METADATA: WHAT IT IS AND WHY YOU SHOULD CARE

by Johnette Hassell, Ph.D., CEDS and Jack Molisani

Until Edward Snowden unleashed his allegations about the US and UK collecting phone information on millions of their citizens, the word metadata was the providence of attorneys and computer forensic/eDiscovery nerds, such as these authors. And while the world may be aware of the term, few truly understand the breadth and pervasiveness of computer metadata.

In this article we will discuss what computer metadata is, explain its importance in investigations and litigation, and provide a variety of examples.

ABOUT ELECTRONICALLY STORED INFORMATION

When we discuss metadata in general and metadata as evidence in a lawsuit in particular, we are discussing what is generally called Electronically Stored Information, ESI. The need to collect, process and produce ESI has caused substantial changes in the way ESI is handled when compared to more tangible kinds of evidence. In particular, ESI must be handled in ways that preserve and protect metadata.

WHAT IS METADATA?

When electronic devices store information, the files used normally contain the information itself (such as a digital photograph) plus additional information about what is stored in the file. For example, the time a photo was taken and other information is typically stored along with the actual photo.

This addition information is called *metadata*, because it is data *about* the data.

Word processing documents may contain information about the last edit, as Word Perfect does, or about the username of the document's creator, as Microsoft Word does.

Metadata, however, is not limited to files in a computer or camera. The US and the UK say they weren't col-

lecting or storing the actual telephone calls, only the metadata about the calls. (You see such metadata each month when you read your phone bill: the numbers you called and how long each call lasted.) If your call was made with a smartphone, the metadata probably also contains the location from where you made the call.

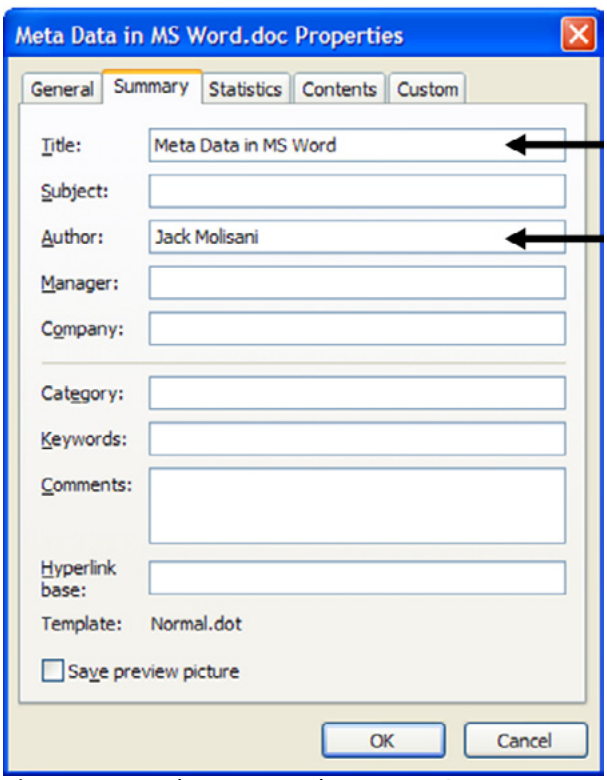

Figure 1. *Metadata in a Word 2003 2007 Document*

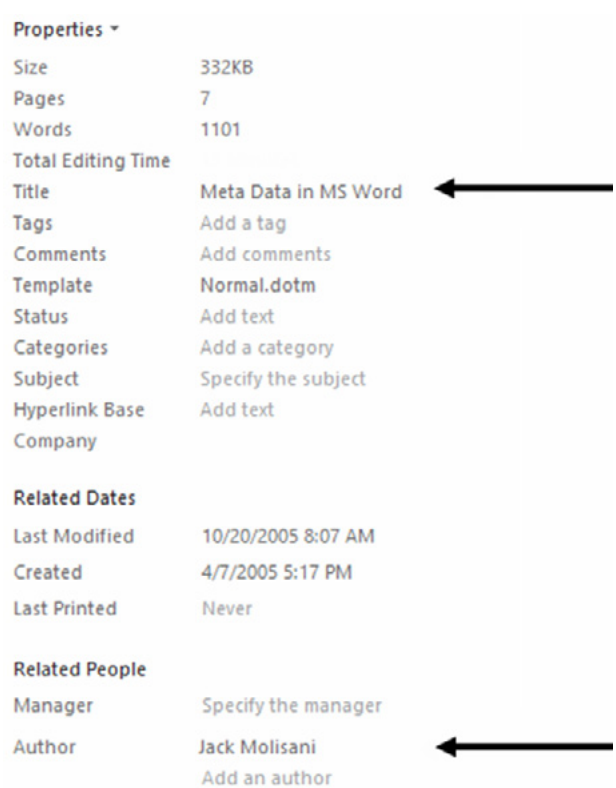

Figure 2. *Metadata in Word 2010 Document*

COMMON TYPES OF METADATA

While governments requesting information about its citizens' phone usage certainly provides a high-profile example of how metadata can be used (or misused), let's look at two common types of metadata: the metadata in office documents (such as MS Word and Excel files) and digital photographs.

TYPICAL MICROSOFT DOCUMENTS

You may know that metadata in documents contain easy-to-see information such as the name of the author, the company name, and certain dates. We say "easy to see" because you can see and even change that information from within the program.

To see a simple example of this information, open a Microsoft Word 2003 or 2007 document and select *Properties* from the *File* menu. A dialog similar to Figure 1 will appear showing some of this information, such as the document Title and Author.

For Microsoft Word 2010, select the Info tab on the File menu: Figure 2.

A document created on a corporate PC might display more information, such as the company name and the name of a corporate template (if any). See Figure 2 for a typical example.

While you may have known you can change what appears in the Author field, you may not know that the metadata often includes hidden information, such as the name of *previous* authors who edited the document and the names of the printers used to print the document.

To see the remainder of the metadata stored in a Word file:

- In Microsoft Word, select *Open... from the File menu*.
- From the *Files of type* drop-down list, *select Recover Text from Any File (*.*)* and then select and open a Word document, as seen in Figure 3.
- When the file opens, page down to the bottom of the file to see metadata such as the following (what you see will vary): Figure 4.

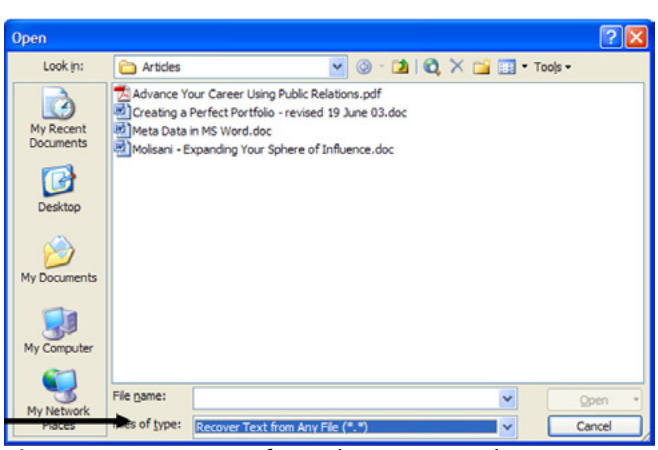

Figure 3. *Using Microsoft Word to View Metadata*

In Figure 4, above, you can see the name the document originally had ("Administrative details 305 198.doc") and where it was located (on a machine with user name "Johnette Hassell").

Figure 5 shows this document was then saved under a new name ("Administrative details 305. doc") in a folder on a different computer ("E:\cs305. fall.01" on the computer named "hassell"):

There is more information you can recover, but this gives a good example of the type of data Microsoft Word stores. Such information might be critical evidence in a lawsuit, where the metadata might show how an accused party saved a company's design document to an external hard drive, edited it on a home PC, then edited it again on a computer at his/her new (and competing) employer.

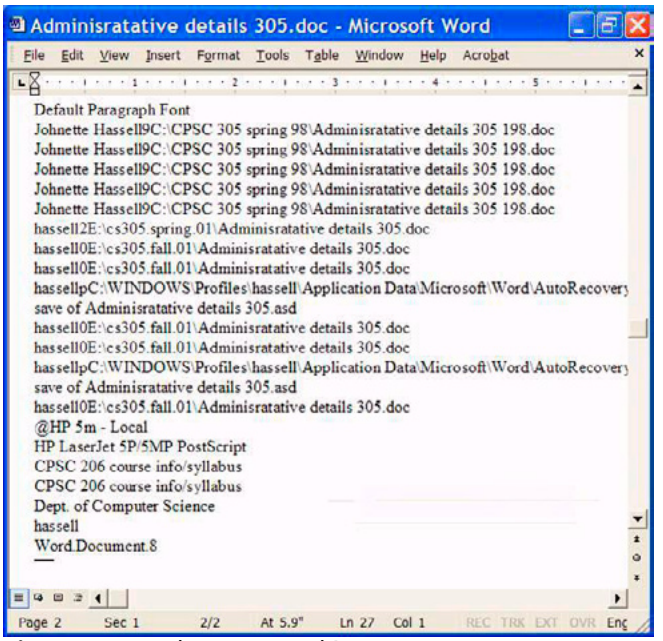

Figure 4. *Metadata in a Word Document*

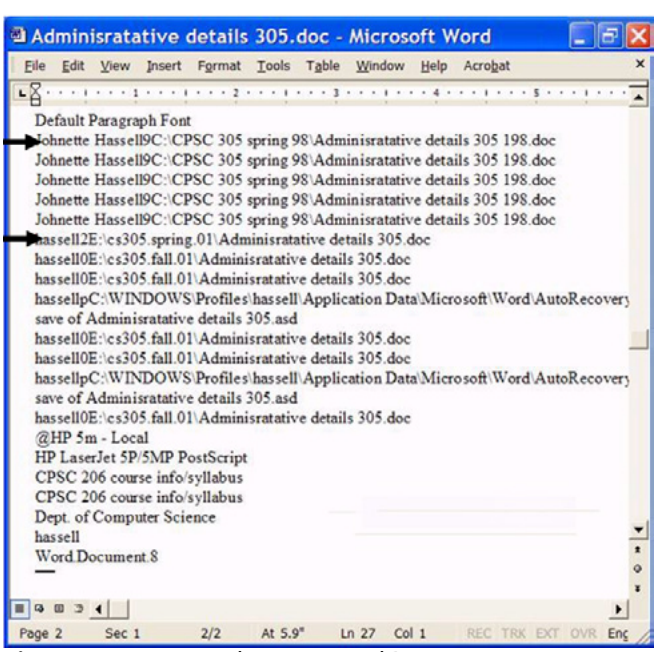

Figure 5. *More Metadata in a Word Document*

DIGITAL PHOTOGRAPHS

Digital cameras (including those on smartphones) record metadata such as the date and time a photo was made. This data is recorded even if the photographer has turned off "Show Dates" in the photo. These time stamps may be useful, for example, to police who performed a drug bust, where they need to show the exact time of the raid and seizure.

COPY MACHINES

Few people know that modern copy machines/ printers work by making an image of a page and then printing the image. [1, 2] Fewer yet know that such machines often *retain copies of recently printed documents on an internal hard drive*, including the date and time each document was submitted, the username who printed the document, and the computer from which the document was sent. These types of metadata are useful in both trade secret cases (in which an employee is accused of theft) and in espionage cases (showing who stole classified documents).

CBS News, in preparation for an investigative report on copy machines, bought 4 used copy machines. In examining their hard drives, they found a list of targets in a major drug raid (from the Buffalo Police Narcotics Unit), 95 pages of pay stubs with names, addresses and social security numbers and $40,000 in copied checks (from a New York construction company), and 300 pages of individual medical records, included everything from drug prescriptions, to blood test results, to a cancer diagnosis (from Affinity Health Plan, a New York insurance company). [3]

SMARTPHONES

The photographs taken by smartphones may have additional metadata, such as the location where the photograph was made. Just ask Highinio Ochoa. He was a hacker known as "w0rmer" [sic] and worked with a hacking group "CabinCr3s." He was, allegedly, responsible for releasing the personal information of scores of police officers throughout the United States.

The FBI found him because he posted a photo of his scantily-clad girlfriend. The photo (made with a smartphone) contained the GPS coordinates of where it was made. These coordinates led to his girlfriend's location, and, eventually to him. [4, 5]

HOW COMPUTERS USE METADATA

Have you ever tried to open an email attachment and received an error message saying the computer doesn't know what program to use to open the file, or perhaps your system tried to open the file and says the file didn't contain what it expected to see?

Computers (such as PCs running MS Windows) "know" what type of information should be in a file two ways: the ".xxx" ending on the file name called the file's extension, and an internal code within the file.

For example, a document stored in Adobe Acrobat format normally ends in ".pdf" (portable document format), such as "MyDocument.pdf" (Figure 6).

Windows uses the extension .pdf to know what program to use to read the file (in this case, Adobe Acrobat.)

If you were to look at that same file using a simple text editor (like Notepad), you would see that the very first characters in the file are "%PDF". This file *signature* identifies that the type of information stored in the file, in this case "PDF"). See Figure 7 for an example of such a signature.

A person can, however, change the file extension in an attempt to hide something from plain view. For example, a spy might rename a spreadsheet recording a list of bribes from "MyBribeList.xls" (Excel spreadsheet) to "My home movie.mov" (a movie format).

A person who tries to open this "movie" will get a message saying the file cannot be opened. While an observer may assume such a file contains a movie based on the .MOV extension, modern forensic tools can indicate when a file extension does not match what the file really contains, as identified by the file signature. (Renaming the file extension in an attempt to mask what's inside the file is a technique child pornographers frequently use in an attempt to hide illegal photographs.)

GUARDING AND PRESERVING METADATA

Now that you knowing about metadata, what should do?

SHARING YOUR METADATA?

First, exercise caution when sharing any documents you work with, especially when sharing them with people outside your organization. Can you remove metadata or otherwise protect it from prying eyes? There are tools that can do this, to a certain extent. [6] But don't forget there is also imbedded metadata, data that is harder to change.

PRESERVE METADATA

If you are an IT professional and your company is involved in a lawsuit (or even if you *think* your company *might* be involved in a lawsuit), you must take steps to ensure the metadata of ESI in your control is not altered. You may also need to set up mechanisms so that other employees can set aside ESI that need to be preserved. Merely turning on a computer makes more than 160 changes to a computer's hard drive(s). Many of those changes are to the dates on files, dates that may be crucial to a case. Cell phones, once they are turned on, make continuous changes to their storage areas. If your organization is faced with litigation, immediately consult with your corporate attorney and a reputable eDiscovery or computer forensic specialist about the best way to preserve all your ESI, including metadata.

If you are an attorney in litigation, be aware of metadata in your client's productions and include metadata in your requests for production. The federal rules of discovery are clear about metadata, but state rules may vary. See the Kroll Ontrack, [7] and K&L Gates [8] websites for up to date information on individual states' rules.

Metadata may have much to tell someone interested in your business. One real estate attorney handled lucrative casino properties. Many of his clients did not want others to know of their interest in such properties. Unfortunately, the attorney used a boilerplate proposal document, repeatedly saving it under different clients' names. The metadata revealed the names of interested parties going back several years; and many of those clients were competitors.

SEARCH FOR METADATA

Ordinary search tools, such as Windows' and Google's search features, do not recognize nor search the "hidden" metadata in files. There are, however, a limited number of tools that allow people to examine the metadata in files. But if you need to search a large collection of electronically stored information (as is often necessary in litigation), use a certified eDiscovery consultant who can help find potential evidence that might be in the metadata.

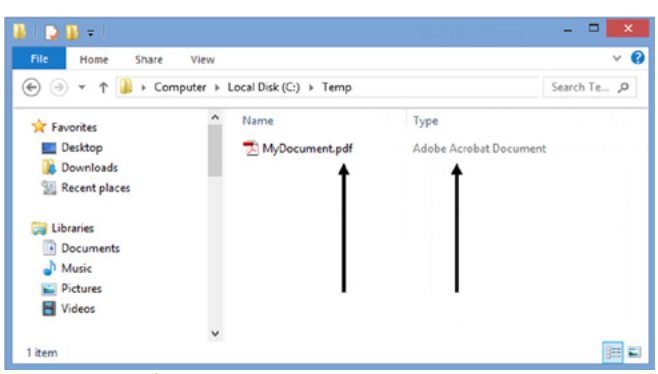

Figure 6. *.pdf Document Extension*

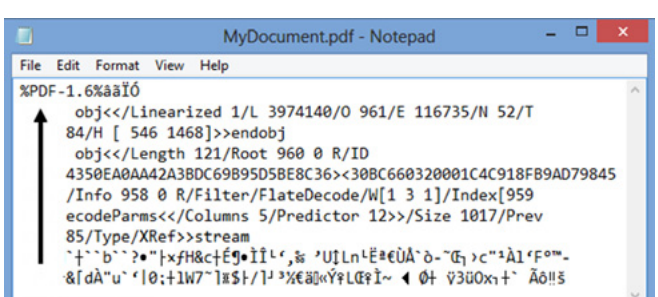

Figure 7. *.pdf Document Signature*

SEARCH FOR ANOMALIES

There are tools available that can alter metadata such as the time a photograph was taken or the date a file was last modified. If you are involved in a lawsuit where dates are important, an eDiscovery specialist can use forensic tools to determine if relevant data, including metadata, were changed (purposely or accidentally).

MAINTAIN DIGITAL CHAIN OF CUSTODY

Since ESI is easily changed by even simple, innocent acts such as opening a file or booting a computer, special care is needed in managing ESI. Preserving the original media (such as a memory card from a camera, the hard drive(s) from a computer, or the files in a smartphone) is the best way to preserve data. The processes of insuring the integrity of potential evidence is known as *maintaining chain of custody*.

Other than the original media itself, currently, the best way to preserve electronic media is for a forensic specialist to make a *valid forensic image* [9], an exact bit-by-bit copy of the item in question. Such images preserve everything on the media (including all metadata) and are regularly accepted in court proceedings as valid evidence. There are numerous tools for making such images. Using appropriate tools, these images can be examined without worry about changing the original evidence.

THE POWER OF METADATA

Information stored in metadata can make or break a case if your company is ever sued (or, in turn, if your company needs to sue a competitor). IT departments are usually the first to be contacted internally when litigation is known or contemplated. Uninformed handling of potential evidence may inadvertently lose or modify important data, including metadata.

Be aware of metadata: what it is, where it is, how to preserve it, when to (and when not to) delete it.

A law suit can be won or lost on metadata alone. Use it to your advantage.

REFERENCES

[1] *http://bucks.blogs.nytimes.com/2010/06/01/why-photocopiers-have-hard-drives/?_r=0.*

[2] *http://bucks.blogs.nytimes.com/2010/05/20/the-identity-theft-threat-from-copiers/?scp=1&sq=copier&st=cse.*

[3] *http://www.cbsnews.com/stories/2010/04/19/eveningnews/main6412439.shtml.*

[4] *http://gizmodo.com/5901430/these-breasts-nailed-anonymous-hacker-in-fbi-case.*

[5] *http://www.dailymail.co.uk/news/article-2129257/Higinio-O-Ochoa-III-FBI-led-Anonymous-hacker-girlfriend-posts-picture-breasts-online.html.*

[6] See *http://en.wikipedia.org/wiki/Metadata_removal_tool.*

[7] *http://www.krollontrack.com/resource-library/rules-and-statutes/.* (Double click on desired state.).

[8] *http://www.ediscoverylaw.com/promo/state-district-court-rules/.*

[9] Demystifying Computer Forensics, Louisiana State Bar Association, J. Hassell, Ph.D. and S. Steen, December 1999.

ABOUT THE AUTHORS

Dr. Johnette Hassell has 30 years experience in computer-related litigation support. A retired computer science professor, she is a court-recognized expert in computer forensics, eDiscovery, computer science, and data recovery. She is a Certified eDiscovery Specialist, and serves on the Association of Certified eDiscovery Specialists (ACEDS) exam and exam preparation committees. Her work is published in law and technical journals and she is a highly sought-after lecturer in CLE courses. As President and CEO of Electronic Evidence Retrieval, Dr. Hassell provides consulting services ranging from early case assessment through testimony: http://www.ElectronicEvidenceRetrieval.com.

Jack Molisani is a Computer Engineer with almost 30 years experience in software engineering, technical communication, and eDiscovery/computer forensics. He is a Fellow of the Society for Technical Communication and the Executive Director of The LavaCon Conference on Content Strategy: https://lavacon.org.