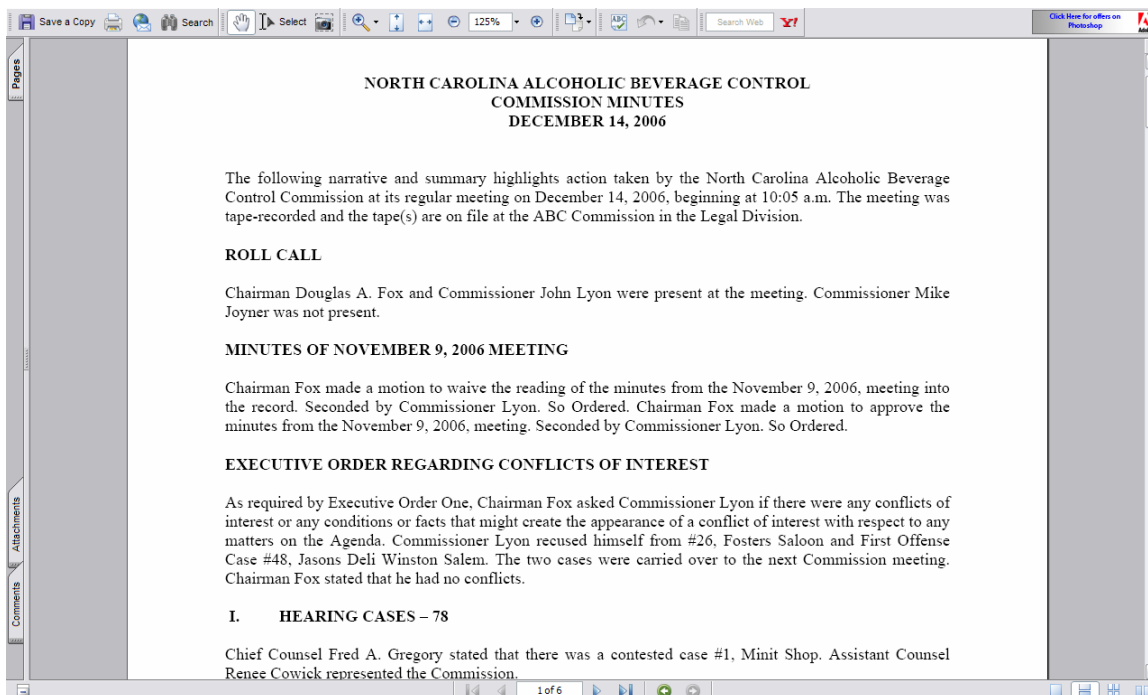


FIGURE 4 ALCOHOLIC BEVERAGE CONTROL COMMISSION MINUTES, DECEMBER 2006¹⁸

	Retail Sales	Mixed Beverage Sales	Wine Sales	State Excise Tax	Mix Bev Tax-Dept Revenue	Mix Bev Tax-DHR	Rehab Tax	Wine Sales Tax	Cost of Liquor Sold	Cost of Wine Sold	Operating Expenses	Interest Income	Other Income	Other Exp.
ABC Board														
Alamance Municipal	5,978,410	1,270,060	0	1,341,566	130,377	13,038	27,292	0	3,922,194	0	1,275,096	9,672	1,771	-5,163
Albemarle	1,713,960	214,247	0	9,133	361,862	21,793	2,179	7,461	1,048,904	5,514	387,922	4,893	0	9,911
Andrews	501,265	0	0	4,198	96,191	0	0	2,242	294	280,053	3,363	88,159	960	0
Angier	716,723	0	0	137,696	0	0	0	3,370	0	399,976	0	135,880	1,625	0
Asheville	12,736,173	5,380,396	58,086	3,288,710	501,729	50,173	61,877	1,976	9,570,332	43,942	3,314,484	66,998	1,232	346
Beaufort County	3,234,610	305,703	7,372	666,823	30,252	3,025	14,958	594	1,955,032	0	727,580	5,683	0	0
Belville	1,316,446	0	0	252,383	0	0	5,794	0	736,687	0	205,719	1,410	0	0
Bertie County	830,341	0	0	159,033	0	0	3,831	0	460,253	0	197,624	3,620	1,914	0
Bessemer City	489,877	0	2,591	93,611	0	0	2,720	174	274,021	1,988	105,000	802	0	0
Black Mountain	1,315,874	0	0	252,121	0	0	5,253	0	731,098	0	266,941	1,044	0	0
Blowing Rock	819,541	777,254	6,873	279,362	73,813	7,391	4,562	437	812,071	0	194,007	12,669	0	0
Bolling Spring Lakes	432,202	1,929	0	83,165	225	22	1,773	0	243,001	0	79,628	0	518	0
Boone	3,702,828	28,802	24,565	715,291	2,653	265	13,461	1,720	2,130,284	0	515,174	0	0	0
Brevard	1,852,208	395,680	22,522	417,350	37,818	3,782	7,531	1,473	1,211,299	15,429	374,326	517	107	0
Brunswick	392,265	0	0	74,954	0	0	1,984	0	216,904	0	72,902	469	0	0
Brunswick County	1,569,914	344,771	0	353,467	36,839	3,684	7,027	0	1,029,530	0	329,531	0	882	18,465
Bryson City	1,196,607	146,112	8,498	250,994	14,093	1,037	5,263	602	748,285	0	246,375	4,667	0	0
Bunn	609,446	0	742	116,922	0	0	2,577	52	341,216	0	106,613	0	0	0
Calabash	1,114,510	233,889	11,471	247,414	29,624	2,962	4,946	750	728,992	7,837	217,576	6,516	949	0
Camden County	931,976	0	2,964	178,394	0	0	4,121	194	515,921	1,510	180,512	3,202	0	0
Canton	840,614	20,517	4,566	165,576	2,232	223	3,765	0	479,610	3,114	176,821	0	2,385	13,874
Carteret County	7,285,129	1,939,389	2,174	1,696,302	197,719	19,772	31,347	152	4,961,947	1,517	1,342,968	19,792	0	4,698
Caswell County	1,509,805	0	0	289,115	0	0	7,195	0	808,693	0	350,559	1,640	0	0
Catawba County	10,007,561	2,290,675	36,549	2,271,425	231,680	23,168	45,728	2,391	6,861,309	21,695	2,187,271	6,952	0	0
Chatham County	1,613,461	60,944	11,170	319,748	5,027	503	5,505	782	943,589	0	351,844	991	0	470
Cherryville	728,324	1,793	2,429	139,107	224	23	3,358	165	408,485	1,531	163,473	2,734	0	0
Chowan County	891,696	74,130	1,259	182,499	7,356	736	3,940	88	554,087	0	202,037	2,795	0	2,150
Clinton	1,170,970	86,061	0	237,408	8,856	886	5,875	0	689,919	0	196,840	3	0	0
Columbus	448,594	8,076	5,372	86,809	911	91	1,805	377	256,739	3,438	99,079	0	295	0
Concord	6,341,448	1,407,315	39,310	1,433,893	143,302	14,330	28,129	2,572	4,239,817	0	1,374,142	62,363	14,198	0
Coolidge	562,565	8,080	1,625	109,767	852	85	2,483	106	316,575	0	141,944	113	3,600	2,410
Craven County	5,136,741	1,158,255	24,607	1,162,483	118,627	11,862	23,724	1,610	3,405,781	0	977,912	44,123	0	0
Cumberland County	13,176,666	5,005,808	50,191	3,295,729	504,517	50,451	70,533	3,513	9,944,641	31,903	2,763,336	83,578	1,159	0
Currituck County	2,937,949	326,417	55,481	614,612	30,347	3,035	10,621	3,629	1,801,751	25,142	438,441	1,412	0	0
Dare County	8,301,617	2,850,258	149,551	2,032,305	291,235	29,123	34,267	11,204	6,033,418	0	1,163,693	58,495	592,731	0

FIGURE 5 ALCOHOLIC BEVERAGE CONTROL REVENUE REPORT¹⁹

¹⁸ The URL for this document on the live Web is <http://reports.ncabc.state.nc.us/uploads/resources/504800ab581e4fab95b128579403e1d6.pdf>.

Once the selection of the records was complete, I manually harvested them from the live Web using WAW on February 6, 2007. I then ingested the two records into the Digital Archive and, with the assistance of a cataloger at the Library, I created MARC catalog records for each item in the State Library Catalog based on the corresponding MARS records (see Appendix C for the existing MARS records for each series and Appendix D for the catalog records created for this study).²⁰ In order to better simulate an actual description of these series including both paper and electronic records, I updated the holdings information and did not include the processing information in the MARS records in the catalog records. To further the illusion, the records in the library catalog also included links to other archived records in the series that were not part of the study. The record for the revenue reports contained links to two years worth of reports, and the record for the minutes included links to three months of minutes.²¹

The Commission's entire Web site was also automatically ingested into the North Carolina State Government Web Site Archives on February 2, 2007 using Archive-It, during the regularly-scheduled quarterly crawl run by the State Records Center. Because the Web sites captured during an Archive-It crawl are immediately made available through the Wayback Machine, no additional work was necessary to make the records for this study publicly accessible.

¹⁹ The URL for this document on the live Web is <http://reports.ncabc.state.nc.us/uploads/resources/114c19d793b34380978c7fd82407d6c2.pdf>.

²⁰ Although a MARC catalog record is not the exact equivalent of a MARS record, or other similar records in a database of archival access records, WAW primarily supports the creation of MARC records for the materials it harvests. In addition, I was not able to access the MARS database to add records, but I was given permission to add these records to the Library's catalog.

²¹ One of the drawbacks of the MARC records for WAW-harvested documents is the inability to give the links for each document a meaningful name. See the catalog records in Appendix D for an example. In this study, the instructions directed the participants to click on the links for the first document in each record.

I created instructions for each of the two conditions to direct participants in how to locate the sample records (Appendix A and Appendix B contain one set of instructions for each access condition). Participants using the records as captured by WAW were instructed in how to find the correct catalog records in the North Carolina State Library Catalog and then instructed to click on the links leading to the records in the Digital Archive. Participants using records captured with Archive-It were instructed in how to find the records starting from the North Carolina State Government Web Site Archives portal homepage and going through the Wayback Machine, which required them to navigate through several archived WebPages to reach the records themselves (see Appendix E for screenshots of the navigation steps). Although the two sets of instructions were very different, they were designed to simulate two probable methods for discovering and retrieving these records. It is inevitable that participants would gain some understanding of the content and context of the records in the process of finding them, so following the instructions was an integral part of the experience of evaluating and analyzing the records and should be seen as part of the overall presentation of the records in the archival environment.

Before seeking the approval of the Institutional Review Board and recruiting any participants, I tested each version of the instructions and questionnaire on either a family member or friend to ensure that the participants would be able to follow the instructions to find the records and that the questions were clearly worded.

COLLECTING THE DATA

Participants were recruited from the North Carolina State Archives and State Library. At the Archive, every visitor must register before entering the reading room.

For a period of two months, an employee at the State Archives asked everyone registering to participate in this study, following the script in Appendix G. Employees at the State Library also asked visitors to participate in the study using the same script. In addition, to increase participation rates, I recruited in person at the Archives for two days toward the end of the two-month period. Those visitors who agreed to participate were given a set of instructions and a questionnaire and asked to complete the questionnaire during their visit to the archives. The questionnaire was self-administered, but Library and Archives staff members were available to answer any procedural questions.

Each participant received one of the four versions of the instructions described above in Table 2, but all participants answered the same questions after examining the same two records. Before beginning the study, participants signed the informed consent letter in Appendix F. The signed consent letters were segregated from the completed questionnaires. Participants did not provide any demographic information about themselves, the only personally indentifying information being the signature on the consent letter. The questionnaires themselves contain no personally indentifying data.

Twenty-three researchers completed the questionnaire, including one who returned the questionnaire by mail. Because each participant was given a random set of instructions and some participants did not return their questionnaire, there were fifteen participants who completed the study using the Web Site Archives portal as their initial access point and eight who used the State Library Catalog.

ANALYZING THE DATA

Table 3 below provides an overview of the survey questions. In analyzing the data, I was looking for differences between the responses of the participants in each

condition to determine how the access method affected their experience in using Web archives. I also looked for similarities to describe researchers' overall experience using this kind of material.

Section	Questions
Sections A and C	1. Who created this document?
	2. When was this document created?
	3. What type of record is this?
	4. Who has preserved this record?
Sections B and D	5. In a sentence or two, explain what event or activity this document describes.
Sections E and F	6. I understand the document I have been asked to look at.
	7. I understand the activity that the document is representing.
	8. I understand the context of the document I have been asked to look at.
	9. I trust that this document is what it purports to be.
	10. I was able to easily locate the information I needed to answer the questions in Section A of this questionnaire.
	11. The presentation of this document effectively conveyed the information I needed to answer the questions in Section C of this questionnaire.
Section G	12. In a sentence or two, describe in your own words whether you found the task of answering the questions in sections A through D easy or difficult, and why.
	13. What else would have helped you to answer those questions?
	14. Do you have any other comments on this experience?

TABLE 3 OVERVIEW OF SURVEY QUESTIONS

For the first four sections of the survey instrument, containing the factual questions about the content and context of the documents (questions 1-5 in Table 3), I coded each participant's response as either reflecting or failing to reflect the information provided in the bibliographic record (see Appendix H for examples of accurate and

inaccurate responses), and then gave each a composite score based on the number of answers that accurately reflected the information in the bibliographic record. I calculated the mean and median composite score for each access condition to determine whether one access condition led to a higher score than the other. Although, the determination of whether an answer was appropriate or not was a subjective one, the accurate answers were essentially the same for each access condition.²² Participants received no score for unanswered questions.

The next two sections (E and F) asked the participants to respond to statements about their confidence in the documents and their understanding of them on a 4-point Likert scale, with 1 being strongly disagree and 4 being strongly agree (see questions 6-11 in Table 3).

To compare the results for the two conditions, I looked for differences in the accuracy and completeness of the answers to the factual questions and in the levels of confidence, understanding, and trust as expressed by the Likert scale responses. I performed T-tests (assuming unequal variances) to determine if the differences were statistically significant. I also calculated the correlation between accuracy in answering the factual questions and high confidence, understanding, and trust. All the calculations were performed in Excel.

²² Thus, for example, I did not consider the name of the individual from the ABC Commission's legal department who signed the minutes to be an acceptable answer to the question of who created the document, but because the signature was on the document itself and visible to all participants regardless of access condition, this determination should not affect the relative scores for participants using either the Catalog or the Wayback Machine.

For the responses to the three open-ended questions in the last section of the survey (questions 12-14 in Table 3), I looked for patterns in the responses, such as frequently-used words or common suggestions, to observe whether there were notable differences in the way participants described their experiences in their own words, depending on the access method they used.

In addition to comparing the results from each condition, I searched for general patterns that could apply to both conditions, trying to find responses that were common across conditions.

FINDINGS

COMPARISON OF RESPONSES BETWEEN CONDITIONS

Participants in the library catalog group tended to be more accurate in their answers to factual questions about the documents and their context. The most notable differences in the responses of the two groups were in identifying who had preserved the documents and the activity the revenue report documented (See Table 5 for the percentage of participants answering each question correctly). For example, library catalog participants were more likely to answer that the State Library had preserved the documents than their counterparts in the Wayback Machine group were to answer that the State Archives had preserved the documents. Overall, the scores ranged from 1 to 10 points, but the average accuracy score for the former group was 1.50 points higher than the average score for the latter group, a difference that was found to be statistically significant at a threshold of .05 (See Table 4). In addition, the standard deviation of the accuracy scores of the participants who used the library catalog was smaller than that of

the participants who used the Wayback Back, indicating the accuracy scores tended to cluster more around the mean for the first group.

	Wayback Machine			Library Catalog			
Accuracy Score	Mean	Median	Std. Deviation	Mean	Median	Std. Deviation	T-Test
		6.13	6	1.92	7.63	7	1.3

TABLE 4 MEAN AND MEDIAN ACCURACY SCORES FOR EACH CONDITION

	Wayback Machine		Library Catalog	
	Minutes	Report	Minutes	Report
Who created this document?	80%	87%	88%	88%
When was this document created?	100%	73%	100%	75%
What type of record is this?	80%	20%	88%	25%
Who has preserved this record?	13%	27%	50%	63%
In a sentence or two, explain what event or activity this document describes.	93%	40%	100%	88%

TABLE 5 PERCENTAGE OF CORRECT RESPONSES PER QUESTION FOR EACH CONDITION

Participants in the library catalog group were also more likely to answer the factual questions with answers that drew on contextual clues rather than the contents of the documents themselves. For instance, three of the eight participants in the library catalog group mentioned that it was the Deputy Commissioner's Office of the ABC Board that created the revenue report, as indicated in the title field of the bibliographic records, while only one of the fifteen participants in the other group mentioned the

Department of Commerce as a creator of the revenue report, even though the ABC Board is part of the department, as indicated by the Web Archives Portal.²³

In general, participants were confident that they understood the documents and the activity that produced them, regardless of which method they used to access them (See Table 6 below and Appendix I). Looking at the responses to each document separately, all of the median responses were 3 or above, and the mean responses were almost all 3 or above. Neither presentation method elicited many negative responses to the questions in sections E and F, implying that users were generally satisfied with their ability to accomplish the study tasks in both environments. In almost all cases, mean responses pertaining to the revenue report were less positive than mean responses pertaining to the minutes. This difference may be explained by the relative unfamiliarity of the report, and because the minutes included the word “minutes” directly within the text near the top of the document.

When the responses for the minutes and the report were combined, the median response for each question was the same for both access conditions (see Table 6 below). The mean aggregate responses were also quite similar, and none of the differences between the means for each access condition were statistically significant, though the reactions of the Wayback Machine group participants were slightly more positive when the statements related to trust, understanding the context, the effective presentation of the information, and the ability to find the information they needed.

²³ Although these answers would vary depending on access condition and thus, did not affect a participant's accuracy positively or negatively, it is interesting to note when participants gave more detailed responses to the factual questions.

	Wayback Machine			Library Catalog			T-Test
	Mean	Median	Std. Deviation	Mean	Median	Std. Deviation	
Understand Document	3.52	4	0.57	3.50	4	0.80	0.93
Understand Activity	3.57	4	0.57	3.59	4	0.67	0.89
Understand Context	3.30	3	0.75	3.09	3	0.81	0.35
Trust Document	3.37	3	0.56	3.23	3	0.61	0.40
Able to Find Information	3.10	3	0.82	2.86	3	0.71	0.28
Information Presented Effectively	3.13	3	0.82	3.00	3	0.76	0.55

TABLE 6 MEAN AND MEDIAN RESPONSES TO THE QUESTIONS IN SECTIONS E AND F (AGGREGATED) FOR EACH ACCESS CONDITION (1=STRONGLY DISAGREE, 2=DISAGREE, 3=AGREE, 4=STRONGLY AGREE)²⁴

In the responses to the open-ended questions at the end of the survey, one-third of the respondents for each condition were neutral and/or ambivalent on the question of whether it was easy or difficult to find the information they needed to answer the questions in the first sections of the questionnaire.²⁵ Most of these respondents singled out particular questions as being more difficult to answer than others or commented that it was more difficult to answer the questions about the revenue report. Users of the Wayback Machine were more likely to describe their experience as being easy, with nine of the fifteen respondents using the word easy or an equivalent to characterize their

²⁴ One participant using the State Library Catalog did not answer the questions for the revenue report.

²⁵ A neutral attitude is defined here by a statement containing a neutral adjective to describe the participant's experience, such as "ok." An ambivalent response describes one part as being difficult and another part as being easy.

experience and only one using the word difficult.²⁶ On the other hand, only one of the eight library catalog respondents described the experience of accessing and interpreting the records as easy. Three used the word difficult in response to the question. One member of this group commented in Section G that she “answered the questions but didn’t feel complete confidence” (Participant 36).

Correlation between answering the questions accurately, and trust in the documents and understanding the documents and their context was low (see Table 7 below). Participants with a high aggregate score on the factual questions in the first four sections of the survey instrument were less confident in their understanding of the documents, and much less positive about their ability to find the information they needed and the effectiveness of the presentation. Generally, participants in the library catalog group answered the questions more accurately, but were less likely to say they understood the context, trusted the documents, and were able to find information in an effective way, indicating that a user’s confidence in her ability to understand a document and its context may not always be warranted and may not be an accurate indicator of her actual understanding. Similarly, individuals who considered the tasks easy were not necessarily as accurate in their answers as those who considered the tasks difficult.

²⁶ It must be noted that participant responses contained reactions to the characteristics of the documents themselves and the survey questions in addition to the experience of using the archived records.

	Accuracy Score
Understand Document	0.41
Understand Activity	0.31
Understand Context	0.41
Trust Document	0.17
Able to Find Information	0.02
Information Effectively Conveyed	0.06

TABLE 7 CORRELATION BETWEEN THE ACCURACY SCORE AND THE QUESTIONS IN SECTIONS E AND F

GENERAL RESPONSES ACROSS CONDITIONS

Overall, trust in the documents was lower than might have been expected, given that users found the documents either through the State Library Catalog or the Government Web Site Archives portal: the mean response being closer to “agree” than “strongly agree” and the median response being “agree.” When combined with the responses to the question of who preserved the documents (overall, participants answered this question correctly only 32% of the time), however, this lower confidence makes more sense and reflects an uncertainty about Web archiving in general. Users not understanding who had preserved the document might be likely to question its authenticity and reliability.

Participants using the catalog to access the documents did not generally use the information contained in the catalog record to answer questions about the documents. The catalog record for the revenue report identified the form/genre as “report,” but only three of eight responses to the question “What type of record is this?” included the word

report.²⁷ Other responses included spreadsheet, statistical data, and chart. Six of the eight responses to this question for the minutes included the word “minutes.” This difference may be explained by the relative unfamiliarity of the report, and because the minutes included the word “minutes” directly within the text near the top of the document. None of the responses exactly duplicated the information in the catalog record relating to record type.

Only two participants, one from each access condition group noted the relationship between the ABC Commission and the Department of Commerce, although all participants were exposed to this information through the donor information in the catalog record or through the navigation of the Web Site Archives. Although this relationship was not directly addressed in the survey, it is an important part of the context for these records.

Participants in both groups expressed frustration with the lack of information on the documents themselves and in the access methods. A summary of these comments is provided in Appendix J. Frequently, responses to the questions 12 and 13 in section G included comments related to having more information available within or near the top of the documents themselves. Frequent examples are date/time of creation (six respondents) and author/creator (five respondents). These comments may have been prompted, at least in part, by the participants being specifically asked questions about document creators and dates of creation.

²⁷ This question caused some confusion with the survey testers. Originally the question asked about the “document” type, but some testers thought that document type might refer to the file format. After pilot testing and before I recruited any actual participants, I changed the word to “record” to clarify that the question referred to the content and form of the document rather than the file format. A few participants in the study did seem to still be confused by the question despite the change in terminology.

One unexpected finding was that a few of participants using the documents in the Digital Archive seemed confused about who had created the electronic records. In response to the question of who created the minutes, one participant answered that the North Carolina Alcoholic Beverage Control Commission created the document, but that she did not know who created the digital version (Participant 6). Because they had not seen these documents in their original context on the Web, these researchers may have thought that the State Library (or OCLC, as two responses suggested) had received paper copies of these documents and digitized them.

DISCUSSION

The results of this study provide some indication that there is value to archival description in Web archives. Study participants with exposure to the archivist-created access aids were better able to answer questions about the context and content of the records, even with the admittedly minimal series description provided in the MARS catalog. The complete archived Web site did not, in fact, contextualize the record well enough for the researchers to quickly understand the record, although perhaps if they had had more time to spend on the site as a whole, they would have demonstrated a more complete understanding of the two study documents.

If archivists plan to rely on the “technocentric” approach to Web archiving, they will need to work to influence the practices of Web site creators to better contextualize the records they contain from their original creation and placement on the Internet. Although, it may not always be possible for archivists to influence the practices of Webmasters, when there is an organizational or legal mandate to collect these records, archivists may have the ability to effect change in this area.

In comparison to the findings of Spindler and Pearce-Moses in their study of users' understanding of archival catalog records when they were only exposed to the catalog records and not the documents represented by those records, the participants in this study did not generally rely on the statement of responsibility in the catalog record to answer the question about the record creator, suggesting that the information in catalog records may be forgotten once the user begins to examine the archival materials (Spindler and Pearce-Moses 1993, 336). Many of the comments participants made in the last section of the questionnaire would indicate that archivists and archival systems creators could better serve researchers if they displayed metadata on the same screen as the archived materials. Participants relied heavily on the information contained in the document and on the screen display rather than any contextualizing information they encountered before they viewed the records.²⁸ Researchers using paper records are able to keep a finding aid or other access aid next to them as they work with primary sources for quick reference. They might also benefit from having this ability in an electronic environment as well. One respondent, for example, suggested a "citation heading" to be displayed along with the document.

In the Wayback Machine group, a participant commented that, "[i]n general, both documents assumed that the reader knew what he or she was looking at" (Participant 38). With respect to the primary users of the documents, this assumption would be natural, and probably correct, because the records were an integral part of their normal business

²⁸ It is telling that three participants in the library catalog condition group mentioned OCLC in their answers when asked who preserved the records. The only indication of OCLC's involvement in the preservation activity is the small OCLC icon at the top of the documents and the name in the URL for the document.

activities. Secondary users, seeing these records in an environment that could easily be confused with the original use environment, do not necessarily have the background knowledge that the primary users have, but are in a similar use environment.

The confusion evidenced by some participants illustrates a point Margaret Hedstrom made about the usability and accessibility of electronic archives. She wrote, “[u]ntil most members of society feel as comfortable with electronic evidence as they do with traditional forms of documentation, archivists will have a responsibility to help users evaluate, understand, and interpret new documentary forms” (Hedstrom 1998, 15). Many participants in the study did not understand the nature of Web archiving and did not understand how the documents in the study were preserved and by whom. One comment from section G is particularly revelatory in this context:

I was not sure as to whether I should take extra time to find these answers or not. But I suppose if I had to look elsewhere on the web site for names & positions, I would be looking for info that would all be different in the year 2150. I suppose the issue then becomes, - "How thorough will the web archives be?" - will I be able to find out who had the jobs of creating & preserving documents 143 years earlier? (Participant 13).

These results indicate that archivists and third-party providers of Web archiving services need to do a better job of indicating to users how the original record has been preserved in order to maintain its authenticity and reliability. Even if the form of the record itself remains unchanged and the document genre remains recognizable, the format of the record will still present challenges to the user as she tries to evaluate and interpret it in its new archival environment.

Archivists have long debated whether and how to provide subject access to historical records. If, in fact, users are able to fully understand the evidential aspects of a

record even while it remains in its traditional Web site context, preserving and presenting records in this way could go a long way toward providing subject access, without a lot of extra work on the part of the archivist processing the records, especially if search engines for Web archives improve. Participant comments tended to confirm that more complete description and presentation in the record's original environment could significantly reduce the amount of arrangement and description required once the records are accessioned into an archival environment and provide more effective subject-based access. Nonetheless, the difference between how well the participants answered the factual questions in this study implies that archivists may still need to focus at least some of their attention on providing contextual information to users.

LIMITATIONS

Web sites are a much more complicated form of information than this study might seem to imply. This study involved two traditional types of records available on the Internet in PDFs. Although taking these minutes and report as an example of records on the Internet facilitates the study of user reactions because they do fit into a more traditional notion of records and how they should be presented to users, they are by no means representative of all the types of records available online. Other types of digital records on the Internet may include databases, interactive tours and programs, the Web sites themselves, and even the records of how users interacted with the Web sites. The documents used in this study could be considered "non-Web archives" that have been "created and organized independently from the Web," which means that capture and processing techniques and procedures for this type of record could have very different characteristics from the techniques and procedures used to archive "Web-served"

documents (Masanès 2006, 32-37). This study does not address access to all these types of records, and in truth, the results may not apply well to those that are radically different from the paper-based records on which most archival processing and descriptive practices were based.

This study also does not address the wide variety of ways in which even these two documents could be presented to users. The focus in this experiment has been on the two most common ways of providing access to archived Internet resources in the United States. However, as noted above, other methods do currently exist; and with the way these resources are stored and made available over the Internet through the use of unique URLs, archivists and librarians can create a variety of user interfaces for providing access. For instance, an archivist could create an EAD finding aid for a Web site or a series of related Web sites and embed links to the archived Web sites on the finding aid, bypassing the Wayback Machine interface and the library catalog altogether. Or she could create a Web page with a table listing all the URLs for a particular collection and providing a brief description and/or some metadata for each. Another option would be to create a database to house the URLs of the archived Web sites and their related metadata that would allow for more effective searching when collections begin to grow very large. This option would be very similar to the traditional cataloging process and fully-implemented EAD, but the format and types of metadata allowed would not be limited as they are in a library catalog or EAD finding aid. The archivist could use Dublin Core metadata and include preservation metadata that does not fit into the MARC model if she chose to do so, for instance.

Another option yet would be to use a search engine to index the archived Web pages and allow searching within a collection, within a repository, or across repositories. Developing search engines to be used with the Web site archives environment is a high priority for those involved in the design and creation of Web archiving tools.

How archivists chose to present archived Web sites and other Web records and make them discoverable and accessible in the future could follow any of the models described here, any combination of these models, and other models not described and not even considered yet. Unfortunately, this study could only test two current practices in this area, but more research will be needed to test the viability of other models as they develop and are more widely used.

Another issue that is not widely discussed in the literature dealing with access to digital records relates to preservation metadata. Currently when archivists discuss preservation metadata, they are discussing it in the context of the actual preservation needs of the records and documents themselves. Preservation metadata in this context helps the archivist to ensure that the files are being appropriately cared for and to assess their continued reliability and authenticity in the archive. What is absent from this discussion is how best to explain to users what steps have been taken to preserve the files. Users will need this information in order to make their own judgments concerning the reliability and authenticity of the materials they are using, but it is far from clear how much information and what kind of information they will need. At the moment, this metadata is usually invisible to the average user of digital records, but as time passes and archivists are required to perform more preservation activity to keep these records accessible, the changes made to the files themselves and the systems in which they are

and were used will be important to users. This study did not attempt to address the problem of the presentation of preservation metadata.

The needs of users today are not necessarily the needs of users tomorrow. Researchers of the future may use Web site archives in ways that are completely different from the way researchers of the present use paper archives. If archivists know little about how people actually use discovery and access tools, they know even less about how they will use them. Adrian Brown suggests that users, having become accustomed to full-text search tools, will primarily want these tools for accessing Web archives as well, but current search engines do not even begin to address issues of context, authenticity, and reliability in the ways that they index and rate relevance to a given query (Brown 2006). At the moment, they do not effectively accommodate searches for information for a specific time period, either past or present (Bar-Ilan 2006). Search engines may indeed supplant finding aids as the access aid of choice for archival research, but search engines too have much room for improvement if they are to be optimized for this type of research.

Finally, the small number of researchers recruited for this study are not necessarily representative of all potential Web archives users. The small sample size for this study precludes generalizing the results to the larger population of archives users and making detailed statistical inferences. While the researchers at the North Carolina State Archives are certainly representative of some archives users, they do not represent all types of current archives users. Furthermore, just as changes will occur in how archives are used, the people using these materials will likely change. In fact, archivists are actively promoting these changes and see the ability to reach a new audience as one of

the key advantages of networking technologies, but new audiences may require new ways of doing business.

CONCLUSION

Archivists did not conduct many formal user tests, if they conducted any at all, when the finding aid in its most common form was first being adopted. There have been some user tests of EAD finding aids and MARC AMC records in recent years, which are a step in the right direction. However, providing access to archived Web sites, and archived electronic records in general, may well require a completely new kind of access aid. If current trends are any indication, archivists will not be creating traditional finding aids for born-digital collections. At a time when old descriptive practices are being reevaluated, it is an opportune moment for archivists to conduct more tests into user needs and practices.

It is difficult for archivists to truly understand archival access aids in their role of providing intellectual access for users. It is far easier to test discoverability and usability with various access aids. Nonetheless, archivists should not ignore what Luciana Duranti has described as the most important role of access aids today – that of explaining the contextual and provenancial background of records. The findings from this research into description for Web archiving do not conclusively point to one particular representation method as being more effective than the other at fulfilling this essential function of archival description, but do provide some insight into the strengths and weaknesses of each. Moreover, this study does not attempt to consider the contextual requirements for

the archived Web as a “hypermedia collectively edited or a global cultural artifact” (Masanès 2006, 1).

Web archives, at least in the near future (though perhaps not by the year 2150), will need to provide researchers with more information about the nature of records captured from the Internet. To do so may require displaying more technical and preservation metadata to users. In addition, as suggested with respect to recordkeeping systems, archivists might find that the best use of their time lies in working with records creators (and Webmasters) to provide better contextualization at the point of creation, either automatically or manually, rather than undertaking traditional description of Web archives after accessioning them into an archival repository.

Participants’ generally positive responses to the records accessed through the Wayback Machine, as well as evidence of their ability to understand these records suggest that “technocentric” methods for processing and description do have a place in the archives, but participants’ understanding of context was clearly improved by manually-created bibliographic records. A large number of the participants suggested that the access system displays could be improved. Ultimately, the more important question may be how to present descriptive information, rather than what descriptive information to present.

The question of how researchers use access aids to understand records after they have found them deserves further study. More research into how the users of primary source materials use finding aids and other contextual information to help them interpret records and documents is essential if archivists want to create effective access aids for both paper and digital records.

More than one study participant made unsolicited comments on the utility and importance of having these records preserved and available online. One noted, “It is good that such material is being preserved on the Internet” (Participant 17), while another wrote, “If this sort of document availability (online access) becomes a reality, it will be a tremendous benefit to researchers” (Participant 7). As this “sort of document availability” does become a reality, archivists still have an important role to play as mediators in ensuring that it will, in fact be, of tremendous benefit.

BIBLIOGRAPHY

- Altman, Burt, and John Nemmers. "The Usability of On-line Archival Resources: The Polaris Project Finding Aid." *The American Archivist* 64, no. 1 (2001): 121-131.
- Anderson, Ian. "Are You Being Served: Historians and the Search for Primary Sources." *Archivaria* 58 (2004): 81-129.
- Bailey, Catherine. "Archival Theory and Electronic Records." *Archivaria* 29 (1989): 180-196.
- Bar-Ilan, Judit. "False Web Memories: A Case Study on Finding Information about Andrei Broder." *First Monday* 11, no. 9 (2006): http://firstmonday.org/issues/issue11_9/barilan/index.html, February 27, 2007.
- Brown, Adrian. *Archiving Websites: A Guide for Information Management Professionals*. London: Facet, 2006.
- Chalfant, Don, and Kathy Jordan. "Archival Processing of Electronic Records." Presentation at Best Practices Exchange: Libraries and Archives in the Digital Era. Chandler, AZ, May 2-4, 2007.
- Coyle, Karen, and Diane Hillmann. "Resource Description and Access (RDA), Cataloging Rules for the 20th Century." *D-Lib Magazine* 13, no. 1/2 (2007): <http://www.dlib.org/dlib/january07/coyle/01coyle.html>.
- Dow, Elizabeth, David Chesnutt, William Underwood, Helen Tibbo, Mary-Jo Kline, and Charlene Bickford. "The Burlington Agenda: Research Issues in Intellectual Access to Electronically Published Historical Documents." *The American Archivist* 64, no. 2 (2001): 292-307.
- Dryden, Jean. "Archival Description of Electronic Records: An Examination of Current Practices." *Archivaria* 40 (1995): 99-108.
- Duff, Wendy. "Will Metadata Replace Archival Description: A Commentary." *Archivaria* 39 (1995): 33-38.
- Duff, Wendy, and Catherine Johnson. "Where Is the List with All the Names? Information-Seeking Behavior of Genealogists." *The American Archivist* 66, no. 1 (2003): 79-95.
- Duff, Wendy, and Penka Stoyanova. "Transforming the Crazy Quilt: Archival Displays from a Users' Point of View." *Archivaria* 45 (1998): 44-79.

- Duff, Wendy, Barbara Craig, and Joan Cherry. "Finding and Using Archival Resources: A Cross-Canada Survey of Historians Studying Canadian History." *Archivaria* 58 (2004): 51-80.
- Duranti, Lucianna. "Origin and Development of the Concept of Archival Description." *Archivaria* 35 (1993): 47-53.
- Eubank, Kelly, and Kristen Martin. "The North Carolina State Government Web Site Archives." *Hypertext*, 2007: forthcoming.
- Forde, Helen. "Access and Preservation in the 21st Century: What Has Changed?" *Journal of the Society of Archivists* 26, no. 2 (2005): 193-200.
- Gilliland-Swetland, Anne. "Popularizing the Find Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups." *Journal of Internet Cataloging* 4, no. 3/4 (2001): 199-225.
- Greene, Mark, and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist* 68, no. 2 (2005): 208-263.
- Grotke, Abbie, and Janice Ruth. "Selecting and Managing Content Captured from the Web: Expanding Curatorial Expertise and Skills in Building Library of Congress Web Archives." Presentation at DigCCurr 2007: An International Symposium in Digital Curation. Chapel Hill, NC, April 18-20, 2007.
- Guenther, Rebecca, and Leslie Myrick. "Archiving Web Sites for Preservation and Access: MODS, METS and MINERVA." *Journal of Archival Organization* 4, no. 1/2 (2006): 141-166.
- Guercio, Maria. "Principles, Methods, and Instruments for the Creation, Preservation, and Use of Archival Records in the Digital Environment." *The American Archivist* 64, no. 2 (2001): 238-269.
- Hallgrimsson, Thorsteinn. "Access and Finding Aids." In *Web Archiving*, edited by Julien Masanès, 131-151. Berlin: Springer, 2006.
- Hamburger, Susan. "How Researchers Search for Manuscript and Archival Collections." *Journal of Archival Organization* 2, no. 1/2 (2004): 79-97.
- Haworth, Kent. "Archival Description: Content and Context in Search of Structure." *Journal of Internet Cataloging* 4, no. 3/4 (2001): 7-26.
- Hedstrom, Margaret. "Building Record-Keeping Systems: Archivists Are Not Alone on the Wild Frontier." *Archivaria* 44 (1997): 44-71.
- Hedstrom, Margaret. "Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible." *Archivaria* 36 (1993): 53-63.
- Hedstrom, Margaret. "How Do Archivists Made Electronic Archives Usable and Accessible?" *Archives and Manuscripts* 26, no. 1 (1998): 6-22.
- Hedstrom, Margaret, Christopher Lee, Judith Olson, and Clifford Lampe. "'The Old Version Flickers More': Digital Preservation from the User's Perspective." *The American Archivist* 69, no. 1 (2006): 159-187.

- Hensen, Steven. "Archival Cataloging and the Internet: The Implications and Impact of EAD." *Journal of Internet Cataloging* 4, no. 3/4 (2001): 75-95.
- Hill, Amanda. "Serving the Invisible Researcher: Meeting the Needs of Online Researchers." *Journal of the Society of Archivists* 25, no. 2 (2004): 139-148.
- Hostetter, Christina. "Online Finding Aids: Are They Practical." *Journal of Archival Organization* 2, no. 1/2 (2004): 117-145.
- Hutchinson, Tim. "Strategies for Searching Online Finding Aids: A Retrieval Experiment." *Archivaria* 44 (1997): 72-101.
- Jatowt, Adam, Yukiko Kawai, Satoshi Nakamura, Yutaka Kidawara, and Katsumi Tanaka. "Journey to the Past: Proposal of a Framework for Past Web Browser." In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, 135-144. New York: ACM Press, 2006.
- Jimerson, Randall. "Archival Description and Finding Aids." *OCLC Systems & Services* 18, no. 3 (2002): 125-129.
- Kahle, Brewster. "Editor's Interview: The Internet Archive, an Interview with Brewster Kahle." *RLG DigiNews* 6, no. 3 (2002): <http://www.rlg.org/preserv/diginews/diginews6-3.html#interview>.
- Kelly, Brian. "WebWatch: Surfing Historical UK University Web Sites." *Ariadne* 35 (2003): <http://www.ariadne.ac.uk/issue35/web-watch/>.
- Kimpton, Michelle, and Jeff Ubois. "Year-by-Year: From an Archive of the Internet to an Archive on the Internet." In *Web Archiving*, edited by Julien Masanès, 201-212. Berlin: Springer, 2006.
- Lack, Rosalie. "The Importance of User-Centered Design: Exploring Findings and Methods." *Journal of Archival Organization* 4, no. 1/2 (2006): 69-86.
- Lee, Hur-Li. "What is a Collection?" *Journal of the American Society of Information Science and Technology* 51, no. 12 (2000): 1106-1113.
- MacNeil, Heather. "Metadata Strategies and Archival Description: Comparing Apples to Oranges." *Archivaria* 39 (1995): 22-32.
- Masanès, Julien. "Web Archiving: Issues and Methods." In *Web Archiving*, edited by Julien Masanès, 1-53. Berlin: Springer, 2006.
- Meissner, Dennis. "First Things First: Reengineering Finding Aids for Implementation of EAD." *The American Archivist* 60, no. 4 (1997): 372-387.
- Menne-Haritz, Angelika. "Access - The Reformulation of an Archival Paradigm." *Archival Science* 1 (2001): 57-82.
- Murray, Kathleen, and Mark Phillips. "Collaboration, Best Practices, and Collection Development for Born-Digital and Digitized Materials." Presentation at DigCCurr 2007: An International Symposium in Digital Curation. Chapel Hill, NC, April 18-20, 2007.
- OCLC. "Web Archives Workbench BETA, User Guide (version 2.0)." January 2007.

- Pearce-Moses, Richard, and Joanne Kaczmarek. "An Arizona Model for Preservation and Access of Web Documents." *DttP: Documents to the People* 33, no. 1 (2005): 17-24.
- Proffitt, Merrilee. "How and Why of User Studies: RLG's RedLightGreen as a Case Study." *Journal of Archival Organization* 4, no. 1/2 (2006): 87-110.
- Prom, Christopher. "Reengineering Archival Access through the OAI Protocols." *Library Hi Tech* 21, no. 2 (2003): 199-209.
- Rauber, Andreas, Andreas Aschenbrenner, and Oliver Witvoet. "Austrian Online Archive Processing: Analyzing Archives of the World Wide Web." *Research and Advances Technology for Digital Technology : 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002. Proceedings*. Berlin: Springer, 2002. 16-31.
- Roe, Kathleen. *Arranging & Describing Archives & Manuscripts*. Chicago: The Society of American Archivists, 2005.
- Roth, James. "Serving Up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids." *The American Archivist* 64, no. 2 (2001): 214-237.
- Spindler, Robert, and Richard Pearce-Moses. "Does AMC Mean "Archives Made Confusing"? Patron Understanding of USMARC AMC Catalog Records." *The American Archivist* 56, no. 2 (1993): 330-341.
- Sweet, Meg, and David Thomas. "Archives Described at Collection Level." *D-Lib Magazine* 6, no. 9 (2000):
<http://www.dlib.org/dlib/september00/sweet/09sweet.html>.
- Tibbo, Helen. "Primarily History in America: How U. S. Historians Search for Primary Materials at the Dawn of the Digital Age." *The American Archivist* 66, no. 1 (2003): 9-50.
- Wallace, David. "Managing the Present: Metadata as Archival Description." *Archivaria* 39 (1995): 11-21.
- Yakel, Beth. "Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users?" *Journal of Archival Organization* 2, no. 1/2 (2004): 63-77.
- Yakel, Beth, Seth Shaw, and Polly Reynolds. "Creating the Next Generation of Archival Finding Aids." *D-Lib Magazine* 13, no. 5/6 (2007):
<http://www.dlib.org/dlib/may07/yakel/05yakel.html>.

APPENDIX A PARTICIPANT INSTRUCTIONS (WEB SITE ARCHIVES)

For the next fifteen to twenty minutes, imagine that you are an individual living in the year 2150 and you are interested in finding out about government oversight of alcoholic beverage sales in North Carolina around the turn of the 21st century.

You know that around this time, the North Carolina State Government began putting a lot of its records on the Internet, so you are going to use material that the government archived in this format as part of your research. You find out from the staff at the State Archives that this material is available through the North Carolina State Government Web Site Archives.

1. Go to the Web Site Archives at <http://www.ah.dcr.state.nc.us/archives/Webarchives/index.html>.
2. Choose the “Browse by Agency” option from the navigation bar on the left side of the page.
3. Select “Commerce, Dept. of” from the list of agencies.
4. Select the link for the Alcoholic Beverage Control Commission.
5. You see a list of dates and you decide to start with February 2, 2007. Clicking on this date takes you to the Alcoholic Beverage Control Commission homepage.
6. Click on the “Reports” button on the right side of the page.
7. From the list of reports, choose “Legal reports,” and from that list, select “Commission Meeting Minutes.”
8. Select the item entitled “November 2006,” which will bring up a PDF document.

After examining the document, answer the questions on the next page. You may use any information you are able to find in the document and on the rest of the Web site to complete the questionnaire.

Please do not leave the Wayback Machine for the “live Web.” This means you should only use Web sites found within in the North Carolina State Government Web Site Archive. You will be able to tell whether a given Webpage is part of the archive by checking the Web site address for the phrase “Wayback.archive-it.org.” If the phrase appears in the address, you are still in the Wayback Machine.

Section A – Answer the following questions about this document as completely as possible.

Who created this document?

When was this document created?

What type of record is this?

Who has preserved this record?

Section B – Answer the following question about this document.

In a sentence or two, explain what event or activity this document describes.

Please continue onto the next page and follow the instructions there.

1. Go back to the North Carolina State Government Web Site Archives at <http://www.ah.dcr.state.nc.us/archives/Webarchives/index.html>.
2. Choose the “Browse by Agency” option from the navigation bar on the left side of the page.
3. Select “Commerce, Dept. of” from the list of agencies.
4. Select the link for the Alcoholic Beverage Control Commission.
5. Click on the link for February 2, 2007. Clicking on this date takes you to the Alcoholic Beverage Control Commission homepage.
6. From the navigation bar on the left side of the screen, click on the link for “Local ABC Boards.”
7. From the navigation bar on the left side of the screen, click on the link for “Revenues.”
8. Click on the link for “July 1, 2005-June 30, 2006” under the title “Annual Revenue by ABC Board,” which will bring up a PDF document.

After examining the document, answer the questions on the next page. You may use any information you are able to find in the document and on the rest of the Web site to complete the questionnaire.

Please do not leave the Wayback Machine for the “live Web.” This means you should only use Web sites found within in the North Carolina State Government Web Site Archive. You will be able to tell whether a given Webpage is part of the archive by checking the Web site address for the phrase “Wayback.archive-it.org.” If the phrase appears in the address, you are still in the Wayback Machine.

Section C – *Answer the following questions about this document as completely as possible.*

Who created this document?

When was this document created?

What type of record is this?

Who has preserved this record?

Section D – *Answer the following question about this document.*

In a sentence or two, explain what event or activity this document describes.

Section E – *Read the following statements and indicate whether they describe your experience answering the questions above for the first document.*

I understand the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I understand the activity that the document is representing.

Strongly agree Agree Disagree Strongly disagree

I understand the context of the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I trust that this document is what it purports to be.

Strongly agree Agree Disagree Strongly disagree

I was able to easily locate the information I needed to answer the questions in Section A of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

The presentation of this document effectively conveyed the information I needed to answer the questions in Section A of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

Section F – Read the following statements and indicate whether they describe your experience answering the questions above for the second document.

I understand the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I understand the activity that the document is representing.

Strongly agree Agree Disagree Strongly disagree

I understand the context of the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I trust that this document is what it purports to be.

Strongly agree Agree Disagree Strongly disagree

I was able to easily locate the information I needed to answer the questions in Section C of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

The presentation of this document effectively conveyed the information I needed to answer the questions in Section C of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

Section G – Please respond to the following questions.

In a sentence or two, describe in your own words whether you found the task of answering the questions in sections A through D easy or difficult, and why.

What else would have helped you to answer those questions?

Do you have any other comments on this experience?

Thank you for your participation!

APPENDIX B PARTICIPANT INSTRUCTIONS (STATE LIBRARY CATALOG)

For the next fifteen to twenty minutes, imagine that you are an individual living in the year 2150 and you are interested in finding out about government oversight of alcoholic beverage sales in North Carolina around the turn of the 21st century.

You know that around this time, the North Carolina State Government began putting a lot of its records on the Internet, so you are going to use material that the government archived in electronic formats as part of your research. You find out from the staff at the State Archives that this material is available through the State Library Catalog.

1. Go to the State Library Catalog at <http://go.dcr.state.nc.us/cgi-bin/PWebrecon.cgi?DB=local&PAGE=First>.
2. Select the Author browse option on the basic search page of the library catalog and type “Alcoholic Control” into the Find This field. Click on the Begin Search button.
3. From the list of authors returned, select “Alcoholic Control, Board of.”
4. From the list of titles returned, select the first title (Chairman’s Office: Detailed Minutes), and click on the title link.
5. Click on the “Detailed Record” button.
6. On the catalog record displayed, click on the top-most link in the record, which will bring up a PDF.

After examining the document, answer the questions on the next page. You may use any information you are able to find in the document and in the catalog to complete the questionnaire.

Please do not leave the Digital Archive and catalog environment. This means you should only use the Web site with the black OCLC Digital Archive header at the top of the page or the library catalog.

Section A – *Answer the following questions about this document as completely as possible.*

Who created this document?

When was this document created?

What type of record is this?

Who has preserved this record?

Section B – *Answer the following question about this document.*

In a sentence or two, explain what event or activity this document describes.

Please continue onto the next page and follow the instructions there.

1. Go back to the State Library Catalog search page at <http://go.dcr.state.nc.us/cgi-bin/PWebrecon.cgi?DB=local&PAGE=First>.
2. Select the Author browse option on the basic search page of the library catalog and type “Alcoholic Control” into the Find This field. Click on the Begin Search button.
3. From the list of authors returned, select “Alcoholic Control, Board of.”
4. From the list of titles returned, select the second title (Deputy Commissioner's Office: Public Revenue from Alcoholic Beverages File), and click on the title link.
5. Click on the “Detailed Record” button.
6. On the catalog record displayed, click on the top-most link in the record, which will bring up a PDF.

After examining the document, answer the questions on the next page. You may use any information you are able to find in the document and in the catalog to complete the questionnaire.

Please do not leave the Digital Archive and catalog environment. This means you should only use the Web site with the black OCLC Digital Archive header at the top of the page or the library catalog.

Section C – Answer the following questions about this document as completely as possible.

Who created this document?

When was this document created?

What type of record is this?

Who has preserved this record?

Section D – Answer the following question about this document.

In a sentence or two, explain what event or activity this document describes.

Section E – Read the following statements and indicate whether they describe your experience answering the questions above for the first document.

I understand the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I understand the activity that the document is representing.

Strongly agree Agree Disagree Strongly disagree

I understand the context of the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I trust that this document is what it purports to be.

Strongly agree Agree Disagree Strongly disagree

I was able to easily locate the information I needed to answer the questions in Section A of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

The presentation of this document effectively conveyed the information I needed to answer the questions in Section A of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

Section F – *Read the following statements and indicate whether they describe your experience answering the questions above for the second document.*

I understand the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I understand the activity that the document is representing.

Strongly agree Agree Disagree Strongly disagree

I understand the context of the document I have been asked to look at.

Strongly agree Agree Disagree Strongly disagree

I trust that this document is what it purports to be.

Strongly agree Agree Disagree Strongly disagree

I was able to easily locate the information I needed to answer the questions in Section C of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

The presentation of this document effectively conveyed the information I needed to answer the questions in Section C of this questionnaire.

Strongly agree Agree Disagree Strongly disagree

Section G – *Please respond to the following questions.*

In a sentence or two, describe in your own words whether you found the task of answering the questions in sections A through D easy or difficult, and why.

What else would have helped you to answer those questions?

Do you have any other comments on this experience?

Thank you for your participation!

APPENDIX C MARS RECORDS FOR ABC COMMISSION RECORD SERIES

NORTH CAROLINA
State Archives
Manuscript and Archives Reference System (MARS)

Full Record Display [Help](#) | [Exit](#)

Quick Navigation: [Browse/Search](#) | [Find Resources](#)

[view relationships](#) [Previous](#) | [Next](#)

Deputy Commissioner's Office: Public Revenue From Alcoholic Beverages File

Title:	Deputy Commissioner's Office: Public Revenue From Alcoholic Beverages File
Provenance:	Class: State Records [Collection] Group: Alcoholic Beverage Control Commission Record Group
Years:	1958-1970
Creator:	Alcoholic Control, Board of Alcoholic Control, Board of
Location:	ORC.101-H.7
MARS Id:	38.11 (Series) RS Number: 320
Genres / Forms:	Reports
Quantity:	1 Records Center Box(es)
Arrangement:	Unprocessed.
Scope / Contents:	Reports received from counties detailing revenues collected and methods of revenue distribution. %7ETransferred Dates Floor Row/Shelf Spaces Volume to Archives Transf. 1/4/82 1958-1970 ORC 101-H 7 1 box 6-20-88
Index Terms:	Subjects: Revenue Alcoholic Beverages Taxation Corporate Names: Alcoholic Control, Board of
Source / Donor:	Commerce, Dept. of

FIGURE 6 MARS RECORD FOR THE PUBLIC REVENUE FROM ALCOHOLIC BEVERAGE FILE IN THE ALCOHOLIC BEVERAGE CONTROL COMMISSION RECORD GROUP

NORTH CAROLINA
State Archives
Manuscript and Archives Reference System (MARS)

Full Record Display [Help](#) | [Exit](#)

Quick Navigation: [Browse/Search](#) | [Find Resources](#)

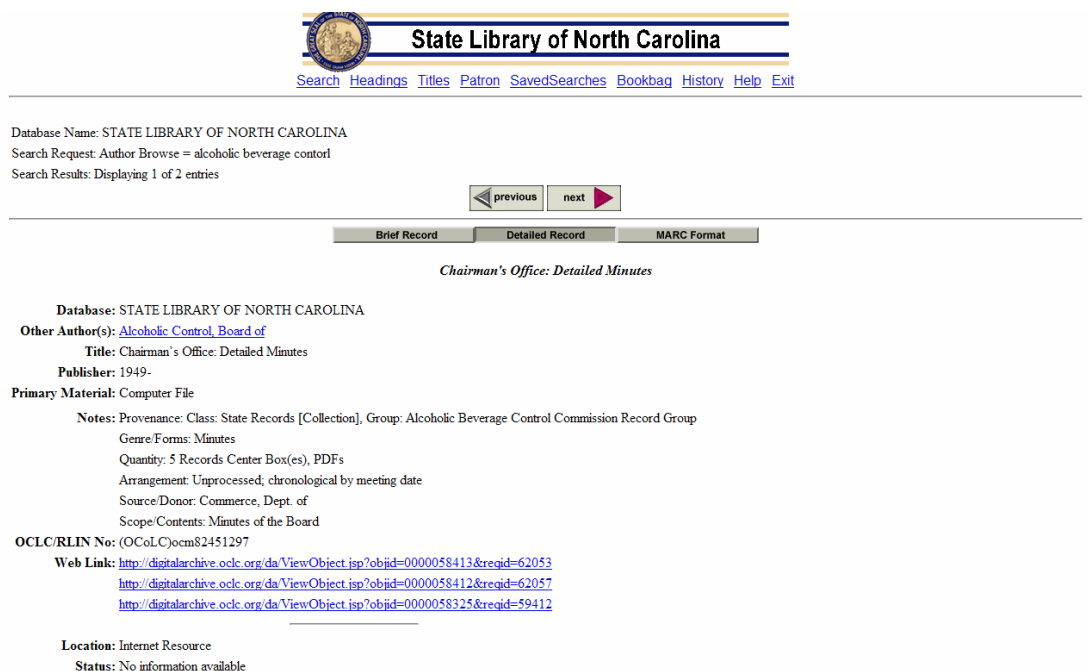
[view relationships](#) [Previous](#) | [Next](#)


Chairman's Office: Detailed Minutes

Title:	Chairman's Office: Detailed Minutes
Provenance:	Class: State Records [Collection] Group: Alcoholic Beverage Control Commission Record Group
Years:	1949-1958
Creator:	Alcoholic Control, Board of Alcoholic Control, Board of
Location:	ORC.105-D.27
MARS Id:	38.12 (Series) RS Number: 319
Genres / Forms:	Minutes
Quantity:	1 Records Center Box(es)
Arrangement:	Unprocessed; chronological by meeting date.
Scope / Contents:	Minutes of the board. %7ETransferred Dates Floor Row/Shelf Spaces Volume to Archives Transf. 9/16/80 1949-1958 ORC 105-D 27 1 box 6-20-88
Index Terms:	Corporate Names: Alcoholic Control, Board of
Source / Donor:	Commerce, Dept. of
Other Copies:	Minutes (1951-1952) (Box 1) (SRC.Basement.Range 12)

FIGURE 7 MARS RECORD FOR THE DETAILED MINUTES SERIES IN THE ALCOHOLIC BEVERAGE CONTROL COMMISSION RECORD GROUP

APPENDIX D STATE LIBRARY CATALOG RECORDS FOR ABC COMMISSION RECORD SERIES



 **State Library of North Carolina**

[Search](#) [Headings](#) [Titles](#) [Patron](#) [SavedSearches](#) [Bookbag](#) [History](#) [Help](#) [Exit](#)

Database Name: STATE LIBRARY OF NORTH CAROLINA
 Search Request: Author Browse = alcoholic beverage control
 Search Results: Displaying 1 of 2 entries

◀ previous next ▶

Brief Record **Detailed Record** MARC Format

Chairman's Office: Detailed Minutes

Database: STATE LIBRARY OF NORTH CAROLINA
Other Author(s): [Alcoholic Control, Board of](#)
Title: Chairman's Office: Detailed Minutes
Publisher: 1949-
Primary Material: Computer File


Notes: Provenance: Class: State Records [Collection], Group: Alcoholic Beverage Control Commission Record Group
 Genre/Forms: Minutes
 Quantity: 5 Records Center Box(es), PDFs
 Arrangement: Unprocessed; chronological by meeting date
 Source/Donor: Commerce, Dept. of
 Scope/Contents: Minutes of the Board

OCLC/RLIN No: (OCoLC)ocm82451297

Web Link: <http://digitalarchive.ocl.org/da/ViewObject.jsp?objid=0000058413&reqid=62053>
<http://digitalarchive.ocl.org/da/ViewObject.jsp?objid=0000058412&reqid=62057>
<http://digitalarchive.ocl.org/da/ViewObject.jsp?objid=0000058325&reqid=59412>

Location: Internet Resource
Status: No information available

FIGURE 8 LIBRARY CATALOG RECORD FOR THE PUBLIC REVENUE FROM ALCOHOLIC BEVERAGE FILE IN THE ALCOHOLIC BEVERAGE CONTROL COMMISSION RECORD GROUP



 **State Library of North Carolina**

[Search](#) [Headings](#) [Titles](#) [Patron](#) [SavedSearches](#) [Bookbag](#) [History](#) [Help](#) [Exit](#)

Database Name: STATE LIBRARY OF NORTH CAROLINA
 Search Request: Author Browse = alcoholic beverage control
 Search Results: Displaying 2 of 2 entries

◀ previous next ▶

Brief Record **Detailed Record** MARC Format

Deputy Commissioner's Office: Public Revenue From Alcoholic Beverages File

Database: STATE LIBRARY OF NORTH CAROLINA
Other Author(s): [Alcoholic Control, Board of](#)
Title: Deputy Commissioner's Office: Public Revenue From Alcoholic Beverages File
Publisher: 1958-
Primary Material: Computer File

Notes: Provenance: Class: State Records [Collection], Group: Alcoholic Beverage Control Commission Record Group.
 Genres/Forms: Reports
 Quantity: 7 Records Center Box(es), PDFs
 Arrangement: Unprocessed
 Source/Donor: Commerce, Dept. of
 Scope/Contents: Summaries of reports received from counties detailing revenues collected and methods of revenue distribution.

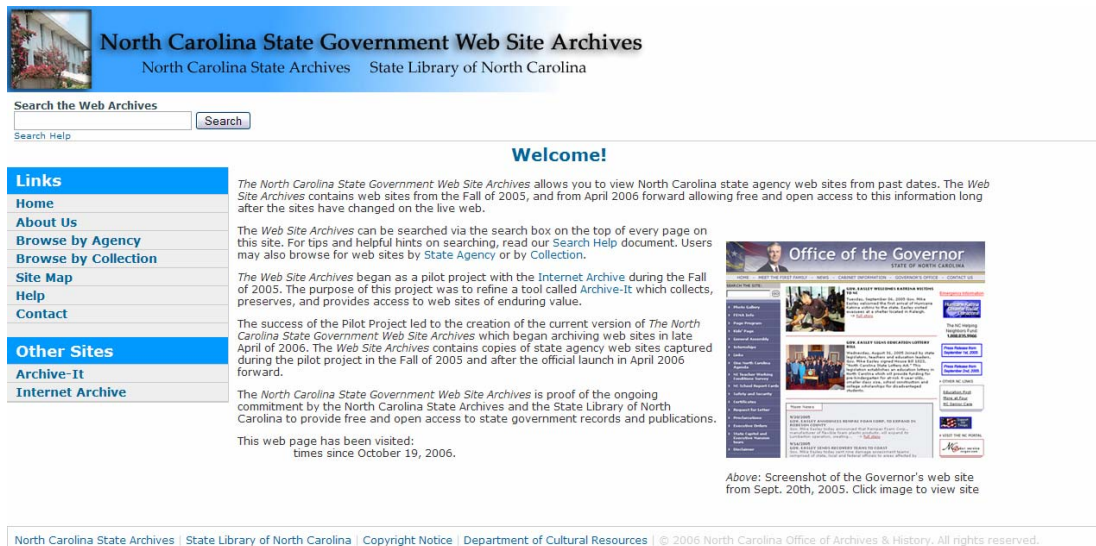
OCLC/RLIN No: (OCoLC)ocm82452091

Web Link: <http://digitalarchive.ocl.org/da/ViewObject.jsp?objid=0000058411&reqid=62046>
<http://digitalarchive.ocl.org/da/ViewObject.jsp?objid=0000058324&reqid=59410>

Location: Internet Resource
Status: No information available

FIGURE 9 LIBRARY CATALOG RECORD FOR THE DETAILED MINUTES SERIES IN THE ALCOHOLIC BEVERAGE CONTROL COMMISSION RECORD GROUP

APPENDIX E NORTH CAROLINA STATE GOVERNMENT ARCHIVES PORTAL AND WAYBACK MACHINE PAGES FOR THE ABC COMMISSION WEB SITE



North Carolina State Government Web Site Archives
North Carolina State Archives State Library of North Carolina

Search the Web Archives

Search Help

Welcome!

Links
[Home](#)
[About Us](#)
[Browse by Agency](#)
[Browse by Collection](#)
[Site Map](#)
[Help](#)
[Contact](#)

Other Sites
[Archive-It](#)
[Internet Archive](#)

The North Carolina State Government Web Site Archives allows you to view North Carolina state agency web sites from past dates. The Web Site Archives contains web sites from the Fall of 2005, and from April 2006 forward allowing free and open access to this information long after the sites have changed on the live web.


The Web Site Archives can be searched via the search box on the top of every page on this site. For tips and helpful hints on searching, read our Search Help document. Users may also browse for web sites by State Agency or by Collection.

The Web Site Archives began as a pilot project with the Internet Archive during the Fall of 2005. The purpose of this project was to refine a tool called Archive-It which collects, preserves, and provides access to web sites of enduring value.

The success of the Pilot Project led to the creation of the current version of The North Carolina State Government Web Site Archives which began archiving web sites in late April of 2006. The Web Site Archives contains copies of state agency web sites captured during the pilot project in the Fall of 2005 and after the official launch in April 2006 forward.

The North Carolina State Government Web Site Archives is proof of the ongoing commitment by the North Carolina State Archives and the State Library of North Carolina to provide free and open access to state government records and publications.

This web page has been visited:
times since October 19, 2006.



Above: Screenshot of the Governor's web site from Sept. 20th, 2005. Click image to view site

North Carolina State Archives | State Library of North Carolina | Copyright Notice | Department of Cultural Resources | © 2006 North Carolina Office of Archives & History. All rights reserved.

FIGURE 10 NORTH CAROLINA STATE GOVERNMENT WEB SITE ARCHIVES PORTAL



North Carolina State Government Web Site Archives
North Carolina State Archives State Library of North Carolina

Search the Web Archives

Search Help

Browse Archives by State Agency

Links
[Home](#)
[About Us](#)
[Browse by Agency](#)
[Browse by Collection](#)
[Site Map](#)
[Help](#)
[Contact](#)

Other Sites
[Archive-It](#)
[Internet Archive](#)

Browse sites in the Web Site Archives by the state agency that created them. Web sites listed under each agency are not an attempt to provide a comprehensive list of all governmental agencies nor do they reflect the organizational hierarchy. Some departments include almost all subordinate units and programs under a single web site (e.g. The Department of Revenue) while others have separate web sites for multiple units and programs (e.g. The Department of Health and Human Services). Only those units or programs that have separate web sites are listed under each department.

- Administration, Dept. of
- Administrative Hearings, Office of
- Administrative Office of the Courts
- Agriculture and Consumer Services, Dept. of
- Crime Control and Public Safety, Dept. of
- Commerce, Dept. of
- Community Colleges System Office
- Correction, Dept. of
- Commissions, Councils, Foundations, and Trusts
- Cultural Resources, Dept. of
- Environment and Natural Resources, Dept. of
- General Assembly
- Governor, Office of the
- Health and Human Services, Dept. of
- Insurance, Dept. of

FIGURE 11 NORTH CAROLINA STATE GOVERNMENT WEB SITE ARCHIVES PORTAL, BROWSE BY AGENCY

North Carolina State Government Web Site Archives
 North Carolina State Archives State Library of North Carolina

Search the Web Archives

Search Help

Department of Commerce

Click on the links below to browse all versions of the web sites captured in the *North Carolina State Government Web Site Archives*.
 Not all web sites in the Web Site Archives are currently available on the live web. As such, there may not be versions captured after a certain date. We recommend that you try a search if you can't find what you're looking for.

Links	<p>Department of Commerce http://www.nccommerce.com</p> <p>Alcoholic Beverage Control Commission http://www.ncabc.com</p> <p>Board of Science and Technology http://www.ncscienceandtechnology.com</p>
Other Sites	<p>Business Recruitment http://www.investnc.com</p> <p>Division of Community Assistance http://www.ncdca.org</p> <p>Division of Employment and Training http://www.ncdet.com</p> <p>Economic Development Information System http://cmedis.commerce.state.nc.us</p> <p>Employment Security Commission http://www.ncesc.com</p> <p>Film Office http://www.ncfilm.com</p> <p>Industrial Commission http://www.comp.state.nc.us</p> <p>International Trade Division http://www.exportnc.com</p> <p>JobLink Career Center http://www.ncjoblink.com</p> <p>North Carolina Japan Office http://www.ncjapan.com</p>

FIGURE 12 NORTH CAROLINA STATE GOVERNMENT WEB SITE ARCHIVES PORTAL, DEPARTMENT OF COMMERCE SITES

ARCHIVE-IT **North Carolina State Government Web Site Archive Web Archive (North Carolina State Archives)** INTERNET ARCHIVE **Wayback Machine**

Enter Web Address: All

Searched for <http://www.ncabc.com> 23 Results

[Look up URL](#) in general Internet Archive web collection

Note some duplicates are not shown. [See all](#).
 * denotes when site was updated.

Search Results for Jan 01, 2005 - Jun 14, 2007		
2005	2006	2007
14 pages	5 pages	2 pages
Sep 20, 2005 * Sep 21, 2005 * Sep 22, 2005 * Sep 23, 2005 * Sep 24, 2005 * Sep 25, 2005 * Sep 27, 2005 * Sep 28, 2005 * Sep 29, 2005 * Oct 06, 2005 * Oct 13, 2005 * Oct 20, 2005 * Oct 27, 2005 * Nov 11, 2005 *	Apr 28, 2006 * May 28, 2006 * Jun 01, 2006 * Jul 05, 2006 * Nov 02, 2006 *	Feb 02, 2007 * Apr 27, 2007 *

[Home](#) | [Copyright © 2005, Internet Archive](#) | [Terms of Use](#) | [Privacy Policy](#)

FIGURE 13 WAYBACK MACHINE, NORTH CAROLINA ALCOHOLIC BEVERAGE CONTROL COMMISSION



FIGURE 14 NORTH CAROLINA ALCOHOLIC BEVERAGE CONTROL COMMISSION HOMEPAGE, WAYBACK MACHINE

APPENDIX F STUDY CONSENT LETTER

What are some general things you should know about research studies?

You are being asked to take part in a research study. To join the study is voluntary.

You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies. However, your participation is extremely valuable in helping to understand the experience of archives users and in improving that experience in the future.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?

The purpose of this research study is to learn about how people make sense of archived electronic records available through the Internet. Archivists are beginning to preserve important records and documents available on the Internet, and these materials will require them to consider new ways of making archival materials available to the public. The results of this study will help archivists present these materials more effectively to users.

How many people will take part in this study?

If you decide to be in this study, you will be one of forty people in this research study.

How long will your part in this study last?

Participation in this study will require approximately 15-20 minutes. You should complete the study here at the Archives, using the computers in the reading room. Once you turn in your questionnaire, you will have completed your portion of the study.

What will happen if you take part in the study?

If you decide to take part in this study, you will be looking at two Web sites and answering questions about what you find on them and about your experience answering the questions.

There will be two groups of study participants and participants will be assigned to one group or the other based on the order in which they agree to participate. Each group will look at the same two documents and answer the same questions. However the way in which the documents are presented will be different for each group.

What are the possible benefits from being in this study?

Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study. However, this research is intended to benefit the entire community of archives users and professionals.

What are the possible risks or discomforts involved from being in this study?

There are no foreseeable risks or discomforts if you agree to participate in this study.

How will your privacy be protected?

No personal information will be collected for this study and your identity will not be linked to your study responses in any way. Participants will not be identified in any report or publication about this study.

Will you receive anything for being in this study?

You will not receive any compensation for taking part in this study.

What if you have questions about this study?

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researchers listed on the first page of this form.

What if you have questions about your rights as a research participant?

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Participant's Agreement:

I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

Signature of Research Participant

Date

Printed Name of Research Participant

APPENDIX G STUDY RECRUITMENT SCRIPT

Instructions for the recruiter are in italics.

Please read the following to every researcher registering at the North Carolina State Archives:

A graduate student studying archives at the School of Information and Library Science at UNC-Chapel Hill is conducting a study on archiving Web sites and would like to have researchers here at the State Archives participate.

Are you at least eighteen years old? (may be omitted if researcher appears to be at least thirty years old)

*If the researcher says **no**, thank them for their time.*

*If the researcher says **yes**, continue as follows:*

This study will only take about 15-20 minutes and can be completed on the computers in the search room. It involves looking at a couple of Web sites and completing a questionnaire based on your experience. Your participation is entirely voluntary, but your assistance will be greatly appreciated and will contribute to a better understanding of how to make this material available to researchers in the future.

Would you be willing to participate?

*If the researcher says **no**, thank them for their time.*

*If the researcher says **yes**, hand them the top packet of papers on the pile and continue as follows:*

Please read the consent letter you will find on the top of the packet and sign it before you begin the study. When you are done, please return the consent letter and survey here at the front desk.

APPENDIX H CODING OF RESPONSES TO QUESTIONS IN SECTIONS A THROUGH D

Questions	Correct Answers (Wayback Machine)	Correct Answers (Library Catalog)	Incorrect Answers (Examples from both conditions)
	Minutes		
Who created this document?	Alcoholic Beverage Control Commission	Alcoholic Beverage Control Commission	Dottie Taylor or other individual name
When was this document created?	Nov. 9, 2006 (or following the Nov. 9, 2006 meeting)	Nov. 9, 2006 (or following the Nov. 9, 2006 meeting)	N/A
What type of record is this?	Minutes	Minutes	PDF, Summary of a meeting
Who has preserved this record?	State Archives	State Library	ABC Commission, State of NC, OCLC
In a sentence or two, explain what event or activity this document describes.	Any description of the meeting and/or topics discussed	Any description of the meeting and/or topics discussed	N/A
	Revenue Report		
Who created this document?	Alcoholic Beverage Control Commission	Alcoholic Beverage Control Commission	County and City ABC Boards
When was this document created?	July 2005-June 2006, after June 2006, or annually	July 2005-June 2006, after June 2006, or annually	2000s
What type of record is this?	Report	Report	Spreadsheet, Chart, P+L, State Record, Statistical Data
Who has preserved this record?	State Archives	State Library	ABC Board, OCLC

Questions	Correct Answers (Wayback Machine)	Correct Answers (Library Catalog)	Incorrect Answers (Examples from both conditions)
In a sentence or two, explain what event or activity this document describes.	Revenue from liquor sales in North Carolina	Revenue from liquor sales in North Carolina	Liquor sales in NC, Taxes on alcoholic beverages

TABLE 8 EXAMPLES OF CORRECT AND INCORRECT ANSWERS TO THE FACTUAL QUESTIONS

**APPENDIX I SUMMARY OF REPOSSES TO QUESTIONS IN
SECTIONS E AND F FOR EACH DOCUMENT**

Minutes							
	Wayback Machine			Library Catalog			
	Mean	Median	Std. Deviation	Mean	Median	Std. Deviation	T-Test
Understand Document	3.67	4	0.49	3.63	4	0.52	0.85
Understand Activity	3.67	4	0.49	3.63	4	0.52	0.85
Understand Context	3.47	3	0.52	3.50	3.5	0.54	0.89
Trust Document	3.33	3	0.62	3.38	3	0.52	0.87
Able to Find Information	3.40	3	0.63	3.13	3	0.64	0.34
Information Presented Effectively	3.47	4	0.64	3.00	3	0.76	0.16

**TABLE 9 SUMMARY OF THE REPOSSES TO THE QUESTIONS IN SECTIONS E AN F FOR THE MINUTES
(1=STRONGLY DISAGREE, 2=DISAGREE, 3=AGREE, 4=STRONGLY AGREE)**

Revenue Report							
	Wayback Machine			Library Catalog			
	Mean	Median	Std. Deviation	Mean	Median	Std. Deviation	T-Test
Understand Document	3.36	3	0.63	3.43	4	0.98	0.86
Understand Activity	3.47	4	0.64	3.57	4	0.79	0.76
Understand Context	3.13	3	0.92	2.86	3	0.90	0.52
Trust Document	3.40	3	0.51	3.14	3	0.69	0.40
Able to Find Information	2.79	2.5	0.89	2.71	3	0.76	0.85
Information Presented Effectively	2.80	3	0.86	3.00	3	0.82	0.61

TABLE 10 SUMMARY OF THE REPOSSES TO THE QUESTIONS IN SECTIONS E AN F FOR THE MINUTES (1=STRONGLY DISAGREE, 2=DISAGREE, 3=AGREE, 4=STRONGLY AGREE)

APPENDIX J SUMMARY OF COMMENTS FROM SECTION G

These comments indicate that many participants felt like the information contained in the documents and their context did not allow them to quickly find the information they needed to answer the survey questions.²⁹

Questions	Responses
Condition: Wayback Machine	
In a sentence or two, describe in your own words whether you found the task of answering the questions in sections A through D easy or difficult, and why.	I found answering the questions about the second document more difficult - the date produced was not available on the document and I would have assumed this is just part of a much longer report. There were no clues on the previous page. (33)
	It is possible that I could have searched the ABC web site to locate the answers, but as I had already spent 20 minutes on the survey, I decided that would not be proper. (13)
What else would have helped you to answer those questions?	More info on the site. (45)
	Names and positions of who created and preserved document, and time it was created included on document itself. (13)
	Material regarding date of creation placed prominently at the top of the page next to the agency involved. (1)
	If the documents were time-stamped. (3)
Do you have any other comments on this experience?	References of who created the doc and why (19)
	As an older person, I find navigating about on the computer more difficult than probably most people do. The information regarding the website (agency, dates etc.) would have been easier for me to see if there had been a box near the top that included that information. (1)
	Can bring up follow-on questions about person who created these documents and how they went thru the process. (36)
Condition: Library Catalog	

²⁹ With the exception of a few minor spelling and punctuation corrections, these responses are verbatim from the survey instrument. The participant's survey number is indicated in parentheses after the quotations.

Questions	Responses
<p>In a sentence or two, describe in your own words whether you found the task of answering the questions in sections A through D easy or difficult, and why.</p>	<p>I'm an attorney, so I'm used to ferreting out information! However, neither document was a model of clarity in terms of answering the questions. In general, both documents assumed that the reader knew what he or she was looking at. That might have been true with the minutes, but probably not with the graph. (38)</p> <hr/> <p>Difficult - Not familiar with the agencies or context. No summary information about author, source was in the document itself. I answered the questions but didn't feel complete confidence. (36)</p> <hr/> <p>Answering the questions in sections A through D was difficult because the chart with the information was poorly labeled. (48)</p>
<p>What else would have helped you to answer those questions?</p>	<p>A statement at the beginning/top of each document which clearly listed the answers to the questions you asked: creator of document, date created, purpose of document. (38)</p> <hr/> <p>I think it is easier to understand documents with more context and background. (2)</p> <hr/> <p>Had to back up through screens to view and remember the author, date...Having a "citation heading" would have helped. (28)</p>

TABLE 11 SUMMARY OF COMMENTS FROM SECTION G