North Carolina Department of Cultural Resources

Division of Historical Resources

Archives & Records Section

Government Records Branch

Metadata as a Public Record in North Carolina: Best Practices

Guidelines for Its Retention and Disposition

November 2010

1 INTRODUCTION

1.1 PURPOSE

This document is intended to offer guidance to government employees about the management and retention of metadata under North Carolina's public records laws and state and federal discovery rules for civil litigation. The goal is to provide information and guidance for public employees to understand their responsibilities and liabilities related to metadata.

1.2 DEFINITION OF METADATA

Metadata is structured information that describes, explains, and/or locates an electronic file. Metadata provides answers to questions like "what is it," "where did it come from," and "who created it"?

Different disciplines have developed special definitions for the term "metadata." In the legal community, the term has been defined as "a variety of data associated with electronic documents or files."[1] In the library and archival communities, the term has a slightly different definition—data used to describe the context, content, and structure of materials.

Examples of metadata include a file's name, its author[s], dates of creation, last access or modification dates, and its modification history. Most metadata is automatically generated when a file is created and saved. This includes information that is logged by a computer in order to store and retrieve a file or aid in its searchability. Other metadata is manually generated by the user in order to provide context and information about the file. Metadata can assist a software application by, for example, automatically understanding the language in which a document is typed or through the utilization of cell formulae in spreadsheets.

---

[1] Millonzi, Kara, "Is Metadata a Public Record?: Part 1", posted Thursday, March 4, 2010, accessed July 17, 2010. UNC School of Government, Coats' Cannons: NC Local Government Law Blog, http://sogweb.sog.unc.edu/blogs/localgovt/?p=1984
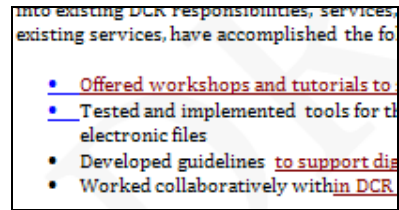
1.3 TYPES OF METADATA (AS IDENTIFIED BY COURTS)

North Carolina courts have not directly addressed issues related to metadata in either the civil discovery context or the public records context. There are several court cases in other states, however, that have addressed litigants' civil discovery requirements related to metadata. And, to date, a few courts in other jurisdictions have analyzed the application of their states' public records law requirements to metadata. These courts generally have catalogued the various types of metadata in one of three categories. Although slightly misaligned with archival and academic categories of metadata, the courts' distinctions may prove legally significant in analyzing how metadata must be treated by litigants in civil litigation and whether or not metadata, or certain types of metadata, constitute a public record and therefore must be managed according to the General Statutes and applicable agency records retention and disposition schedules.[2] The courts have identified the following three categories:

A. Substantive (Application) Metadata

Any changes to an electronic file produce what is called "substantive metadata." Substantive metadata is automatically captured within a file when the file's content is changed after changes or additions are manually made. For instance, when a user employs "track changes" in a Microsoft Word document, the application stores information about those changes as substantive

Substantive metadata is information stored about the content of a file, such as the "track changes" feature in documents.
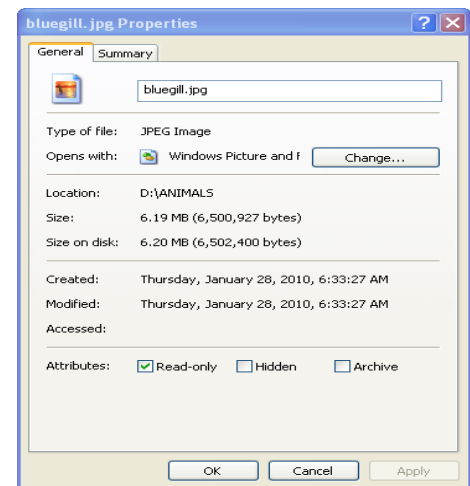
System metadata for a jpeg image file, displayed through properties window of the file.

---

[2] www.stateschedules.ncdcr.gov.

metadata. This metadata remains connected to the file and continues to be added to that file as changes are made throughout its life. Thus, the history of a file's entire existence—from creation to the end of its active life—can be interpreted through the substantive metadata.

Substantive metadata can be thought of as part of the interior structure of a home. Even if you paint over the wallpaper, the history of the house remains beneath it (you can scrape off the paint and still see the wallpaper).
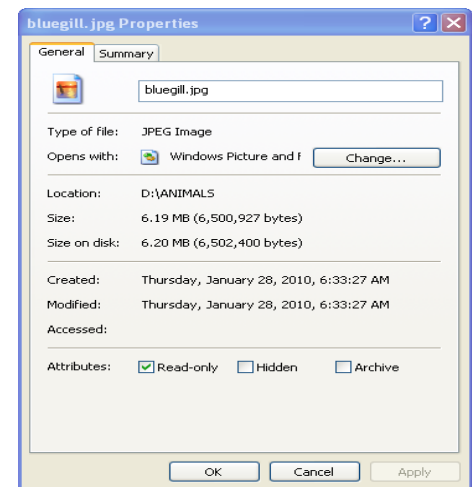
B. System Metadata

System metadata is information about an electronic file that is created automatically by a computer. It is information about the file's structure that the computer needs to locate and render it. Date and time of creation, e-mail header data, file name, file type (the suffix, such as ".doc"), author[s] names or initials, and date of last modification or access are examples of system metadata. Like substantive metadata, system metadata continues to be updated as changes and transfers are made throughout an electronic file's life.

System metadata for a jpeg image file, displayed through properties window of the file

System metadata is more like the frame of the house. The foundation of the house has to "understand" how the frame comes together for it to be sufficiently strong and remain standing. If a significant change is made to the frame, this will likely affect the house's overall structure/foundation. The same is true of system metadata. If a significant change is made to the system metadata, the computer may not know what to do with the file anymore.

For example, if the file type—which identifies the file's basic structure—is changed manually by renaming the file (e.g., ".jpg" is renamed ".gif"), the computer may not know how to render—or open—that file any longer.
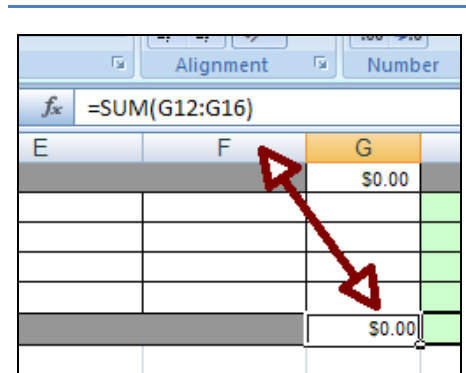
Some system metadata are hidden from normal view, and must be opened or accessed in special ways depending on the type of file.

C. Embedded Metadata

Embedded metadata generally is manually created by the user. Embedded metadata includes information such as cell formulae in spreadsheets, hyperlinks, and some database information. Users can also augment a file's embedded metadata, if they choose, by adding information to the properties portion of some. This metadata may not be visible simply by looking at the file, but it does provide details about the file's functionality, usability, and comprehension.

Embedded metadata is information that is manually created by the user to help identify or clarify the content of a file. An example, in the form of formula data, can be seen above.

While these three categories are fairly clear, there currently are not precise answers as to which metadata categories as a whole are or are not included in the public record in North Carolina. Consequently, the discussion about what type(s) of metadata does or does not need to be retained must remain at the level of the specific metadata element (i.e., cell formulae, file name, date modified, etc.).

2 LEGAL ISSUES OF METADATA

2.1 PUBLIC RECORDS STATUTES

Currently, no North Carolina case law addresses the issue of metadata as public record. However, North Carolina broadly defines a public record, and it is very probable that, because of this broad definition, at least some metadata may constitute a public record in North Carolina. G.S. 132-1 defines public records as including "all . . . electronic data-processing records . . . regardless of physical form or characteristics . . . made or received pursuant to law or ordinance in connection with the transaction of public business."[3]

The value of certain metadata is probably the core issue with regards to public records requirements. Some metadata are crucial for a computer to render a file and for individuals to understand its contents, and therefore should be considered part of the record itself and subject to the public records law. In other words, any metadata

---

[3] http://www.ncleg.net/EnactedLegislation/Statutes/HTML/ByChapter/Chapter_132.html

required or useful in making a public record "complete" may be created, managed, and considered part of the public record. Employees should most likely keep and manage this data in the same manner and length of time they retain the corresponding record.

Examples of <u>potentially essential</u> (required) metadata include:

- E-mail header information which details the path a message takes. This metadata is usually hidden from a user's view and does not necessarily print when e-mail is printed.
- File creator name
- Date created
- Title (stored as the file name)[4]
- Cell formulae for spreadsheets created with spreadsheet software where the number in the cell is generated by a formula that is typically not displayed, e.g. spreadsheets generated in Microsoft Excel™.

Information that is considered integral to providing authenticity and context, such as e-mail headers, will likely be considered substantive to the record. Whereas, information that is not useful in providing authenticity and context, such as track changes will likely not be considered substantive to the record. Following this logic, employees can make a few assumptions about required metadata. For instance, the following information may be integral to providing authenticity and context for the record, and thus may need to be created and retained:

- Creator name
- Creation date
- Title (stored as the file name)
- Cell formulae for spreadsheets

The following pieces of information <u>may</u> be nonessential to the record; thus, they may not need to be retained:

- Date of last modification
- Track changes
- Other substantive metadata

---

[4] Please refer to Best Practices for File-Naming on North Carolina Government Records Web site, http://www.records.ncdcr.gov/erecords/default.htm#guide

Employees should consult the General Schedule for State Agency Records or their agency records retention and disposition schedules for further guidance regarding potential retention of this type of metadata.[5] Employees may also wish to consult David Lawrence's *Public Records Law for North Carolina Local Governments.*[6]

Just as public records in North Carolina fall under the Department of Cultural Resources (NCDCR) retention and disposition requirements, essential metadata connected with a public record are subject to the same requirements and should be managed accordingly. Guidance can be found in the *General Schedule for State Agency Records at* http://www.records.ncdcr.gov/schedules/GS_2009_updateamendment_20090831B.pdf. North Carolina employees should maintain essential metadata (as described herein) as part of the public record. It is important to note that metadata does not typically print and is sometimes "hidden", thus making it more challenging in terms of management and production. This means that employees should not necessarily consider printing and interfiling documents as a retention technique.

✔ **IMPORTANT: These recommendations are not meant to direct public employees to create information or public records they would not normally create. Creating and maintaining all metadata is not feasible, useful, or advised. Employees should commit to maintaining only the metadata that is essential for a file's current use and/or retention.**

This is particularly important because any metadata—including modification date, track changes, and other substantive metadata—may be subject to e-discovery. Any public employee may be required to turn over any or all files (along with the related metadata) if requested by the courts during a legal dispute.

The challenge lies in the fact that this area of law is still developing on a case-by-case and context-by-context basis. Therefore, this document, and the guidelines provided herein, may change at any time as the courts provide more specific guidance.

---

[5] Please refer to the General Schedule for State Agency Records on the North Carolina Government Records Web site at http://www.records.ncdcr.gov/stateagy.htm#gs Item GS 130, Data Documentation File.

[6] Lawrence, David M., *Public Records Law for North Carolina Local Governments,* 2009.

2.2 BEST PRACTICES FOR RECORDKEEPING

Good electronic recordkeeping requires the work of the employees who create, name, and manage files as well as knowledgeable information technology employees. Often the devices on which files are managed are maintained and managed by IT employees. Their job is to manage the hardware, maintain it, back it up, and move files when new hardware needs to be installed. Electronic files, when moved from one storage device to another, inherit the date when they are moved or copied to another storage device. In order to ensure that critical information is not lost, IT professionals should ensure that critical information is not lost when data is transferred. Employees can also help in this effort. Good file naming can help to ensure that creation date is not lost by making the date part of the file name. Also, employees should choose file formats that are supported and documented. For more guidance on formats and file naming, please consult Electronic Records Web page at
http://www.records.ncdcr.gov/erecords/default.htm

File Naming Tips:
- Use logical names in your schema
- Include the date of the record in the file name
- Do not leave empty spaces in file names. If needed, use an underscore between the words, e.g. file_naming_schema_201008

2.3 ELECTRONIC DISCOVERY REQUIREMENTS

"Discovery" is a term that is used to describe the period before a trial in a civil lawsuit when each party may request documents and other evidence from the other. Electronic Discovery ("E-discovery") is a term that describes the electronic discovery process—the gathering of electronic information from each of the parties in a lawsuit. When responding to e-discovery requests, employees should consult the attorney for his or her state agency or local government.

For more information, the University of North Carolina's School of Government provides guidance and discussion at http://sogweb.sog.unc.edu/blogs/localgovt/?s=metadata. For agencies dealing with litigation or that may experience future litigation that involves an e-discovery request, an excellent resource that provides a detailed analysis of this

topic is the School of Government document called "Electronic Discovery in North Carolina."[7]

3 SUMMARY

Neither North Carolina statutes nor her courts have specifically addressed the issues of and requirements for metadata. Therefore, this document serves as a best practices approach to providing for North Carolina employees. Employees should understand their responsibility and liability under G.S. 132[8] regarding metadata and its subsequent retention, management, and potential delivery.

At a minimum, employees should input and retain the following four fields of metadata:

- Creator name
- Creation date
- Title (stored as the file name)[9]
- Cell formulae for spreadsheets

As more information on metadata retention becomes available, public employees in North Carolina will receive notice from the Department of Cultural Resources. Any questions should be directed to agency legal counsel and Cultural Resources staff.

The Department of Cultural Resources offers guidance on electronic records and digital formats. The Government Records Branch of the North Carolina State Archives offers workshops and online tutorials on records of all types—paper, electronic, e-mail, and imaging. To learn more or sign up for a class, please visit www.records.ndcr.gov. In addition, the State Library of North Carolina offers guidance. Please visit http://www.digital.ncdcr.gov/cdm4/about.php

---

[7] http://www.sog.unc.edu/pubs/electronicversions/pdfs/ediscovery09.pdf

[8] http://www.ncleg.net/EnactedLegislation/Statutes/HTML/ByChapter/Chapter_132.html

[9] Please refer to Best Practices for File-Naming on North Carolina Government Records Web site, http://www.records.ncdcr.gov/erecords/default.htm#guide