

**A White Paper on the Status of
North Carolina Digital State Government Information**

**North Carolina State Government Information:
*Realities and Possibilities***

November 2003

Prepared by:
Kristin Martin, Digital State Documents Librarian
Jan Reagan, Head, Documents Branch

State Library of North Carolina
North Carolina Department of Cultural Resources

Prepared as part of *The Access to State Government Information Initiative*
Funded through a Library Services and Technology Act Statewide Leadership Grant

**North Carolina State Government Information:
*Realities and Possibilities***

Table of Contents

Executive Summary	1
Introduction.....	1
I. Access to State Government Information Initiative.....	3
II. The Status of North Carolina State Government Information	5
Publication vs. Record.....	5
New Formats and Presentations.....	7
Born Digital vs. Digitized Information.....	9
III. Changes in State Government Publishing Practices	11
Publication Formats.....	11
Reasons for Change	13
IV. New Challenges to Permanent Public Access.....	14
Current Information: Challenges to Access.....	14
Historical Information: Challenges to Access and Preservation.....	17
Technical Barriers to Permanent Public Access.....	18
Non-Technical Barriers to Permanent Public Access.....	21
Challenges within North Carolina State Government	23
V. Addressing the Challenges: Potential Solutions	25
Digital Preservation Approaches.....	25
Preservation Initiatives: National Level	26
Preservation Initiatives: State Level	27
Access Initiatives: Federal and State.....	27
Conclusion: A Call for Action.....	29
Bibliography	32
Appendix A: Content and Purpose of State Information	35
Appendix B: Survey Tool.....	36
Appendix C: Searches for State Government Information	48
Appendix D: Glossary of Terms.....	51

Executive Summary

Access to public information is vital to any democracy. Government information keeps citizens informed and enables them to participate in their government and hold it accountable. State government information is no exception and, for this reason, the State of North Carolina must ensure permanent public access to all state information regardless of format.

Currently, the State Library and the State Archives and Records Section in the North Carolina Department of Cultural Resources are legally mandated by General Statute to preserve printed state publications and government records for permanent public access. Traditional definitions of “publication” and “record” in state government provide a clear distinction between the two for purposes of public access, collection, management, and preservation of information. New technologies that enable state agencies to produce and disseminate information directly through the Internet now allow previously “unpublished” information to be included on webpages for public access, thus blurring the distinction between “publication” and “record.” For this reason, current definitions of “publication” and “record” may need to be reconsidered in order for the state to effectively manage government information in digital formats.

In many cases, printed counterparts no longer exist for digital state information. State publications and records existing solely in digital formats – born digital information – pose challenges to the traditional systems within state government designed to collect, manage, and preserve information for easy public access and long-term use. In order to address these challenges, the State Library obtained an LSTA Statewide Leadership Grant in 2002 and embarked on a three-year project to research digital information issues, gain a better understanding of current publishing practices in state agencies, and develop solutions for managing state information in digital formats. Led and staffed by the State Library, the Access to State Government Information Initiative is a collaborative effort with the State Data Center, State Archives and Records, and a core Work Group of primary stakeholders consisting of information providers (state agency staff), information facilitators (librarians, archivists, records managers, technology specialists, state data specialists), and end-users.

In 2002, project staff commenced the research phase of the *Initiative*. Staff conducted literature and web searches for information regarding the collection, management, and preservation of digital information as well as “best practices” in other countries, states, and the

federal government. As the agency responsible for ensuring public access to state publications, the State Library focused its investigation on the changes taking place in the production and dissemination of publications, rather than all state government information. Staff examined a sample of 10 executive branch agency websites, reviewed nearly 2,000 agency publications in print and digital formats, and conducted interviews with 76 state agency personnel representing 27 agencies to obtain this information.

The trend in state government is definitely to produce fewer printed publications and more digital information via the Internet. Improvements in technology and state budget cuts are driving this transition from print to digital and, as a result, born digital publications now comprise approximately half of all publications produced by state agencies. For the most part, agencies acknowledge the advantages of digital information and agree that, even if printing budgets improve, print will no longer be the preferred format for publications.

Research shows that although there is a significant amount of state information on the Web, finding it can be challenging. Standard search engines such as Google, have limited indexing capabilities and may not “crawl” or search for state information in databases and dynamically generated pages in the “invisible” or “deep” Web. Also, constant design changes and updating of information on websites often leads to broken links and frustration for users. Making state information easier to find through standard search engines and customized access tools must be a goal for North Carolina state government.

The most troubling concern about digital government publications is whether historical information will be available in the future. Criteria for removing publications from agency websites range from content to server space considerations to terms of political office. No state agencies have policies in place that address the issue of long-term preservation and, as a result, public access to digital publications taken off the Internet is problematic. Possible solutions for preserving and accessing historical digital publications offline may be some type of centralized repository, or series of distributed repositories, searchable from a central location.

There are also a number of technical and non-technical barriers to permanent public access that must be overcome. Digital information has a short life span for three technological reasons: media degradation, hardware obsolescence, and software obsolescence. Digital storage media degrades more quickly than paper and can quickly become unreadable. Software and hardware platforms, necessary to translate digital information into a human-readable format, become

obsolete as new technology replaces older programs and storage devices. Beyond the technical issues, other problems hinder efforts to preserve digital information. State librarians and archivists have the interest and responsibility to preserve information, however, they lack the resources needed to create and implement a digital preservation strategy. The principles of librarianship and archival theory that guide these professionals in managing materials in tangible formats also enable them to tackle the difficult issues of the digital world. The lack of experience managing digital information makes it difficult, however, to convince policy makers and funding sources to allocate resources toward the effort.

Unfortunately, there are currently no “best practices” to emulate and no definitive solutions to implement. The complexities of digital information and the volatile nature of the technology that generates it complicate the realization of solutions. Research is underway, however, in the United States and around the world, to determine methods for providing permanent public access to digital information. Different approaches to preservation, including migration and emulation of digital information, are being considered. The Library of Congress is leading the National Digital Information Infrastructure and Preservation Program, a project of nearly \$100 million to address the challenges of digital preservation. At the state level, about three-fifths of the states have begun to address the need for permanent public access to digital government information. Additionally, states and the federal government are addressing the need for improved access to current government information through enhanced indexing and searching tools such as GILS, the Government Information Locator Service.

The amount of digital state government information increases daily and the probability of it disappearing is high. For this reason, North Carolina state government must act now to develop a digital information strategy to prevent further loss of valuable publications and records. Stakeholders must work together to start laying the groundwork for sustaining ongoing efforts to realize workable solutions for ensuring the existence, availability, and usability of government information over time, regardless of format. As the Library of Congress states, “action is needed now, not some time in the future; and everyone—from creators to custodians—must contribute to the solution and learn to operate fluently in a world of constant and unpredictable change.”

Introduction

State government information is valued and widely used by the constituents and citizens of North Carolina who depend on accurate and reliable current and historical information and data. A variety of users including students, educators, businesses, historians, farmers, legislators, local government, journalists, and others seek information produced by state government and expect it to be available for their use. State government produces information that touches upon nearly every aspect of life in the state of North Carolina. Research indicates the scope of state information is broad and the content and purpose varied, much like the array of constituencies served by this information.

Current state government information is necessary for the proper functioning of North Carolina society. It is needed to participate in society (e.g., obtain a driver's license); properly conduct business; provide services; and comply with state law. Historical information collected by both the State Library and the State Archives and Records Section in the North Carolina Department of Cultural Resources has enduring value and significance as a vital source of evidence of government activities and decisions over time. It remains an important source of corporate memory for the government and the people of North Carolina (Appendix A).

The State Library and the State Archives and Records are legally mandated to manage and preserve printed state publications and government records, respectively, to ensure permanent public access to this information. The Government Records Branch in the Archives and Records Section sets retention schedules and collects and manages public records, transferring those records of enduring value to the State Archives for permanent public access. The State Library fulfills its duty through the North Carolina State Documents Depository System, which provides for the collection, cataloging, and distribution of government publications to libraries across the state and the State Library collection. The State Library serves as “the official, complete, and permanent depository for all State publications.”¹

Over the last six years, new technologies have enabled government agencies to publish and distribute digital information directly via the Internet. More recently, budget cuts have forced some state agencies to eliminate printed documents altogether. The result is a new breed of

¹ *North Carolina General Statutes*, Chapter 125-11.5-11.13: Libraries, Article 1A: “State Depository Library System, 2002.

information—born digital information—that poses challenges to the traditional systems within state government designed to collect, manage, and preserve information for easy public access and long-term use.

The State Library and the State Archives and Records Section have begun to address the challenges of digital state government information by developing access tools for finding state information on the Web and developing guidelines for indexing databases and maintaining and preserving records of web-based activities. State agency participation and compliance with these guidelines and recommendations, however, has been minimal and digital state information is in jeopardy of being lost to the public. State government must concentrate its efforts and resources towards realizing solutions for accessing and managing state information in all formats, including born digital.

I. Access to State Government Information Initiative

In 2002, the State Library obtained an LSTA Statewide Leadership Grant for year one of the *Access to State Government Information Initiative*, to research issues and develop solutions for managing state information in digital formats.² Led and staffed by the State Library, the *Initiative* is a collaborative effort with the State Data Center, State Archives and Records, and a core Work Group of primary stakeholders consisting of information providers (state agencies), information facilitators (librarians, archivists, records managers, data specialists), and end-users. Project staff, collaborators, and stakeholders will work together to assess options and reach consensus on state government’s approach to identifying, collecting, preserving, and providing continued access to state government information in all formats.

Phase I of the *Initiative* was devoted to research, the results of which provide the foundation for this paper and the *Initiative* itself. A noted decline in the number of printed state publications received in the North Carolina State Publications Clearinghouse over the last six years (Table 1), and the increase in information available on state agency webpages provided the impetus for the State Library to propose this research phase.³

Table 1: Titles Cataloged and Added to the State Documents Collection – 1997 vs. 2003

<u><i>New Titles (printed documents) Cataloged and Distributed through the North Carolina State Documents Depository System</i></u>		<u><i>New Monographs and Serials Issues (printed documents) Added to the North Carolina State Documents Collection in the State Library (permanent depository collection)</i></u>	
1997	819	1997	8,345
2002/03	413	2002/03	4,264
% change (est. 6 yrs.)	50 % fewer titles	% change (est. 6 yrs.)	51 % fewer titles

² Library Services and Technology Act, 1997 provides federal funds through Statewide Leadership Grants, to support state level, change-oriented initiatives that have broad, statewide impact.

³ The North Carolina State Publications Clearinghouse, established in the State Library by G.S. 125-11, serves as the conduit between state agencies and state depository libraries for the receipt, processing, and distribution of state publications.

Project staff conducted literature and web searches for information regarding the collection, management, and preservation of digital information as well as “best practices” in other countries, states, and the federal government. Staff conducted interviews with 76 state agency personnel representing 27 agencies to obtain information regarding publishing practices and trends within state agencies. (Appendix B: Survey Tool). Agency personnel were able to provide only “best guess” estimates to quantitative survey questions, as no authoritative data exists regarding publishing practices and methods. In order to verify and fortify the data estimates collected in the interviews, project staff examined a sample of 10 executive branch agency websites and reviewed nearly 2,000 agency publications in print and digital formats. Using data from the various research components, staff were able to approximate percentages for publishing practices, formats, methods and the like. It is important to note, however, these approximate numbers suffice only to indicate trends and do not represent definitive data.

As mentioned earlier, the State Library is the agency responsible for ensuring public access to state publications. For this reason, the Phase I project research focused on the changes taking place in the production and dissemination of state publications, rather than all state government information. Staff met and worked with State Archives and Records staff to gain perspective on the current status of state records and insight into agency perspectives on the management of digital information.

During the course of the research, it became apparent that not only are printed publications shifting to digital formats, but also new formats and presentation options brought forth by the Internet are blurring the traditional distinctions between publications and records. New technologies allow state agencies to disseminate and provide access to information in ways that were not considered practical or even possible in printed and other tangible formats.

II. The Status of North Carolina State Government Information

Publication vs. Record

Traditional definitions of “publication” and “record” in state government provide a clear distinction between the two for purposes of public access, collection, management, and preservation of this information. New digital presentations, however, challenge the State’s traditional programs for collecting, managing, and preserving state publications and records. Because of this, current definitions of “publication” and “record” may need to be reconsidered in order for the state to effectively manage government information in digital formats.

Traditionally, librarians manage government publications produced for public dissemination, while archivists and records managers handle “unpublished” government records. Digital publishing now allows agencies to easily include previously “unpublished” information on webpages for public access, thus blurring the distinction between “publication” and “record.” In some cases, it is unclear which agency, the State Library or the State Archives and Records, is now responsible for ensuring the continued existence of this information.

Widely used and accepted dictionary definitions of “publication” and “record” refer to both as printed or written works or materials. For example, *The American Heritage Dictionary of the English Language* defines a “publication” as “an issue of printed material offered for sale or distribution” and a “record” as “an account, as of information or facts, set down especially in writing as a means of preserving knowledge.”⁴

Definitions for “publication” in the North Carolina General Statutes limit the scope to printed materials only. North Carolina General Statute 125-11 defines a “state publication” as “any document prepared by a State agency or private organization, consultant, or research firm, under contract with or under the supervision of a State agency.” The same statute defines a “document” as “any printed document including any report, directory, statistical compendium, bibliography, map, regulation, newsletter, pamphlet, brochure, periodical, bulletin, compilation, or register, regardless of whether the printed document is in paper, film, tape, disk, or any other format.”⁵ All such publications are currently collected and managed by the State Library for public access.

⁴ *American Heritage Dictionary of the English Language*, 4th ed. (Boston: Houghton Mifflin, 2000).

⁵ *North Carolina General Statutes*, Chapter 125-11.5-11.13.

There are currently no provisions or definitions in G.S. 125-11 for collecting and managing documents not published in printed formats.

The definition provided for “public record,” in North Carolina General Statute 132-1, is somewhat broader: “all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance or in connection with the transaction of official business by any agency.”⁶ State Archives and Records specifically addresses the issue of digital records in *The North Carolina Guidelines for Managing Public Records Produced by Information Technology Systems*, published in April 2000. The Guidelines define “electronic records” as records “requiring the aid of electronic technology to make the record readable or otherwise comprehensible by ordinary human sensory capabilities.”⁷ While State Archives and Records have been researching methods to collect and preserve digital records, most of their recommendations have been issued in the form of guidelines. Retention schedules still require printed copies for records of enduring value.

Webpages perhaps present the greatest challenge to the traditional definition of “publication” and the state’s systems for preserving and ensuring access to government information. Webpages are technically “published” when broadcast over the Internet, but often contain information traditionally considered to be a record. In addition, webpages often lack clearly defined boundaries, making it difficult to collect, manage, and preserve this digital information for long-term public access. For instance, does each page constitute a separate document? Or, is each file or image a separate document? What is the relationship between the various pages? Should PDF files within a site be treated as publications, while the rest of the website is treated as a record?

Web-enabled databases also pose challenges to traditional means for collecting and preserving state information. Currently, State Archives and Records considers stand-alone databases produced and maintained by state agencies to be public records. Web-enabled databases, however, allow users to extract specific information from agency databases according to selected criteria and produce reports for downloading or printing. Users can also seamlessly

⁶ *North Carolina General Statutes*, Chapter 132: Public Records, 2002.

⁷ Division of Archives and History, “North Carolina Guidelines for Managing Public Records Produced by Information Technology Systems,” (Raleigh: Department of Cultural Resources, 2000), 1, <http://www.ah.dcr.state.nc.us/e-records/manrecrd/manrecrd.htm> (accessed November 12, 2003).

link to websites outside the pages created by the database. Data within the database is now merely one of the components of a more complex presentation of government information that has yet to be defined as a publication or a record for collection, management, and preservation purposes.

Librarians, archivists, and records managers in state government must work together to reconsider and redefine what constitutes a publication vs. a record for purposes of retention and preservation. Fortunately, a decision about how to view or define digital government publications and records need not be determined prior to investigating possible solutions for managing digital information. This issue, however, is one that should be addressed by the *Access to State Government Information Initiative*.

New Formats and Presentations

Today, state government produces and disseminates information in tangible and intangible formats. Tangible formats have physical presence and form, such as printed documents, printed and written records, photographs, videotapes, and CD-ROMs. Intangible formats, on the other hand, have no physical presence and form and may include information originating and existing in cyberspace, web-enabled databases, digital documents, and e-mail messages. Libraries and archives currently maintain systems for collecting, managing, and preserving state government information in tangible formats; however, information in intangible formats poses challenges to these systems. For this reason, information formats must now be considered when describing and defining publications and records for purposes of collection, management, and preservation.

The presentation of information as a discrete entity or an integrated part of a greater whole is also a critical component in describing and defining state publications and records. Information can exist as a discrete document or record that has meaning and value on its own, such as a book, journal issue, a digital document in PDF format, or a database that contains and relies upon a defined set of data. Discrete presentations may also be self-sustaining, clearly defined parts of something larger, such as a monograph on a specific topic within a general topic series (e.g., the Department of Labor individual farm safety pamphlets published as part of the series, *On the Farm: Health and Safety Tips*). Traditional library cataloging, which focuses on describing the object at hand, works well for discrete presentations, since items can be easily identified and described without extensive reference to external information sources.

Information can also exist as an integrated part of an intricately connected whole, composed of related parts which derive additional meaning and value from the whole. Integrated information is “linked” to other resources and derives meaning and value from its relationship with these other resources. It usually does not stand alone and may require artificial boundaries to be imposed upon it. Integrated presentations may include webpages with numerous links or databases accessible through web-based search interfaces. In a tangible format, individual letters in a series of correspondence that rely upon other letters in the series for meaning qualify as integrated documents. Archivists, when describing such collections, rely upon the principles of original order, provenance, and series-level description. An understanding of the order, the collector, and the information contained within the collection as a whole provides more value than individual descriptions of each letter or item in the collection.

Integrated presentations of information in intangible formats, particularly webpages, add another level of complexity for description and preservation because of the difficulty in determining boundaries and the interweaving of discrete objects within an integrated presentation. Peter Lyman, professor at the School of Information Management & Systems, University of California, Berkeley, aptly describes the current situation in his observation that “the librarian tends to look at the content of a webpage as the object to be described and preserved. The computer scientist tends to look at the Web as a technology for linking information—a system of relationships (hence the name ‘Web’).”⁸ For this reason, librarians, archivists, and records managers must work together to assess integrated information and determine how it can be incorporated into the state’s solutions for preservation and continued access to government information.

⁸ Peter Lyman, "Archiving the World Wide Web," in *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, ed. Amy Friedlander (Washington, D.C.: Council on Library and Information Resources and the Library of Congress, 2002), 47, <http://www.clir.org/pubs/abstract/pub106abst.html> (accessed November 13, 2003).

Table 2: Formats and Presentations Examples

	Tangible	Intangible
Discrete	<ul style="list-style-type: none"> • Printed reports/books • Videos • CD-ROMs • Printed journals 	<ul style="list-style-type: none"> • PDF reports/books • Stand-alone databases • PDF journals
Integrated	<ul style="list-style-type: none"> • Printed letters in a series of correspondence • Volumes of an encyclopedia 	<ul style="list-style-type: none"> • Hypertext webpages • Web-enabled databases • Email

Born Digital vs. Digitized Information

Born digital information is information published and distributed in intangible formats via the Internet for which there is no tangible counterpart (e.g., printed paper documents, microform, or videotape). Born digital state information may include reports, magazines, newsletters, webpages, and datasets published, disseminated, and accessed only through the Internet. Digitized information, on the other hand, is information converted from analog (i.e., printed, tangible formats) to digital formats for dissemination and access via the Internet. For instance, libraries and archives are now scanning older documents, newspapers, and manuscripts to create digital versions available through the Internet. Unlike born digital materials, digitized materials have tangible counterparts that can be managed and preserved. The digital copy is often created to enhance access.⁹

The distinction between “born digital” and “digitized” becomes blurred when, for example, historic footage or older information becomes part of a new digital project.¹⁰ Does the “digital project” constitute a new and distinct “born digital publication” to be collected, managed, and preserved as it is? Or, do the parts of the project, each collected, managed, and preserved in their original tangible format, suffice for long-term access? An even finer distinction can be drawn

⁹ This paper normally uses the term “digital” to describe information in binary code that requires a computer (or other machine) to translate into a human-readable form. “Electronic” encompasses all information that requires technological intervention to be read, which includes some analog formats, such as audiotape. We are specifically concerned with digital information, so the term “digital” is preferred over “electronic,” though the two are close enough in meaning that they may be used interchangeably.

¹⁰ *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, ed. Amy Friedlander (Washington, D.C.: Council on Library and Information Resources and the Library of Congress, 2002), 2, <http://www.clir.org/pubs/abstract/pub106abst.html> (accessed November 13, 2003).

between intangible, born-digital information disseminated solely through the Internet and born-digital information disseminated through the Internet that is also transferred to a tangible, analog format (e.g., print) for distribution. Should the state be concerned with collecting and preserving the born digital version as well as the printed version of this information?

The State Library and the State Archives and Records are particularly interested in finding ways to identify, collect, and manage born digital information for which there is no tangible, printed counterpart. Determining which information on agency websites is born digital and which is digitized, however, may be difficult. In talking with agency personnel, project staff tried to make the distinction between born digital information and digital information that also exists in printed format. Agency personnel, for the most part, have difficulty distinguishing born digital information from all other digital information on agency websites. As a result, it may not be feasible to single out born digital information for access and preservation solutions. Instead, developing solutions for managing all state government information in digital formats for long term access may be more reasonable.

III. Changes in State Government Publishing Practices

Research findings from Phase I of the *Initiative* regarding publishing practices and trends within North Carolina state agencies (see Section I) confirm that state agency publishing is changing and the trend is to produce fewer documents in print and more in digital formats available through the Internet. The information gathered in Phase I serves as the basis for this Section as well as Sections IV and V of this paper.

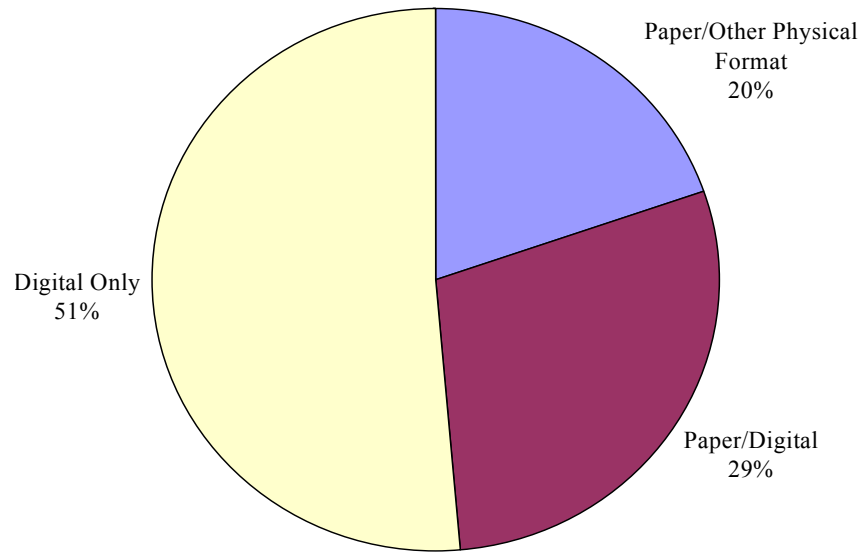
As previously mentioned, research into state agency publishing practices involved reviewing the State Library's state documents collection, examining state agency websites, and interviewing state agency personnel involved in publishing and information dissemination. Results from the various components of the research afforded project staff a good perspective on the state of agency publishing as well as insight into the "workings" of state government. Staff also identified opportunities and barriers in state government that could impact the development and implementation of solutions for permanent public access to state information.

Publication Formats

Using data collected through website examinations, publication reviews, and agency interviews, staff were able to approximate the percentage of state government information being produced and disseminated in printed formats, digital formats, or both. Overall, born digital publications make up approximately half of all publications, while publications solely in paper (or other tangible physical format) make up less than a quarter of those produced by state agencies in 2003. (Figure 1)

Agencies predict the amount of born digital publications will continue to rise. One third of state agencies interviewed predict that in 5 years, 90 % or more of their publications will be produced and disseminated in digital format only. The remaining two-thirds, while predicting an increase in digital-only information, also predict publishing a sizable percentage of publications in multiple formats (e.g., paper and digital).

Figure 1: Publication Formats: 2003



Agencies are currently providing direct access to databases of information through web interfaces and creating dynamically generated webpages.¹¹ Nearly two-thirds of state agencies interviewed indicate they provide access to statistical and directory-type information through web-enabled databases. Examples of statistical information available via web-enabled databases include soil analysis, air quality information, plant and wildlife sightings, mortality rates, and employment and criminal statistics. These databases allow users to manipulate data and create customized reports in a way that cannot be replicated in paper publications. This small but significant collection of government information databases may well prove the hardest type of resource for which permanent public access can be provided.

¹¹ Dynamically generated webpages, as opposed to static webpages, are created on-the-fly, usually from different components in a database that have been called together by a user command. These pages create difficulties for search engines indexing pages, since dynamic pages do not exist until they are called upon, and may change quite frequently as the underlying data is updated.

Reasons for Change

The two main forces driving the transition from paper to digital publications are improvements in digital technology and state budget cuts for agency printing. Some agencies have been specifically mandated by the General Assembly to produce and disseminate information through the Internet instead of printing paper documents.¹² All agencies, however, are feeling the pinch of tight budgets, which often results in a reduction in print publications.

While agencies are unhappy about the size of budget cuts, most feel the advantages of digital publishing go beyond cost savings. Among the advantages are ease of distribution and access, efficient updating of information, manipulation of data, and an expanded audience. A representative from the North Carolina Department of Health and Human Services describes how the Web has actually changed the type of publications the agency produces: “A decade ago we didn’t do many fact sheets, instead relying on longer form publications like reports or brochures. We have found that fact sheets are easy to do and easy for people to understand. They also can be updated more readily than a 30-page publication ... For instance, during the recent SARS activity, we had to update fact sheets almost hourly, as new details became available.” Additionally, as the Web has become ubiquitous, agencies’ target audiences have demanded that information be presented on the Web.

While most agencies believe they would print more items if they had the money, they feel digital dissemination provides too many advantages to be scaled back or abandoned. Responses indicate, however, that certain publications would remain in print, or exist in both print and digital formats. These publications would be geared toward audiences that lack ready access to computers or are uncomfortable using new technologies, or publications better suited and more easily used in paper format, such as maps or calendars. Most agencies agree that using print as the sole format for publications is not the preferred route for the future. They continue to strive, however, to serve traditional audiences for their publications as well as new audiences gained through the Internet.

¹² Session Law 2002-424 Section 14.1 specifically targets the Office of the Governor, the Office of the Lieutenant Governor, the Department of Administration, the Office of the State Auditor, the Office of State Budget and Management, the Board of Elections, the Department of Insurance, the Office of the Secretary of State, the Office of State Treasurer, the Office of Administrative Hearings, the Office of the State Controller, the Department of Cultural Resources, the General Assembly, the Office of State Personnel, the Department of Revenue, and the Rules Review Commission. Available at: <http://www.ncleg.net/SessionLaws/HTML/2001-2002/SL2001-424.html>.

IV. New Challenges to Permanent Public Access

“Permanent public access,” according to the American Association of Law Libraries, is “the process by which applicable government information is preserved for current, continuous, and future public access.”¹³ The North Carolina State Documents Depository System, administered by the State Library, and programs in State Archives and Records fulfill this goal for printed state publications and records. These programs are not structured or staffed at this time to accommodate born-digital information. For the present, state agencies and other producers of state publications are responsible for maintaining permanent public access; however, project research results reveal that agency staff lack the time and resources to consider long-term access to publications by users other than their immediate target audience. Maintaining print-only dissemination of information is no longer a reasonable option because of the advantages offered by digital information formats and presentations. Additionally many types of digital publications cannot be accurately replicated in printed formats. For these reasons, other solutions must be found for ensuring permanent public access to state government information.

Current Information: Challenges to Access

State agencies appreciate the value of up-to-date information on websites and constantly review and update information on their pages. Trends in state agency website management revealed in the research include growth in digital information, modifications to website format, increased multimedia and e-commerce transactional features, and more dynamically generated pages. Over one-third of the agencies interviewed indicate their websites will be undergoing a complete redesign within the next year. As a rule, agencies do not proactively alert users about new digital publications as they are published. The Office of the State Auditor is the only agency interviewed that has a comprehensive notification system in place. The constantly changing nature of government websites contributes to the difficulty of providing easy access to information, both now and in the future.

¹³ Richard J. Matthews et al., *State-by-State Report on Permanent Public Access to Electronic Government Information* (Chicago, IL: American Association of Law Libraries, 2003), 8, http://www.ll.georgetown.edu/aallwash/State_PPAREport.htm (accessed November 17, 2003).

Finding state government information on the Web can be very difficult. Within standard search engines, such as Google or Teoma, government publications are easily lost in the sea of information. As of 2000, the Web was estimated to contain over 4 billion public pages and 550 billion pages in the “invisible” Web, with 7 million new pages added daily.¹⁴ The invisible Web, also called the “deep” Web, consists of sites that are not crawled by search engine spiders, usually because pages are dynamically generated from databases.¹⁵ Standard search engines have limited indexing capability for dynamically generated pages. Their crawlers avoid indexing URLs with question marks and can be stymied by textboxes requiring input. *NC@Your Service*, the state portal for North Carolina, serves as a gateway to the branches and departments of state government; however, it is difficult to find specific state information through the portal without some knowledge of state government organization and hierarchy.¹⁶

Looking solely at static pages, the state of North Carolina’s web presence is estimated at 49.6 gigabytes of information and 457,000 files.¹⁷ Already, this is a formidable amount of information to search. Project research shows the amount of dynamically generated information available on North Carolina state agency websites will only continue to grow, making it more difficult to find information using standard search engines. Currently, in order to access statistical data or other information in databases, users must know of the existence of a database containing the specific data of interest and where it resides within state government webpages. North Carolina Public Records Law, G.S. 132-6.1, requires agencies to index databases to ensure they can be easily discovered.¹⁸ Of the 25 agencies that report to have web-enabled databases, only three (12 %) report that databases are indexed according to the Guidelines issued by State Archives and Records.¹⁹

¹⁴ Lyman, "Archiving the World Wide Web," 38.

¹⁵ *Invisible Web: What It Is, Why It Exists, How to Find It, and Its Inherent Ambiguity* (University of California, Berkeley, August 28 2003), <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html> (accessed September 23, 2003).

¹⁶ The state portal for North Carolina is *NC@Your Service Portal – the Official Website of the State of North Carolina*. See website at: <http://www.ncgov.com>.

¹⁷ Crawl of North Carolina websites, September 2003, using *Preserving Electronic Publications* software developed by the University of Illinois at Urbana-Champaign. The full results of the web crawl are available at: http://pep.library.uiuc.edu/NC_LatestStats20030909.html.

¹⁸ *North Carolina General Statutes*, Chapter 132.

¹⁹ State Public Records Services, "Public Database Indexing: Guidelines and Recommendations," Release 1.1 (Raleigh: Division of Archives and History, 1996), <http://www.ah.dcr.state.nc.us/e-records/pubdata/default.htm> (accessed November 20, 2003).

Making state information easier to find through standard search engines and customized access tools must be a goal for North Carolina state government. A study by IDC, an information technology consulting firm, suggests that half of all online searches conducted by knowledge workers are unsuccessful, potentially costing a company employing 1,000 knowledge workers \$2.5 million a year for wasted time. Additional costs of up to \$5 million annually may be accrued if workers spend time reworking existing information or duplicating efforts when information cannot be found.²⁰ As a test, *Initiative* project staff conducted a series of five searches for government publications from different state agencies in October 2003 (Appendix C). Some of the searches were for dynamically generated information, others for information on static pages. Using Google and the State Portal's search engine, searches were only successful part of the time. Information from databases did not appear in the results of either search engine; state government pages did not appear in some Google searches; and search terms with multiple meanings confused both Google and the State Portal's search engine, resulting in irrelevant hits.

Broken links make finding government information difficult as well. Larry Jackson, a researcher at the University of Illinois at Urbana-Champaign, notes that web-authoring is concerned with style and "look-and-feel," leading to frequent changes and updates on webpages, while "terms like 'government documents' convey, at least to the layman, an expectation of formality, official content, and permanence."²¹ A study of British government websites discovered that 25 % change their URL each year.²² The profusion of website tweaks, overhauls and URL changes common in state government pages, while providing a fresh look and incorporating new technologies, contribute to broken links and user disorientation. The problem is not unique to government information. Even with respected science and medical journals, 13% of the links referencing Internet sources in articles published between 2000 and 2003 no longer connect to the intended resource.²³ Better access tools are needed to assist users in finding information that has been moved or rearranged on websites.

²⁰ Susan Feldman and Chris Sherman, "The High Cost of Not Finding Information: An IDC White Paper," (IDC, 2001), 7-9, http://monkey.biz/Content/Default/Support/Resources/IDC_TheHighCostOfNotFindingInformation_1510.pdf (accessed September 12, 2003).

²¹ Larry S. Jackson, "Statistical Profiles of Web and Metadata Usage by Two U.S. State Governments," in *GSLIS Technical Report ISRN UIUCLIS--2002/6+EARCH* (Urbana-Champaign: University of Illinois at Urbana-Champaign, 2002), http://www.isrl.uiuc.edu/pep/papers/UIUCLIS_2002_6_EARCH/ (accessed May 21, 2003).

²² Rick Weiss, "On the Web, Research Work Proves Ephemeral," *Washington Post*, November 24 2003, A08, <http://www.washingtonpost.com/wp-dyn/articles/A8730-2003Nov23.html> (accessed November 25, 2003).

²³ Robert P. Dellavalle et al., "Going, Going Gone: Lost Internet References," *Science* 302, no. 5646 (2003): 788.

By the same token, older websites are not always updated as frequently as they should be, causing confusion for users. For example, an agency may remove links to older pages, but leave the pages up on the web servers. A Google search may retrieve the older pages or “floating webpages.” Site overhauls may not remove these old, outdated pages, thus creating the possibility of searchers retrieving two websites covering the same material with no indication that one is obsolete. When project staff examined agency websites, they discovered smaller offices and sometimes entire agencies often fail to maintain pages, as evidenced by outdated information and broken external links. The challenge of identifying the date and currency of information on websites leads to additional user confusion. While preserving historical websites is an important goal of the *Initiative*, current and non-current pages must be correctly identified so that users can quickly verify the currency of the information.

Historical Information: Challenges to Access and Preservation

Perhaps the most troubling concern about digital government publications is whether historical digital information will be available in the future. When project staff reviewed 333 currently produced serials from ten agency websites, back issues for 165 titles, or 50 % of the digital serials, were available online. Most agencies report maintaining or planning to maintain back issues of digital serials online, though many are not sure how long issues will remain available. Criteria for determining the length of time information stays online include content or subject of the publication, space consideration on the page or server, and terms of political office. Because the production of online serials is a fairly recent occurrence in state government, loss of valuable historical information is not yet too great, but the state must move swiftly, however, as digital publications begin to age and more back issues disappear from the Web.

Agencies tend to view monographs as more transient than serials, particularly as new editions replace outdated information. Over half the agencies report the length of time a monograph is available through the Web varies according to content and availability of newer editions. Almost three-fourths claim to remove older editions when a new edition becomes available for fear of confusion when information is superseded, outdated, or changed.

Access to historical publications taken offline presents yet another challenge. While state agencies indicate they are receptive to the public requesting to view digitally stored information (in fact, as several pointed out, they must comply with requests as per the North Carolina Public

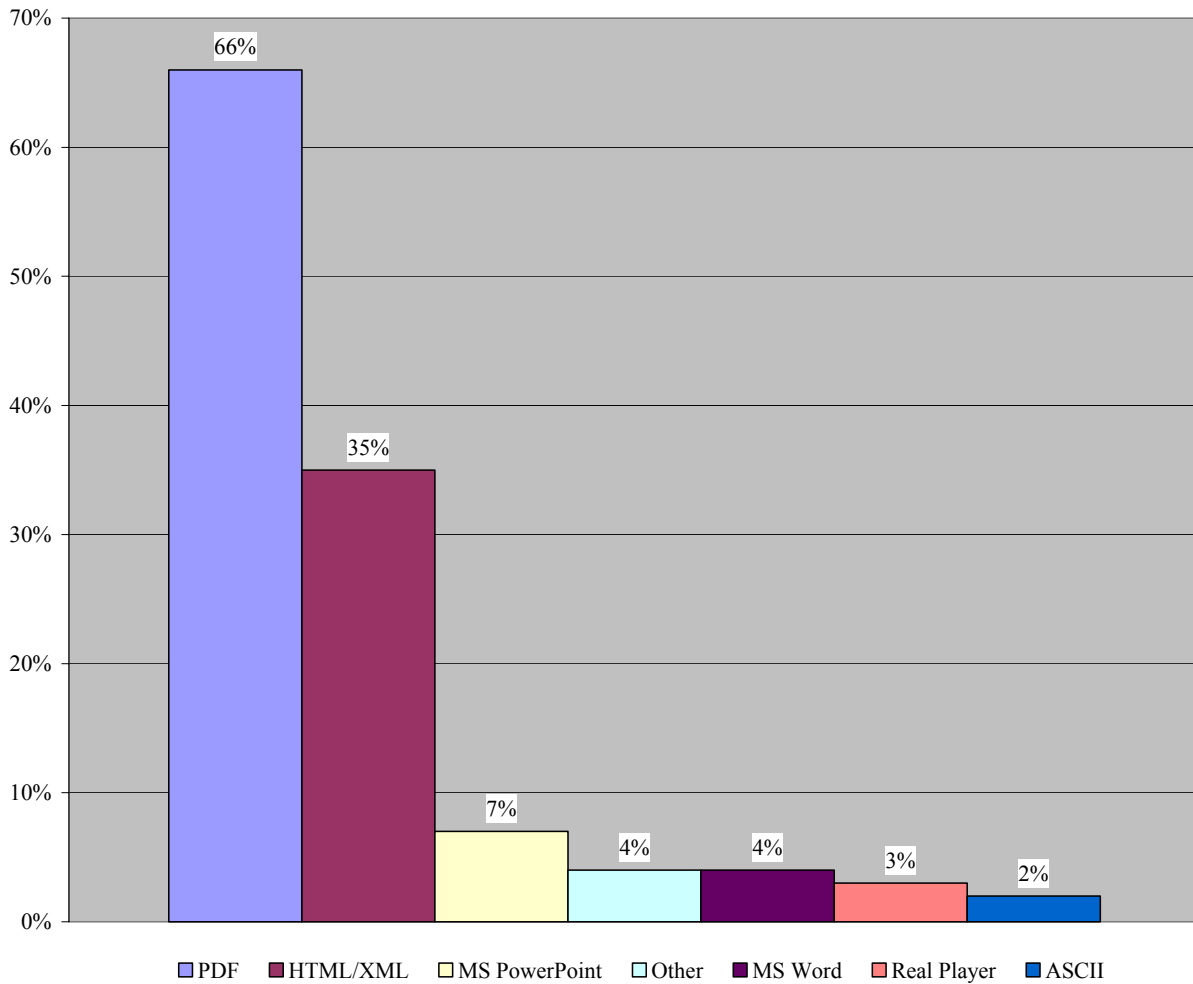
Records Law), there has been little effort to facilitate public access to this information. Almost no agencies maintain an index of digital publications stored offline. Agencies familiar with the North Carolina State Publications Clearinghouse indicate that they rely on the State Library for the historical record. This, of course, is only relevant for publications that still exist in print as well as digital formats. As a result, some type of centralized repository, or series of distributed repositories, searchable from a central location, may be necessary for historical digital information to truly be accessible.

Technical Barriers to Permanent Public Access

Agencies use a variety of file formats for digital publications with PDF and HTML as the most common formats for Internet publishing. Agencies also produce information in the following formats: RTF, GIF, Microsoft PowerPoint, Real Media Player, shape files, compressed/zipped files, audio files in WAV format, SQL and PHP (for databases), FileMaker Pro, MPEG for video, MS Access, and Windows Media Player (Figure 2). Just over one fourth of the agencies interviewed have a policy that provides guidelines for web design (i.e., requirements for style, metatags, and accessibility). In general, decisions about digital formats are left up to the person posting the information to the page, regardless of role (e.g., publication author or creator, graphic designer, or webmaster). The variety of file formats, many of them proprietary, and lack of standards make preservation efforts more difficult.

Approximately 50% of agencies claim they store or would store older serials issues once they are removed from the Web. For monographs, only 30% of agencies store older editions. In some cases, agency personnel have little knowledge of how agency information technology staff handles older digital documents and, by default, assume items are being stored. Methods of storing older publications taken offline are haphazard, with few concrete policies or standards guiding the process. Storage of offline publications might be on individual hard drives, network servers, optical disk, magnetic disk, or magnetic tape.

Figure 2: File Formats of Digital State Publications



None of the agencies interviewed have policies to address the issue of long-term preservation. Some, though, have considered short-term offline storage plans. For example, the Office of the State Auditor intends to keep its audit reports available on the Web for five years. They have not yet formulated a plan beyond this point and hope the State Library can assist in long-term preservation and access to the reports. Other agencies, like the North Carolina Department of Labor, store older digital publications in their design format, more for reprint purposes than for long-term preservation.

Of consequence to the long-term record and functioning of state government is the preservation of digital information. According to Abby Smith, director of the Council on Library and Information Resources, inadequate preservation is “the greatest risk to present and future

access to digital information.”²⁴ Digital information suffers from a short lifespan for three technological reasons: media degradation, hardware obsolescence, and software obsolescence.

In the end, digital information is nothing more than a series of 1s and 0s. There is nothing for the naked eye to see and no way to interpret the set of numbers into meaningful information without the intervention of a machine. Media on which digital information is stored degrades over time much more quickly than paper. Magnetic media and optical disks suffer from “bit rot,” – they lose some of the 1s and 0s stored on their surfaces. At first, the computer can compensate for the loss by filling in the missing pieces, but eventually the loss becomes too great and the information contained within is gone. Paul Conway, director of Information Technology Services at Duke University Library, describes this as “The Dilemma of Modern Media”—as information density for storage formats grows, the life expectancy of the media on which the information is stored decreases.²⁵ Consider the permanence of clay tablets, but imagine the volume of trying to store a DVD’s worth of information on them, let alone the problems of losing the moving picture and having to transcribe the sound.

Usually, even before the media can degrade, hardware obsolescence renders reading digital information impossible. Most computers can no longer read 8-inch and 5 ¼ inch floppy disks because they no longer have a drive that accommodates them. 3 ½ inch diskettes are rapidly going the same route toward obsolescence. DVDs may soon replace CDs. In addition to issues with storage medium, there are problems with software used to create information. Software is upgraded or changed completely, rendering information created in older programs unreadable. Because software is usually proprietary, the code to determine how the software reads information becomes lost as companies go out of business or upgrade their products. Presently, Microsoft is a dominant player in the software environment, but there are no guarantees Microsoft will exist in perpetuity. There is also no assurance that newer software versions will be able to translate information from older programs. Operating systems have also evolved, making documents created on a Commodore 64, once the most popular computer in the United States, unreadable by Windows XP.

²⁴ Abby Smith, "Digital Preservation: An Individual Responsibility for Communal Scholarship," *EDUCAUSE Review*, May/June 2003, 10, <http://www.educause.edu/ir/library/pdf/erm0338.pdf> (accessed September 17, 2003).

²⁵ Paul Conway, "Preservation in the Digital World," (Council on Library and Information Resources, 1996), <http://www.clir.org/pubs/abstract/pub62.html> (accessed September 24, 2003).

In sum, information in paper format, or other format that can be viewed by the human eye (even aided with a magnifying glass), can suffer from benign neglect and still be recovered after decades of disuse. Digital information, on the other hand, requires constant attention or it becomes unrecoverable.

An excellent example of the perils of digital publication is the BBC Domesday Project videodiscs. The project was inspired by a census of England taken by William the Conqueror in 1086, known as the Domesday Books, which today reside in the National Archives. The BBC decided to undertake a similar census in 1986, resulting in two multimedia disks containing maps, photographs, video, and text collected from across the United Kingdom. By 2001, there was only one computer in the National Archives that could still read the disks and the hardware had become fragile. Through heroic preservation efforts the information was saved, but it took 16 months of dedicated effort.²⁶ Such expensive and time-consuming preservation efforts are not feasible for preserving the entire body of digital government information produced by North Carolina state government. Instead, plans must be made now to appraise the value of state information and determine the state's approach to implementing digital preservation solutions.

Non-Technical Barriers to Permanent Public Access

Beyond the technical issues, other problems hinder efforts to preserve digital information. As Peter Lyman points out, "The Web is not stored in attics; it just disappears".²⁷ The average lifespan of a webpage is just 44 days, with only 44 % of webpages found in 1998 still available a year later. Libraries and other memory institutions, which traditionally have owned the physical objects that contain information, instead now provide access to information resources in remote locations. Because the Web is so distributed, no one really feels responsible for its care. Most individuals creating information lack the economic incentives, technical expertise, or the time to preserve their creations. The Web and government information are a public good. Even though

²⁶ Jeffrey Darlington, Andy Finney, and Adrian Pearce, "Domesday Redux: The Rescue of the BBC Domesday Project Videodiscs," *Ariadne*, no. 36 (2003), <http://www.ariadne.ac.uk/issue36/tna> (accessed August 5, 2003).

²⁷ Lyman, "Archiving the World Wide Web," 39.

everyone benefits from its preservation, no single institution or individual feels responsible for the task, instead hoping that someone else may be willing to take on the challenge.²⁸

While archives and libraries have the interest to preserve information, they may be lacking the resources to create a digital preservation strategy. The magnitude of the problem can stymie attempts because it is difficult to determine the proper starting point, know what technical expertise will be needed, find trusted tools to aid in the process, and determine costs. Even when there is a clear mandate to preserve information, the lack of previous knowledge and experience with digital information makes it difficult to convince policy makers and funding sources of the need to allocate resources toward the effort.

The cost of the potential loss of data is also hard to quantify and can take a back seat when pitted against more pressing issues. For example, the United State National Archives and Records Administration (NARA) developed the Electronic Records Archives (ERA) program in an attempt to improve the storage and preservation of electronic records. The General Accounting Office issued a report with concerns that the project's capabilities are not fully established and that "NARA is unable to objectively track the cost and schedule of the ERA project."²⁹ As a result, the Senate Appropriations Committee deferred the ERA's funding for this year over worries that the money would not be wisely spent.³⁰ Similar concerns about overspending and uncertain savings for technology infrastructure for the FBI and the Office of Personnel Management, however, have not resulted in the withdrawal of funds.³¹

²⁸ Brian F. Lavoie, "The Incentives to Preserve Digital Matierals: Roles Scenarios, and Economic Decision-Making," (Dublin, OH: OCLC Online Computer Library Center, INC, 2003), 28-31, <http://www.oclc.org/research/projects/digipres/incentives-dp.pdf> (accessed July 1, 2003).

²⁹ Subcommittee on Technology, Information Policy, Intergovernmental Relations, and the Census, Committee on Government Reform, *Electronic Records: Management and Preservation Pose Challenges*, July 8 2003, 4, <http://www.gao.gov/new.items/d03936t.pdf> (accessed 2003, November 13).

³⁰ Ted Leventhal, "Senate Panel Seeks to Move Funds for E-Archives to Amtrak," *National Journal's Technology Daily*, September 10 2003.

³¹ Stephen Barr, "Savings Uncertain from Electronic Tracking of Employees," *Washington Post*, September 24 2003, <http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A55367-2003Sep23¬Found=true> (accessed November 21, 2003), Larry Barrett, "FBI: Under the Gun," *Baseline*, September 10 2003, <http://www.baselinemag.com/article2/0.3959.1261145.00.asp> (accessed November 21, 2003).

Challenges within North Carolina State Government

Many of the challenges to permanent public access mentioned here exist in North Carolina as well. Again, while state agencies are concerned with making sure their constituents receive the current information they need, their focus is not on maintaining historical information or trying to broaden their audience for current information. Budget cuts and staff reductions, along with increased demand for more information online, have stretched agency resources as far as they can go, without adding the additional responsibility of providing permanent public access to their information. While agencies handling statistical data, like the Employment Security Commission, tend to be more conscious of the value of historical data, agency public information officers (PIOs) generally view preservation of historical information to be a low priority or out of the scope of their job responsibilities.

One agency PIO commented, “I think that one thing the State Library is going to have to come to grips with is the fluid nature of digital publishing. What our web site says one day may be different the next, and we don’t always archive out-of-date material. The kind of historical record that has existed in the form of printed documents won’t be as readily available in the world of digital publishing.” Another admitted feeling overwhelmed with trying to keep on top of current information, and commented, “It would be wonderful if the State Library could take care of older publications.” While the first PIO is correct in saying the State Library will not be able to retain everything, the second one’s plea indicates the State Library may be in the best position to attempt to preserve government information. Relying on state agencies to be responsible for the historical record with no additional support will most surely result in losses of valuable digital state information.

The State Library, however, cannot simply take over the preservation and access of digital state information without the cooperation of the agencies producing the information. While North Carolina publishing practices have never been tightly centralized, the growth of born digital publishing has created even more decentralized publishing practices. Less than a third of agencies report having a centralized publication and distribution system, many mentioning that the ease of posting to the Web has contributed to the lack of centralization. Individual divisions and sections within agencies have a lot of autonomy to produce publications as they see fit, giving them flexibility to respond to their target audiences. This, however, also leads to the lack

of standards for formats, adding to the difficulty of preserving historical information. The proliferation of formats increases the possibility that information will be lost as software becomes obsolete.

The challenges of preservation and access may be better addressed using specialized staff and pooled resources. The results of this approach may well lead to the designation of a central repository for digital government information. Because state agencies have had autonomy in publishing decisions, there may be resistance to the idea of standards and centralized management of their information. Agencies may, instead, prefer to have digital information stored in a distributed fashion. In either case, state agencies and the State Library must work together to ensure the preservation and continued access to historical state information.

V. Addressing the Challenges: Potential Solutions

Digital Preservation Approaches

One approach to preservation is migration, where digital information is periodically reformatted to be accessed using current hardware and software. Over time, this approach can lead to loss of data and formatting as different programs interpret code differently. Migration is currently the principal means by which current preservation products manage digital data.

Another approach is emulation, where newer software platforms are made to emulate older platforms, based on information stored alongside the digital information to be preserved. However, research into emulation is still very exploratory.³² Su Shing-Chen describes the paradox of digital preservation: “On the one hand we want to maintain digital information intact as it was created; on the other we want to access this information dynamically and with the most advanced tools.”³³ Determining how important it is to preserve the original look and feel of a digital resource along with the information within will weigh heavily on preservation choices.

Code is now available for open-source software, which can make access and migration of digital information easier. HTML and other hypertext, for example, are open-source. Unfortunately, PDF, the most popular form for digital North Carolina state documents, is a proprietary format licensed by Adobe Systems, Inc. One possible solution, to which Adobe is agreeable, is to create an archival standard of PDF, known as PDF/A. PDF/A would be a platform independent version of PDF, allowing documents to still be read, even if Adobe should go out of business. The International Organization for Standardization may approve such a format in the near future.³⁴ Standardization of digital formats will facilitate the preservation and access to digital materials by simplifying the number of formats to migrate or emulate and allowing the creation of standard search tools.

³² Daniel Greenstein and Abby Smith, "Digital Preservation in the United States: Survey of Current Research, Practice, and Common Understandings," in *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program* (Washington, D.C.: Library of Congress, 2002), 115, <http://www.digitalpreservation.gov/index.php?nav=3&subnav=1> (accessed November 15, 2003). For more information on emulation, see the CAMILEON website: <http://www.si.umich.edu/CAMILEON/>.

³³ Su-Shing Chen, "The Paradox of Digital Preservation," *Computer*, March 2001, 4.

³⁴ Gail Repsher Emery, "E-Documents Need E-Preservation," *Washington Technology* 17, no. 23 (2003), http://www.washingtontechnology.com/news/17_23/federal/20235-1.html (accessed September 17, 2003). See also Michael Looney, "The Need for Digital Archiving Standards," *Syllabus: Technology for Higher Education*, March 2003, <http://www.syllabus.com/article.asp?id=7362> (accessed March 7, 2003), and Nigel McFarlane, "PDF Keeps It All Nice," *The Sydney Morning Herald*, July 29 2003 (accessed August 4, 2003).

Preservation Initiatives: National Level

Projects and research dedicated to addressing the challenges of long-term access to digital information are underway. The largest digital preservation project in the United States is the National Digital Information Infrastructure and Preservation Program, led by the Library of Congress. Congress appropriated \$5 million to study the problem initially, and has appropriated an additional \$99.8 million for the program that will span at least five years. With partners in the public and private sector, the Library of Congress will try to create an architecture for digital preservation; determine best practices for preservation, both for business models and technology; and institute standards. Web pages, digital periodicals, digital video, digital audio, and other multimedia formats will be considered.³⁵

Other national initiatives that comprise joint efforts between libraries and archives are also in progress. In the United States, the Government Printing Office (GPO) and the National Archives and Records Administration (NARA) forged an agreement in 2003 to jointly ensure free and permanent access to digital federal documents.³⁶ In Canada, the National Library and National Archives have joined to form the Library and Archives Canada in order to maintain Canada's documentary heritage in all formats.³⁷ Other large-scale national programs for digital preservation are underway in Australia, France, the Netherlands, and the United Kingdom.³⁸

Universities and consortia in the United States are also conducting research and creating digital archives to preserve intellectual output. For instance, the Massachusetts Institute of Technology's DSPACE is a university repository system designed to capture, store, index, and

³⁵ *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*, (Washington, D.C.: Library of Congress, 2002), 1-6, <http://www.digitalpreservation.gov/index.php?nav=3&subnav=1> (accessed September 18, 2003).

³⁶ Miriam Drake, "Agreement Ensures Permanent Public Online Access to Government Information," *Information Today NewsBreaks*, August 25 2003, <http://www.infotoday.com/newsbreaks/nb030825-1.shtml> (accessed September 11, 2003).

³⁷ "Canada: Looking Forward to the Digital Future," *Information Retrieval and Library Automation*, June 2003, 1-3.

³⁸ Neil Beagrie, "National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity," (Council of Library and Information Resources and Library of Congress, 2003), <http://www.clir.org/pubs/abstract/pub116abst.html> (accessed September 17, 2003).

distribute the works of MIT professors.³⁹ Other projects around the country are attempting to make searchable repositories of now-defunct websites.⁴⁰

Preservation Initiatives: State Level

States are also involved in efforts to ensure permanent public access to digital government information, though most programs are still in their infancy. According to a study by the American Association of Law Libraries published in June 2003, only Colorado has enacted legislation that explicitly addresses permanent public access to government information,⁴¹ though other states, such as Illinois and Georgia have passed legislation to modify its library depository law to include digital publications.⁴² According to the study, three-fifths of the states have begun to address the need for permanent public access to digital government information in some fashion. Additionally, OCLC (the Online Computer Library Center), a non-profit company, has created a digital archiving service.⁴³ Several states, including Connecticut and Michigan, are using the OCLC service to preserve digital government documents.

Access Initiatives: Federal and State

In addition to preservation initiatives and research, federal and state governments are also addressing the need for easy access to current government information. At the federal level, the National Biological Information Infrastructure (NBII), which provides access to data and information relating to biological resources, is an example of such an effort. The program links information sources from across the nation and around the world, allowing researchers to easily determine what information exists for their field of study. It fulfills e-government goals by facilitating citizen and business interactions with government and saves taxpayer dollars by

³⁹ Vivien Marx, "In DSpace, Ideas Are Forever," *The New York Times*, August 3 2003, 8(L).

⁴⁰ See Appendix C in Patricia Cruse and Chuck Eckman, "Environmental Scan: Preliminary Survey Results (Ver. 3.2)," in *Web-based Government Information Project: a Mellon Funded Initiative of the California Digital Library* (California Digital Library, 2003).

⁴¹ Matthews et al., *State-by-State Report on Permanent Public Access to Electronic Government Information*, 19-20.

⁴² As an example, Georgia's database for state documents, GALILEO, and their depository rules can be found at: <http://www.libs.uga.edu/govdocs/collections/georgia.html>.

⁴³ See website at: <http://www.oclc.org/digitalArchive>.

reducing the possibility of duplicative research and time spent searching for information. It also expands the potential audience for the information.⁴⁴

Another federal initiative is DisasterHelp.gov, which brings together information from different federal agencies about federal response and assistance to natural and man-made disasters.⁴⁵ Both the NBII and DisasterHelp.gov organize information and make it searchable through metadata. Metadata, which literally means “data about data,” provides descriptive information about a resource, such as the author, title, and summary of the content. Metadata also aids in digital preservation, by documenting the software and system information needed to view the digital information.

State libraries across the country, from Washington to Rhode Island, including North Carolina, have worked to create state Government Information Locator Service (GILS), an access tool to aid in the retrieval of current state government information.⁴⁶ Based on Washington State Library’s model, states have created their own version of GILS metadata to facilitate information retrieval. The GILS metadata are placed as metatags in state government webpages. Special search engines, such as Find-It! Illinois, use the metadata to retrieve information. North Carolina has created its own GILS guidelines, in part to fulfill G.S. 132-6.1 of the Public Records Law that requires state agencies to index their databases. The State Library also initiated a project in 1998 that involved the application of NC GILS metadata to state government pages and the development of a customized search engine to facilitate locating North Carolina state government information on the Web. The system prototype of the project, *FIND NC*, was developed from 1998-2001; however, staff changes and budget priorities have prevented further development beyond the prototype stage. States are also using library catalogs and other metadata schemes to facilitate access to digital government information.

⁴⁴ Ron Sepic and Kate Kase, "The National Biological Information Infrastructure as an E-Government Tool," *Government Information Quarterly* 19 (2002): 410. See website at: <http://www.nbii.gov/>.

⁴⁵ See website at: <https://disasterhelp.gov/portal/jhtml/index.jhtml>.

⁴⁶ See websites at: <http://find-it.wa.gov/>, <http://www.finditillinois.org/>, <http://www.find-it.state.ri.us/>, and <http://www.findnc.org>.

Conclusion: A Call for Action

Research conducted through the *Access to State Government Information Initiative* over the past year has confirmed the trend toward digital distribution of government information. Information creators within state agencies, beleaguered from years of fiscal tightness and audience demands for more information more quickly, do not have time or resources to handle permanent access to digital government information. In a survey done by the Library of Congress, one participant observed that people producing information “are too busy creating to become their own archivists.”⁴⁷ Similarly, in state government, project research shows that while agency staff are concerned about getting information to their audience, managing and disseminating publications for most is one of a myriad of duties. State agency input and participation in state preservation and access projects is vital, however, because steps to preserve digital information must be taken at the point of creation. Agency cooperation is critical to the implementation and success of digital access and preservation solutions in North Carolina state government.

State librarians and archivists, traditional keepers of state information, must also be involved and are in the best position to lead efforts to ensure permanent public access to state information in all formats. The State Library and the State Archives and Records, hold a unique position in state government in that their primary purpose is to collect, preserve, and facilitate access to all state agency information. This position affords the Library and the Archives and Records a broad perspective on the needs of user communities and an objectivity that allows them to facilitate discussions between state agencies that have differing interests and priorities for information access and preservation.

Librarians and archivists also have experience in developing and maintaining systems for accessing state information (e.g., catalogs, finding aids) as well as selecting materials for collections which involves making decisions about the long-term value and status of information. According to the Library of Congress, “‘Saving the Web,’ then, is no more feasible or desirable than saving the contents of everything that has ever been put to paper, to film, and to recorded

⁴⁷ *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*, 30.

sound disc across the globe.”⁴⁸ As professional keepers of the historical record for the state, the State Library and the State Archives and Records are able to view state information in a fair and impartial manner and do not place undue weight on any one agency’s output or try to influence the historical record. Both agencies also facilitate access to state government information through catalogs, finding aids, and web-based access tools. As the digital world adds complexity to the tasks of selection, preservation, and access to state information, the principles of librarianship and archival theory that guide state librarians and archivists in managing materials in tangible formats also enable them to tackle the difficult issues of this new information age.

Time for Action

The time to act is now. Even though state agencies have not yet addressed the issue of permanent access to their digital information, the consequences to date are not tragic. Most agency websites began as fledgling ventures and until a few years ago did not contain much in the way of born digital information and publications. Project research indicates agencies claim not to have removed a lot of older information from websites yet because there is still space on web servers for the information. The question now is what happens when that server space is full? What if this information is deleted to make space for current information? Or, what if it is transferred to a CD for storage? The availability of digital records and publications beyond today is not assured and the probability of this information disappearing is high. Because the amount of digital state government information continues to grow on a daily basis, the state must begin addressing the challenges of access and preservation now before this valuable information is lost forever.

Research conducted during Phase I of the *Access to State Government Information Initiative* provides a solid foundation for determining the state’s approach and developing a plan of action for providing permanent public access to digital state government information. Knowledge of the current and probable future state of publishing in state government gained from the research will provide the framework for solutions development. A Solutions Work Group composed of librarians, data specialists, state agency personnel, archivists, digital information experts, and government information specialists will work to explore options for meeting the challenges of digital preservation and access during Phases II and III of the *Initiative*.

⁴⁸ Ibid., 27.

The process for solutions development will not be simple. There is no one “right” method for storing information, no magic bullet that will make all information, or only the important information, instantly accessible now and in the future. There are no “best practices” to emulate and no definitive solutions to implement. The mere “nature of the beast”—the complexities of digital information formats and presentations and the volatile nature of the technology that enables them—makes the process difficult.

This “call for action” is for North Carolina state government to acknowledge the need to deal with the issues of digital state information and start laying the groundwork for sustaining ongoing efforts to realize workable solutions for ensuring the existence, availability, and usability of government information over time, regardless of format. As the Library of Congress states, “action is needed now, not some time in the future; and everyone—from creators to custodians—must contribute to the solution and learn to operate fluently in a world of constant and unpredictable change.”⁴⁹ We couldn’t agree more!

⁴⁹ Ibid., 16.

Bibliography

- American Heritage Dictionary of the English Language*. 4th ed. Boston: Houghton Mifflin, 2000.
- Barr, Stephen. "Savings Uncertain from Electronic Tracking of Employees." *Washington Post*, September 24 2003, 2. <http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A55367-2003Sep23¬Found=true> (accessed November 21, 2003).
- Barrett, Larry. "FBI: Under the Gun." *Baseline*, September 10 2003. <http://www.baselinemag.com/article2/0,3959,1261145,00.asp> (accessed November 21, 2003).
- Beagrie, Neil. "National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity." Council of Library and Information Resources and Library of Congress, 2003. <http://www.clir.org/pubs/abstract/pub116abst.html> (accessed September 17, 2003).
- Building a National Strategy for Preservation: Issues in Digital Media Archiving*. Edited by Amy Friedlander. Washington, D.C.: Council on Library and Information Resources and the Library of Congress, 2002. <http://www.clir.org/pubs/abstract/pub106abst.html> (accessed November 13, 2003).
- "Canada: Looking Forward to the Digital Future." *Information Retrieval and Library Automation*, June 2003, 1-3.
- Chen, Su-Shing. "The Paradox of Digital Preservation." *Computer*, March 2001, 2-6.
- Conway, Paul. "Preservation in the Digital World." Council on Library and Information Resources, 1996. <http://www.clir.org/pubs/abstract/pub62.html> (accessed September 24, 2003).
- Cruse, Patricia, and Chuck Eckman. "Environmental Scan: Preliminary Survey Results (Ver. 3.2)." In *Web-based Government Information Project: a Mellon Funded Initiative of the California Digital Library*: California Digital Library, 2003.
- Darlington, Jeffrey, Andy Finney, and Adrian Pearce. "Domesday Redux: The Rescue of the Bbc Domesday Project Videodiscs." *Ariadne*, no. 36 (2003). <http://www.ariadne.ac.uk/issue36/tna> (accessed August 5, 2003).
- Dellavalle, Robert P., Eric J. Hester, Lauren F. Helig, Amanda L Drake, Jeff W Kuntzman, Marla Graber, and Lisa M Schilling. "Going, Going Gone: Lost Internet References." *Science* 302, no. 5646 (2003): 787-88.
- Drake, Miriam. "Agreement Ensures Permanent Public Online Access to Government Information." *Information Today NewsBreaks*, August 25 2003. <http://www.infotoday.com/newsbreaks/nb030825-1.shtml> (accessed September 11, 2003).
- Subcommittee on Technology, Information Policy, Intergovernmental Relations, and the Census, Committee on Government Reform. *Electronic Records: Management and Preservation Pose Challenges*, July 8 2003. <http://www.gao.gov/new.items/d03936t.pdf> (accessed 2003, November 13).
- Emery, Gail Repsher. "E-Documents Need E-Preservation." *Washington Technology* 17, no. 23 (2003). http://www.washingtontechnology.com/news/17_23/federal/20235-1.html (accessed September 17, 2003).

- Feldman, Susan, and Chris Sherman. "The High Cost of Not Finding Information: An IDC White Paper." 10: IDC, 2001.
http://monkey.biz/Content/Default/Support/Resources/IDC_TheHighCostOfNotFindingInformation_1510.pdf (accessed September 12, 2003).
- Greenstein, Daniel, and Abby Smith. "Digital Preservation in the United States: Survey of Current Research, Practice, and Common Understandings." In *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*, 113-22. Washington, D.C.: Library of Congress, 2002.
<http://www.digitalpreservation.gov/index.php?nav=3&subnav=1> (accessed November 15, 2003).
- History, Division of Archives and. "North Carolina Guidelines for Managing Public Records Produced by Information Technology Systems." Raleigh: Department of Cultural Resources, 2000. <http://www.ah.dcr.state.nc.us/e-records/manrecrd/manrecrd.htm> (accessed November 12, 2003).
- Invisible Web: What It Is, Why It Exists, How to Find It, and Its Inherent Ambiguity*. University of California, Berkeley, August 28, 2003.
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html> (accessed September 23, 2003).
- Jackson, Larry S. "Statistical Profiles of Web and Metadata Usage by Two U.S. State Governments." In *GSLIS Technical Report ISRN UIUCLIS--2002/6+EARCH*. Urbana-Champaign: University of Illinois at Urbana-Champaign, 2002.
http://www.isrl.uiuc.edu/pep/papers/UIUCLIS_2002_6_EARCH/ (accessed May 21, 2003).
- Lavoie, Brian F. "The Incentives to Preserve Digital Materials: Roles Scenarios, and Economic Decision-Making." Dublin, OH: OCLC Online Computer Library Center, Inc, 2003.
<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf> (accessed July 1, 2003).
- Leventhal, Ted. "Senate Panel Seeks to Move Funds for E-Archives to Amtrak." *National Journal's Technology Daily*, September 10 2003.
- Looney, Michael. "The Need for Digital Archiving Standards." *Syllabus: Technology for Higher Education*, March 2003. <http://www.syllabus.com/article.asp?id=7362> (accessed March 7, 2003).
- Lyman, Peter. "Archiving the World Wide Web." In *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, edited by Amy Friedlander, 38-51. Washington, D.C.: Council on Library and Information Resources and the Library of Congress, 2002. <http://www.clir.org/pubs/abstract/pub106abst.html> (accessed November 13, 2003).
- Marx, Vivien. "In DSpace, Ideas Are Forever." *The New York Times*, August 3 2003, 8(L), col. 01.
- Matthews, Richard J., Anne E Burnett, Charlene C. Cain, Susan L. Dow, David L. McFadden, and Mary Alice Baish. *State-by-State Report on Permanent Public Access to Electronic Government Information*. Chicago, IL: American Association of Law Libraries, 2003.
http://www.ll.georgetown.edu/aallwash/State_PPAreport.htm (accessed November 17, 2003).
- McFarlane, Nigel. "PDF Keeps It All Nice." *The Sydney Morning Herald*, July 29 2003 (accessed August 4, 2003).

Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program. Washington, D.C.: Library of Congress, 2002.

<http://www.digitalpreservation.gov/index.php?nav=3&subnav=1> (accessed September 18, 2003).

Sepic, Ron, and Kate Kase. "The National Biological Information Infrastructure as an E-Government Tool." *Government Information Quarterly* 19 (2002): 407-24.

Smith, Abby. "Digital Preservation: An Individual Responsibility for Communal Scholarship." *EDUCAUSE Review*, May/June 2003, 10-11.

<http://www.educause.edu/ir/library/pdf/erm0338.pdf> (accessed September 17, 2003).

State Public Records Services. "Public Database Indexing: Guidelines and Recommendations." Release 1.1. Raleigh: Division of Archives and History, 1996.

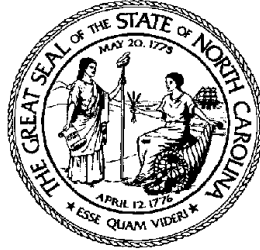
<http://www.ah.dcr.state.nc.us/e-records/pubdata/default.htm> (accessed November 20, 2003).

Weiss, Rick. "On the Web, Research Work Proves Ephemeral." *Washington Post*, November 24 2003, A08. <http://www.washingtonpost.com/wp-dyn/articles/A8730-2003Nov23.html> (accessed November 25, 2003).

Appendix A: Content and Purpose of State Information

Content/Purpose of State Information	Examples	Audience/Users
Government Operations	Audit reports; <i>North Carolina Administrative Code</i> ; Session Laws; State Budget; retirement manuals for state employees; public records that reflect the transaction of official state government business	State employees Legal community Business community Legislators Journalists/Media Educators/Scholars Students Historians
Statistical Information	Unemployment statistics; agricultural production statistics; crime statistics; demographic statistics	Business community Agricultural community Law enforcement Educators/Scholars Students Legislators Journalists/Media Historians State employees
Public/Educational Information for Citizens of North Carolina	Fact sheets about health and environmental hazards; information on tourist attractions and vacation destinations; descriptions of schools and universities; state transportation maps	Citizens Journalists/Media Educators/Scholars Students Historians Legislators
Regulatory Information	Rules governing air quality and waste disposal for factories and businesses; State Port Authority operations information for businesses shipping goods into the state; curriculum requirements for teachers in N.C. public schools; fishing limits for commercial fishermen.	Regulated communities (industries) Business community Legal community Legislators Industry/trade associations Non-profit organizations Journalists/Media Educators/Scholars Students Historians

Appendix B: Survey Tool



State Library of North Carolina
Access to State Government Information Initiative

SURVEY OF STATE AGENCY PUBLISHING PRACTICES

PART A: Contact and Department Information
PART B: Current and Future Publishing Practices
PART C: Born Digital Information and Publications
PART D: Databases

Instructions

Project staff from the State Library will conduct the **Survey of State Agency Publishing Practices** through personal interviews with you and other state agency personnel involved in producing, publishing, and/or distributing state government publications and information. The attached **Survey** will be used as a guide for the interviews.

Please look over the questionnaire before the interview to familiarize yourself with the types of information we are seeking. **You do not need to complete the Survey prior to the interview.** Feel free, however, to make notes on the survey that may be helpful during the interview.

Contact

Kristin Martin
Digital State Documents Librarian
Access to State Government Information Initiative
919-733-3683
kmartin@library.dcr.state.nc.us



***State Library of North Carolina
Access to State Government Information Initiative***

June 2003

Fellow State Government Information Employees:

Thank you for agreeing to participate in the **Survey of State Agency Publishing Practices**. This survey is part of the research component of the *Access to State Government Information Initiative* sponsored by the State Library.

The State Library is the agency legally mandated to facilitate public access to state agency publications under North Carolina General Statute 125-11. The State Library has fulfilled this responsibility since 1987 by receiving copies of all printed publications from state agencies and distributing them to 25 participating libraries across the state for easy public access (i.e., North Carolina State Publications Clearinghouse and North Carolina State Depository Library System). Today's technologies and state budget cuts, however, are changing the way state government information is published and distributed. The result is more digital, Web-based information and less printed paper documents.

In response to the rapidly changing environment in state government publishing, the State Library is leading the *Access to State Government Information Initiative* to better understand the changes and assess the viability of the State Library's programs for ensuring public access to published state government information. The participation and cooperation of state agency personnel, librarians across the state, and digital information professionals is critical to the Initiative's success.

The information gathered from the **Survey of State Agency Publishing Practices** will provide insight into how agencies are producing, managing, and maintaining databases of information and publications in print and digital formats. Most importantly, the research results will help guide and direct the state's efforts to develop solutions for managing digital publications and statistical data to ensure continued public access to state government information in all media and formats.

We look forward to working with you and appreciate your interest and participation.

Jan Reagan
Project Manager, ASGI Initiative
Head, Documents Branch

State Library of North Carolina
jreagan@library.dcr.state.nc.us

PART A: CONTACT AND DEPARTMENT INFORMATION

Please complete the following information

Name

Work Title

Phone Number

Phone Extension

Fax Number

E-mail

Preferred Method for Contact

Department

Physical Location

Mailing Address

City, State, Zip

Brief Description of Duties

Please add the names of any offices for which you handle publications. Please use the office's full hierarchy.

Department

Division

Section (and smaller, if necessary)

Department

Division

Section (and smaller, if necessary)

Department

Division

Section (and smaller, if necessary)

Department

Division

Section (and smaller, if necessary)

PART B: CURRENT AND FUTURE PUBLISHING POLICIES AND PRACTICES

1. Briefly describe the content of the information that is published by your agency. Think about the types of information published (e.g. directories, newsletters, research reports) and any key publications produced by the agency. Consider all types of formats and media (e.g. paper, video, audio, digital).

2. Who is the target audience for your publications? (Check as many as apply)

- other agency staff
- other state government employees
- local governments
- business community
- nonprofit organizations
- legislative members or their staff
- non-governmental specialized research community
- general public
- other: _____

3. Current publishing practices and policies

3a. Briefly describe how the agency determines what information or types of information are published and distributed (e.g. regarding publication content, standards or methods of distribution):

3b. If there is a publishing policy currently in place, please attach the policy and give the date of the policy and the name of the issuing office:

Issuing office: _____

Contact person: _____

Date: _____

4. How are printed publications distributed? (check all that apply)

- Mailing list
- upon request
- State Publications Clearinghouse
- Other: _____

5. North Carolina State Documents Depository Program

5a. Are you familiar with the North Carolina State Documents Depository Program and the State Publications Clearinghouse, run by the State Library of North Carolina (G.S. 125-11)?

yes

no

5b. Does your agency currently send printed publications to the State Library for the North Carolina State Publications Clearinghouse?

yes

no (why not?):

6. Do you consider your agency to have a centralized publication and distribution system (e.g. all publications go through one office) or a decentralized publication and distribution system (e.g. each division, section is in charge of their own publications from start to finish with little to no overall agency oversight)?

- _____ centralized
- _____ decentralized
- _____ other (explain):

7. What percent of the publications (make your best estimate) are published in:

- _____ % paper only
- _____ % other physical format (e.g. video, audio, CD-ROM)
- _____ % web-based digital only
- _____ % more than one format (e.g. paper and digital)

8. How does your agency determine which format(s) to use when creating its publications?

9. What do you think these percentages will be for your agency's publications in the future?

9a. 2006 (three years)

- _____ % print only
- _____ % other physical format
- _____ % web-based digital only
- _____ % more than one format

9b. 2008 (five years)

- _____ % print only
- _____ % other physical format
- _____ % digital only
- _____ % more than one format

10. If you foresee a shift from printed to digital publishing, or other change in publication format, what are the reasons behind that shift?

11. Future publishing practices

11a. Describe how the agency will determine its future publishing practices. (e.g., is there a transition plan for moving print to digital or a strong commitment to continue printed publications?)

11b. If there is a written plan for future publishing practices, please attach the policy and give the date of the plan and the full name of the issuing office:

Issuing office: _____

Contact person: _____

Date: _____

12. How do you see your role changing in regards to agency publications?

13. Do you see any other major changes to your agencies' publishing policies and practices that may not have been covered by answers to the previous questions?

PART C: “BORN DIGITAL” INFORMATION AND PUBLICATIONS

Definition: “Born Digital” -- publications or information that are both created and continue to reside in an electronic environment. Printing or downloading is done at the initiative and convenience of the user. Information dissemination relies upon computer networks and cyberspace for publication, rather than a physical medium, like a book, microform, or CD-ROM.

1. What types of information does your agency place on its website? What is the website’s purpose/function? Who is its target audience?

2. Is there an overall webmaster for the department?

____ yes (name): _____

____ no

3. Are there any other people in your agency (beyond the webmaster) that we should talk to regarding the agency’s webpages? (please give names and titles)

4. Born digital publishing practices

4a. Describe publishing practices specific to born digital information in your agency (e.g. content selection, publication formats, storage requirements, indexing requirements, public access to offline documents):

4b. If there is a formal policy, please attach the policy and give the date of the policy and the full name of the issuing office:

Issuing office: _____

Contact person: _____

Date: _____

5. Digital publishing formats

5a. When publishing documents digitally, which formats are used? (check all that apply)

____ ASCII

____ HTML

____ XML/SGML

____ JPEG

____ Microsoft Excel

____ Microsoft Word

____ PDF

____ TIFF

____ Other: _____

5b. How does your agency choose which format to use?

5c. Are there any standards for determining format? What are the standards?

5d. Describe any special software requirements needed for accessing the documents

6. How is the public notified of new publications available online?
7. Is the public notified if a publication formerly published in print changes to digital format only?
- yes
- no

Questions 8 and 9 relate to periodicals and one-time (monographic) publications, respectively.

Definition: Periodical -- an ongoing publication that has more than one issue and is produced on a regular basis, such a newsletter, magazine, or annual report.

Definition: Monograph -- a one-time publication, such as a book or report. Such a publication might be updated with a new edition, but the new edition would create another one-time publication.

8. Digital periodicals and annuals

8a. Are back issues of digital periodicals and annuals kept online or do you replace older issues with the current issue?

- keep older issues when adding new issues
- replace older issues when adding new issues

8b. If back issues of periodicals and annuals are kept online, how long will they be available?

- less than one month
- 1-2 months
- 3-6 months
- 6-12 months
- 1-2 years
- 3-5 years
- 5-10 years
- 10+ years
- have no control over how long issues are kept online
- varies (please explain how you decide how long to keep back issues online):

8c. Are older issues stored or deleted when they are taken offline?

- stored
- deleted
- don't know
- varies (please explain how you decide whether to store or delete older issues):

9. Digital monographic (one-time) publications

9a. How long are monographic publications available online (through the web)?

less than one month

1-2 months

3-6 months

6-12 months

1-2 years

3-5 years

5-10 years

10+ years

have no control over how long monographs are kept online

varies: (please explain how you decide the length of time monographs are available online):

9b. If a new version (e.g. new edition) of a monograph is produced and published digitally, what happens to the older version of the monograph?

the older version remains online for _____ (give timeframe)

the older version is taken offline when a new version is available

don't know

varies: (please explain how you decide to keep online or remove older versions):

9c. Are older monographs stored or deleted when taken offline?

stored

deleted

don't know

varies (please explain how you decide whether to store or delete older monographs):

10. Offline storage of publications

10a. If older publications are stored offline, please describe how they are stored (e.g. format and storage media):

10b. Do you have a list of digital publications stored offline?

yes (please attach list)

no

10c. Does the public have any access to older publications stored offline?

yes

no

10d. If the public does have access to offline publications, please explain how the access works:

11. Agency publications produced by private contractors

11a. Are there any digital publications produced by private contractors for your agency?

_____ yes

_____ no

11b. If there are such publications, where are they available on the web?

_____ at the state agency website

_____ at the private contractor's website

_____ depends on the publication

11c. How does your agency decide whether the publications are at the state agency website or the private contractor website?

12. Some agency publications may be produced entirely from the agency's own research/data collecting, while other publications may repackage information/statistical data gathered by another state agency, private research group, or the federal government. What percent of the publications (make your best estimate) are:

_____ % researched and published all within the agency

_____ % published by the agency but using repackaged original research by another state agency

_____ % published by the agency but using repackaged original research by a private research group

_____ % published by the agency but using repackaged original research by the federal government

13. Please describe any major changes you foresee happening to the website in the future:

PART D: DATABASES

1. Are there any databases available to the public through the agency's website from which users can extract information (e.g. directories, statistical information)?

- yes
- no

If yes, please answer questions 2-6, otherwise you have finished the survey.

2. Are there any additional people in your agency we should talk to specifically regarding database management? (please list names and titles)

3. Briefly describe the types of information provided through web-enabled databases:

4. Public Database Indexing Guidelines

4a. Are you familiar with the Public Database Indexing Guidelines of G.S. 132 (public records law)?

- yes
- no

4b. Are the databases indexed according to the guidelines?

- yes, using the NC GILS guidelines, which is the best practice for the statewide technical architecture (please attach documentation or provide a link to it)
- yes, using scheme other than NC GILS (please describe scheme and attach documentation or provide a link to it)
- no (why not?)

5. Updates

5a. How often do you update the information contained in the database?

- continuously
- daily
- weekly
- biweekly
- monthly
- quarterly
- semi-annually
- annually
- other: _____

5b. What happens to older information in the database

- it remains in the databases permanently
- it remains in the databases for a set period of time (list timeframe): _____
- it is overwritten
- it may remain in the databases or be overwritten (explain criteria):

5c. Are users alerted when information is overwritten or added to the database?

- yes
- no

6. Reports

6a. Are publications created by using data from databases (e.g. database reports, digital or print format)?

- yes
- no

6b. If reports are currently published, do you believe the agency will continue to do this in the future or leave report creation up to users manipulating the data?

- continue to publish own reports
- only provide raw data and leave report-making up to the users
- provide both data access and publish own reports

Thank you very much for taking the time to participate in this survey.

Appendix C: Searches for State Government Information

The following five scenarios were invented by library staff, with searches performed in October 2003. Searches done at a later date may bring up different results. The first 10 hits from **Google** and the **State Portal** were examined to see if they brought up the relevant state publication.

1. *A concerned parent has recently moved and would like to know more about her son's new elementary school, Carrboro Elementary, in the Chapel Hill-Carrboro School District. Her neighbor told her that there are "school report cards" on the web. So she tries the following search: "report card carrboro elementary chapel hill."*

The actual document, part of the database, *NC School Report Cards*, is at:

<http://www.ncreportcards.org/src/> (contains information collected by DPI for all public schools in North Carolina. The "Report Card" for Carrboro Elementary School, is available at: <http://www.ncreportcards.org/src/schDetails.jsp?pYear=2001-2002&pLEACode=681&pSchCode=304>

Google search: There are links to other area schools, the incorrect elementary school in Chapel Hill, information on the *ABC's of Public Education* (which publishes a separate report card), and links to Chapel Hill-Carrboro Public Schools. The parent could find more information about the school through Chapel Hill-Carrboro School District site, but there is no direct link to the report cards. There is a link to the communications portion of the *NC School Report Cards* site, but no links to the main part of the site. Following hit number 7, <http://www.welcometothetriangle.com>, there are links to *NC School Report Cards* site.

State Portal search: no sites of use, hits link to newspapers, universities, and C-H Transit.

Why the search is difficult: DPI also has *Report Card for the ABCs of Public Education*, designed to comply with the No Child Left Behind Act, so her search brings up those results along with the more detailed *NC School Report Cards*. The *Report Card* is a database and Google does not index it. So the parent had to find it through links from another site.

2. *A business owner is interested in the trends in the unemployment rate in Raleigh over the past two years. He's considering expanding his business, and wants to get a feel for where the labor market is going. He does a search, "unemployment rate Raleigh"*

The Employment Security Commission publishes these labor force statistics available at: <http://www.ncesc.com/lmi/laborStats/laborStatMain.asp>. The business owner could look at the data for the Raleigh-Durham-Chapel Hill MSA, or for Wake County, in a number of different ways.

Google search: the first hit links to a site which has the unemployment rate for the MSA from 1990 – present, though it is missing the searching capabilities of the ESC database. Other hits link to newspaper articles about the current unemployment rate, or to a Raleigh outside of North Carolina.

State Portal Search: links to newspaper articles, as well as some unemployment rates for other parts of the state, but not for Raleigh

Why this search is difficult: In this case the business owner would have found the information on the first hit, but missed out on the flexibility of the ESC site. Because the ESC site is a database, the dynamically produced pages don't get indexed.

3. *A commuter would like to know the status of the work being done on Interstate 40 in Durham. She searches for "I-40 construction Durham."*

The DOT has the *Travel Information Management System (TIMS)*, a database with information on construction projects, organized geographically. County or route number can be used to look up projects: <http://apps.dot.state.nc.us/tims/>. The database provides information on lane closures, detours, and slowdowns, with ranking for severity and information on when the information was posted.

Google search: there are some links to news sources about the construction, but all are old. Other links are mainly irrelevant, dealing with home construction, pedestrian bridges, and the history of interstates. No hits to DOT.

State Portal Search: some newspaper articles about closures during the construction, otherwise irrelevant material. No hits to DOT.

Why this search is difficult: Again the *TIMS* is a database, so it is not indexed.

4. *A farmer has been having difficulties with nematodes in the roots of his crops. He's not sure what the exact pest is, but he looks on the Internet to see what he can find out, since he thinks the state has some resources available to him. The search terms are "nematode root parasite North Carolina"*

The Dept. of Agriculture and Consumer Services has a section in the Agronomics Division devoted to nematodes. There are publications and also information about how to send a sample in for a nematode assay, to diagnose the problem. The site is available at: <http://www.ncagr.com/agronomi/nemhome.htm>

Google search: perhaps some useful information, but no links to DA&CS

State Portal Search: The Nematode Assay section is the first hit.

Why this search is difficult: For once the State Portal comes through with useful information, but in the Google search, the DA&CS gets buried in a mountain of information.

5. *An individual is interested in opening up a fancy restaurant. She wants to serve alcohol with dinner, so she decides to look on the Internet to find out what she needs to do. She searches using four different terms: (1) “alcohol license North Carolina”; (2) liquor license North Carolina”; (3) “alcohol permit North Carolina”; (4) “liquor permit North Carolina”*

The Alcoholic Beverage Control Commission in North Carolina is responsible for issuing permits. Information about the qualifications for receiving a permit, pricing, and duration is available at: <http://www.ncabc.com/Permits/Retail.asp>.

Google search: searches using terms (1) and (2) bring information about driving while intoxicated, not selling alcohol. However using terms (3) and (4) will bring up the ABC Commission in the first few links.

State Portal Search: only search terms under (2) provide a link in the directory to the ABC Commission. When searching all government sites, no search terms provide hits that link to the ABC Commission. Some hits are about DWI, some about enforcement, and completely off topic.

Why this search is difficult: Because of the synonyms, finding the ABC Commission becomes difficult. If the individual didn't try all combinations, links would have only been about driving and enforcement.

Appendix D: Glossary of Terms

Born digital information: information that is created and disseminated in a digital format without an analog or physical counterpart.

Digital information: information which is stored or transmitted as a sequence of discrete symbols from a finite set, most commonly in binary form, which requires the aid of technology in order to be interpreted by human senses.

Digitized information: Information existing in an analog or physical format that is transformed into a digital format.

Discrete object: object or part of an object that has distinct boundaries, whose meaning and value is mostly self-contained.

Dynamic webpage: webpage that is synthesized at the moment, usually generated from components in a database. It does not exist until called upon by users.

Intangible format: format for information not having physical substance and incapable of being touched.

Integrated object: object with poorly defined or artificially defined boundaries, whose meaning and value is dependent upon its relationship with other objects.

Invisible/Deep web: webpages/information available through the web that search engines cannot/choose not to index often because they are dynamically generated from databases, require user input to access, or use scripts that are deliberately ignored by search engines.

Permanent Public Access: information is preserved for current, continuous, and future public access.

Portal: website considered as an entry point to other websites, often through directories and/or a search feature.

Publication: an object designed to communicate information or notify the public, made publicly available.

Record: recorded information, regardless of format, made or received pursuant to law or ordinance or in connection with the transaction of official business.

Static webpage: webpage defined by fixed HTML code that always appears in the same way.

Tangible format: format for information made up of some physical substance capable of being touched, held, and carried.