

# Document Metadata and Computer Forensics

James Madison University Infosec Techreport  
Department of Computer Science

JMU-INFOSEC-TR-2006-003

Jeffrey R. Jones

August 2006

# Document Metadata and Computer Forensics

Spring 2006 Term Paper for CS633 (Computer Forensics)

Jeffrey R. Jones  
INFOSEC Master's Program  
James Madison University

jonesjr@jmu.edu

## ABSTRACT

Metadata contained within documents serves a valid purpose in many circumstances, such as facilitating the collaboration among a group of people. However, many are not aware of the type of information stored with their documents, spreadsheets, and presentations. Due diligence is required by responsible users to ensure that sensitive information is not leaked to third-parties. Until then, forensic investigators could have access to a plethora of hidden document information. This paper examines how metadata is used in PDF documents and documents, spreadsheets, and presentations created in Microsoft Office and OpenOffice.org. Several instances are examined where metadata has led to the discovery of hidden information. This paper also shows how metadata is stored in documents, spreadsheets, and presentations created in the aforementioned applications. Finally, this paper will test and discuss the functionality of several tools available to users and investigators that test for the presence of metadata.

## 1. INTRODUCTION

Metadata is a structured description of objects, which contains certain properties useful to the user as well as the program on which the document was created. More succinctly, metadata is data that describes other data. It is used in office applications to store various types of data ranging from the author's name to the last time the file was printed. Metadata, from its definition, may seem safe to store within the document; however, it can be an incriminating piece of evidence against the document's author, owner, reviewer, publisher, and may provide information regarding the network storage location as well as the unique identifier of the machine on which the document, spreadsheet or presentation was created.

Metadata can reveal sensitive information not deemed for public use, which can be damaging to a company or individual, but at the same time provide a wealth of useful information for a forensic investigator. For instance, an investigator investigating a case of inappropriate material may be able to track down the original owner of the file by simply viewing its metadata.

Forensic analysis of computers and the data contained therein can sometimes be an extremely difficult and time-consuming task. The complexities involved in investigating computer crimes can sometimes cause an investigator or team of investigators to overlook small details that could at least lead the examiner down the right path, if not answer questions in and of themselves.

The amateur and professional forensic analysis of document metadata has been more widely publicized in previous years than at any time in the short history of the computer revolution; specific examples of which are provided in a later

section. In addition, its use as part of the forensic process is included in many computer forensic educational programs throughout the nation. Analyzing document metadata is even briefly discussed in the 2004 edition of the *Forensic Examination of Digital Evidence: A Guide for Law Enforcement* [1].

The forensic analysis of metadata is also being used more extensively in legal cases. For instance, in *Jackson v. Microsoft Corp.*, the defense was able to show that the ex-employee copied, stored, and duplicated proprietary information based in some part off of the date and time stamps stored in the file's metadata. This information was presented to the court, which subsequently dismissed the suit against Microsoft, because of the discovery that the plaintiff misrepresented his actions [2].

The court system as a whole recognizes the usefulness of metadata and its viability and admissibility as evidence. For instance, in 2005, a judge ruled that "when a party is ordered to produce electronic documents ... the producing party should produce the electronic documents with their metadata intact" [3] during a lawsuit against Sprint.

Investigators can use the hidden attributes and information stored with documents to aid in the overall forensic analysis process. This paper will address one small portion of the document forensics process and focus on the discovery, analysis, and usefulness of metadata saved with document, presentation, and spreadsheet applications in wide-use today.

## 2. WHAT IS METADATA?

As mentioned previously, metadata can simply be described as data that describes other data. This is a simplistic definition for a term that is much more complex than at first sight. Metadata has been used for many years in a range of applications. One example of the use of metadata is in web pages on the Internet. Metadata embedded in HTML documents allow a search engine to catalog the topics addressed at the specific site, thereby increasing the effectiveness of search engines in hopes of increasing the traffic to the metadata user's website.

Metadata is also used by relational databases in order to describe the tables, columns, and fields. In this case, the database metadata is referred to as the catalog of the database. In addition, it is used by newer filesystems to store information about the data that is saved on the disk. For instance, the Linux ext3 filesystem logs all filesystem data and metadata changes in a journal. Microsoft filesystems, such as NTFS, also store this type of information. In addition, older filesystems, such as Linux ext2 and Windows FAT16, store metadata file information such as timestamps, size, and data location. Ext2 also has the capability to store permissions information for owners and groups.

Metadata is used extensively by Microsoft Office products, OpenOffice.org applications, and portable document format (PDF) documents in order to store information such as the document's creator, date of creation, date and last time it was saved, the location the document to which the document was saved, as well as contributors to the document.

Metadata is very useful at times and helps to facilitate the collaboration amongst many people. It allows the application, in this case, to keep track of changes and comments made to a document by reviewers. The document author can choose to accept or decline the reviewer recommended document changes and view any comments offered.

A problem with the use of document metadata is that many users and document authors are not well versed in what information is stored with their documents as they share and finally distribute them. Unless specific measures are taken to remove the document's metadata, every time a document, spreadsheet, or presentation is created in Microsoft Office, or OpenOffice.org, as well as PDF documents, metadata is saved alongside the text created by the author.

This type of information could inadvertently disclose sensitive information as well as prove very useful to a computer forensic investigator. Several years ago, metadata was saved with documents without the consent of the creator; however, recent developments and a more concentrated focus on security has lead to the user having a choice of whether to save some of the suspect information with their documents, but only if the user is savvy enough to know about metadata.

### **3. METADATA APPLIED**

This section addresses the specific functional features of documents, spreadsheets, and presentations created in Microsoft Office and OpenOffice.org, as well as PDF documents. In addition, this section will break down some of the risks associated with metadata and to the user and benefits of metadata for the forensic examiner. Finally, this section will also provide examples that have lead to some embarrassing discoveries.

The problem with metadata is not necessarily with the metadata itself. As mentioned previously, metadata has useful purposes; however the problem occurs when users are careless or uninformed. After one highly publicized instance of metadata revealing information not intended for public use, one individual stated that "[t]he real scandal is that after 15 years of using Microsoft Word, I don't know how to turn off 'track changes' [4]."

This problem is widespread and continues to attract much attention. For instance, Simon Byer researched the problem and pervasiveness of metadata in Microsoft Office documents. Through an exhaustive examination of over 100,000 documents, he found that about half of the documents contained between 10 and 50 hidden words, a third of the documents had somewhere between 50 and 500 words, and approximately 10 percent had more than 500 hidden words contained within the document [5].

#### **3.1 Risks (User) / Benefits (Examiner)**

All metadata stored with documents, spreadsheets, and presentations can reveal information about the document that is beneficial to the investigator. However, there are features and capabilities of Microsoft Office, OpenOffice.org, and PDF documents that can more so pose problems for a user but assist a forensic investigator in the performance of his or her duties. The Track Changes feature of Microsoft Office and

OpenOffice.org will be highlighted in this section as well as the Commenting feature, which is common to Microsoft Office, OpenOffice.org, and PDF documents.

Two additional features of Microsoft Office that can provide useful information are the metadata stored with Macros and Fast-Saves. In addition, older versions of Microsoft Office stored computer specific information with its metadata that deserves mentioning, since it revealed much more user-specific information. PDF documents also contain metadata, much of which was stored with the original document, but can also store additional incriminating information.

##### *3.1.1 Track Changes and Commenting*

Microsoft Office and OpenOffice.org have Track Changes and commenting features that are a valuable addition when used properly and diligently. Neither feature is enabled by default in either of the products; however, its use by document creators and reviewers can reveal more information that possibly desired.

The Track Changes feature provides a facility to view the history of all the changes made to a document. If this feature is left enabled by the user, all viewers of the document will have the capability to view all changes made since the last round of changes were accepted by the document creator [6].

Both products' commenting mechanism likewise can reveal the same type of information. It allows users to make comments to the document without changing any of the underlying data content. However, if those comments are left in, each subsequent viewer can review all the comments made by document reviewers. Of greater concern with this feature is that Excel and PowerPoint do not warn a user of the comments that are embedded in a document [6].

When using the Track Changes and commenting features, the name of the reviewer is also stored with any comment or change made to the document. Needless to say, this can be highly undesirable to publicly release this information, as a company may not wish others to know every person who has reviewed the document for release.

As mentioned previously, the document properties can reveal a tremendous amount of information. In both office products addressed in this paper, part of the metadata saved with the document includes the name of the document creator. If this document is used again as a template for another document, the creator's name may stay the same; however, the author may not. In this case, a published document may come from one organization, but in reality originated from another totally separate entity.

##### *3.1.2 Macros and Fast-Saves*

Macros used within office products can be very useful and increase productivity in many situations. However, Microsoft Office adds an additional bit of information to any macro used inside of the document, spreadsheet, or presentation – the name of the macro author [7]. This is a goldmine for the forensic examiner investigating macro viruses that flood the Internet at an alarming rate.

Microsoft Office also has the capability to perform fast-saves, which is a convenient way for the user to ensure that if the computer crashes, a recent backup is just a click away. This can be a wealth of information for the investigator. If the user types up a document, while fast-saves are enabled, and deletes some text, that deleted text is likely to stay with the document throughout its lifetime. Some text is just never deleted [5, 6].

##### *3.1.3 Computer Specific Information*

In Microsoft Office 97, Word, Excel, and PowerPoint stored a unique identifier that could identify the system and installation on which the document, spreadsheet, or presentation was created. Microsoft's reasoning for including this information in the file was to facilitate the cooperation with third-party programs [8]. Microsoft further stated that the Globally Unique Identifier (GUID) number, which has been defined in RFC 4122 [9], could not be used to "identify the author of a document without intimate knowledge of the personal computer on which the document was originally created." However, the method used to create the GUID did not completely coincide with Microsoft's statements and proved valuable during the search for the author of the Melissa virus.

What makes the GUID number so revealing is that it is composed of the computer's Ethernet Media Access Control (MAC) address. There are no two MACs alike and the storage of this type of information could lead investigators with supporting evidence directly to the computer with which the document was created. The major drawback in using this identifier in forensic investigations is that the GUID is only stored once when a document, spreadsheet, or presentation is created. Therefore, if the document was modified by another user, that new user's GUID would not be stored in the document and would only contain that of the original document creator [10].

### 3.1.4 PDF Documents

PDF documents are widely used as a method of distributing documents in a common format readable cross-platform. As such, the capabilities and features of this document format also present a forensic challenge to the examiner and risk to the user. Adobe has a plugin that is designed to work seamlessly with Microsoft Office and OpenOffice.org products. This seamless functionality is of a great benefit to the user, but can also pose a challenge regarding stored metadata.

When a document is converted to PDF format, by default, it is enabled to also store all of the metadata that was stored with the original document, such as version information, the document's creator, as well as change tracking. All of the risks mentioned above regarding the Track Changes and commenting features are also an inherent risk to PDF documents. Adobe also has a commenting feature of its own that can also add to the amount of wealth of metadata stored with the documents.

One additional feature, not enabled by default, when interacting with Microsoft Word is the ability to store the source document inside of the PDF file. It is easy to see that the benefits gained even if the user chose not to copy over to source document's metadata when performing the file conversion would be negated if the user did not understand the capabilities and risks associated with storing the source document with the PDF document.

## 3.2 Metadata – Embarrassing Moments

Metadata can prove to be a useful medium in helping a forensic investigator meet his or her objective while, at the same time, leading to regrettable situations for document creators and reviewers. This section will focus on the more famous discoveries made through the exploitation of document metadata.

### 3.2.1 Merck

The pharmaceutical company Merck has been involved in several lawsuits related to its arthritis drug medication. In December 2005, the *New England Journal of Medicine*

discovered that important information regarding the medication's risk of causing heart attacks was removed from a document sent by Merck. Merck apparently had used the Track Changes feature while preparing the informational paper to be released to the *Journal*. The authors at Merck forgot to accept the changes made to the document and subsequently released it with all the metadata intact. Upon closer investigation, the *Journal* discovered the deletion and subsequently released the information about the Merck's gaffe [11].

### 3.2.2 Democratic National Committee

A document surfaced from the DNC in late October 2005 with some not so flattering things to say about Supreme Court nominee Judge Samuel Alito. After examination of the document's metadata, a few of the author's names were revealed as well as the creation date of the document. More specifically, two userids were discovered and equated to two members of the DNC and the document was created right after Justice O'Connor resigned, meaning that the document was created before Judge Alito was nominated [5].

### 3.2.3 United Nations

In October 2005, the United Nations (UN) released an unedited report regarding Syria's involvement in the assassination of the former Prime Minister of Lebanon. A change was made to the report after it was given to the UN Chief, but it was still visible when an inquiring mind checked the Track Changes function on the document.

Someone changed those responsible for devising the assassination from the brother and brother-in-law of the Syrian President, as well as other high ranking officials, to a generic "senior Lebanese and Syrian officials" statement. The inadvertent disclosure of this information occurred only because the document authors were not careful in checking the types of metadata contained within the document before release [4, 12].

### 3.2.4 SCO Group

Another example of the inadvertent disclosure occurred in 2004 when the SCO group filed a lawsuit against DaimlerChrysler and AutoZone. Once again, the text of the lawsuit had been created using Microsoft Word with Track Changes enabled. The author did not take the steps necessary to remove all the changes and comments before the document was released. It was subsequently discovered that the document was originally prepared with Bank of America as the defendant [4, 13].

### 3.2.4 The "Dodgy Dossier"

One of the more famous faux-pas that has taken place in recent years was brought to light in 2003. The United Kingdom courts summoned documents from its government regarding Iraq's weapons of mass destruction program. Among the documents provided to the courts was one that appeared to be genuine and made a case for Hussein's possession of these weapons. However, upon further examination, it was discovered a large portion of the document was actually taken from a twelve year old PhD thesis. Four of the document authors' names were discovered, putting them under the spotlight in order to ascertain how this information was compiled [14].

### 3.2.5 The Melissa Virus

On March 26, 1999, the Melissa virus was unleashed on the Internet and spread with amazing speed. Millions of users were affected and companies estimated their losses in the millions of dollars. The federal government looked to find

and prosecute the virus's creator because of its widespread impact. This was not an easy task but was aided by two independent persons who discovered some damaging information [15, 16].

These persons looked at the metadata inside of the document containing the virus and discovered the GUID. The investigators conducted further research and finally located a website belonging to a malicious hacker, on which other documents with the same GUID were discovered. While this could not be used in and of itself to pinpoint one individual as the creator, combined with other information, it can definitely help the investigation. In the end, the discovered information was passed to ZDNet, who published several articles, and the federal authorities. This information assisted in efforts to find and convict David Smith of Aberdeen, N.J., for creating the virus.

#### **4. MICROSOFT OFFICE METADATA**

Microsoft Office is the most dominant office product in use today. Because of this fact, the metadata that is used by its component products, Word, Excel, and PowerPoint, has been increasingly scrutinized in previous years and extensively addressed by Microsoft.

According to Microsoft's Knowledge Base [17, 18, 19, 20], the following data can be saved as hidden information inside of Microsoft Office documents, spreadsheets, and presentations:

- Hidden Information (PowerPoint)
- Your name (All)
- Your initials (All)
- Your company or organization name (All)
- The name of your computer (All)
- The name of the network server or hard disk where you saved the document (All)
- Other file properties and summary information (All)
- Non-visible portions of embedded OLE objects (All)
- The names of previous authors (All)
- Document revisions (Word, Excel)
- Document versions (Word)
- Template information (Word, PowerPoint)
- Hidden text (Word, Excel)
- Hidden Cells (Excel)
- Personalized views (Excel)
- Comments (All)
- Globally Unique Identifiers (GUIDs) (All)

Next to each identified type of stored metadata, a designation was made to distinguish the information that may be unique to each Microsoft Office program. Where the metadata category applies to all programs, the designation reflects that fact. Information such as this can prove very helpful to a forensic investigator as well as any other person involved in reconnaissance activities. In either case, a person could possibly determine the author of the document, as well as his or her userid, and the network path location of the original document. This information could be used to map at least a small portion of network's layout.

Microsoft Word, Excel, and PowerPoint have added features that facilitate cooperation among document authors through the use of the Track Changes and Comments features. If a document author publishes a document without taking measures to remove all comments and changes from the document, he or she could reveal all the contributors to the document as well as its original information; all of which could lead to the exposure of data that was not meant to be

seen by anyone, much less the forensic investigator.

It is easy to see how these types of capabilities can be useful to document authors. It can allow the reader to take notice of the creator as well as facilitate collaboration among several authors. In versions prior to Microsoft Office 2003, the problem is that all information is stored in a proprietary binary format. Hence, the user had virtually no way to know what metadata was stored with the document. Even though Microsoft Office 2003 does offer the capability to save the information in the more widely accepted XML format, the binary format is still the default and the non-technical user may not be aware of the extended capabilities in the newer format.

#### **5. OPENOFFICE.ORG METADATA**

OpenOffice.org's office product has increased in popularity over the past several years and has taken a different approach to its file format. OpenOffice.org stores all the document data, to include metadata, in a series of XML files that are associated with the overall filename given by the user.

OpenOffice.org outlines in its XML file format specification [21] the capability to store a lot of metadata inside of its documents, spreadsheet, and presentations to include:

- Generator (All)
- Title (All)
- Description (All)
- Subject (All)
- Keywords (All)
- Initial Creator (All)
- Creator (All)
- Printed By (All)
- Creation Date and Time (All)
- Modification Date and Time (All)
- Print Date and Time (All)
- Document Template (All)
- Language (All)
- Editing Duration (All)
- User-defined Metadata (All)
- Document Statistics (All)

According to the specification document [21], all of the above categories are stored, or have the capability to be stored, with each document, spreadsheet, and presentation created in OpenOffice.org. This is much of the same type of information that is stored by Microsoft Office; however, OpenOffice.org does not make assumptions about what personal information the user would like associated with the file. For instance, if a user does not enter a name in the application's user information widget, the program will not try and get the registration information from the operating system in order to discover the name or initials of the user.

As with Microsoft Office, OpenOffice.org has the capability to store information facilitating the use of the Track Changes utility and commenting features. OpenOffice.org provides easier access to a files metadata data, which facilitates user and examiner discovery, given he or she knows what they are looking for. OpenOffice.org stores the metadata associated with the overall document, spreadsheet, or presentation in a separate plain text file for easier access. The specific format of the metadata and file structure will be discussed in a later section.

#### **6. PDF METADATA**

PDF documents store metadata information also that, as mentioned before, is sometimes copied over from the original

source documents from which the PDF was created. The types of data and format are outlined in the *PDF Reference Manual version 1.6* [22], which specifies that the information is stored inside of a document information library. Furthermore, the information is located inside of an optional Info entry in the trailer of the PDF file.

As the other products mentioned previously, it has the capability to store much of the same information, which includes:

- Title
- Author
- Subject
- Keywords
- Creator
- Producer
- CreationDate
- ModDate

Adobe Systems Incorporated is also part of an initiative to standardize the metadata stored in documents to facilitate greater ease in the storage and searching of information in databases as well as web-based environments. This effort is called the Extensible Metadata Platform (XMP) [23] and one specific purpose is to allow different programs that process Adobe files to add their own types of metadata each step of the way. The type of information stored as metadata will not change, but only the format in which it is stored using a more accessible XML type format.

## 7. METADATA FORMATS

Given that the discussion has focused on the types of information stored and the benefits and disadvantages of metadata in office application products, I will now present how the metadata is stored in each of the programs addressed in this paper. Examples of how the data is stored are presented, when possible, to ease discovery by the forensic investigator.

### 7.1 Microsoft Word, Excel, and PowerPoint (Pre-XML)

Microsoft versions up until very recently only stored the information from Word, PowerPoint, and Excel in a proprietary binary format called the Object Linking and Embedding (OLE) 2 protocol. In this format [24], all information is written in streams that are stored in the binary file as linked lists of file blocks. The metadata is stored for the most part in what is called the Summary Information and Document Summary Information storage information streams within this file. The limited amount of specific information on this format prevents a more detailed analysis of the storage locations as in subsequent sections. This binary format has been used by all versions of Office.

Metadata in these documents is not easily viewed by the forensic investigator, much less the user. One can view the hidden information saved with a document by opening up the file with a hexadecimal viewer. This will reveal some of the hidden information; however, it is not practical solution.

A more common approach to viewing metadata in the older versions of Word, Excel, and PowerPoint is to use a third-party application. Many of these applications have been around for years and can provide a good summary of the information contained within these documents. This paper will discuss some of the third-party tools in a later section. These tools can assist the forensic investigator's efforts to locate metadata in documents, to include those created by older versions of Microsoft Office.

### 7.2 Microsoft Word, Excel, and PowerPoint (New XML Format)

As of Microsoft Office 2003, the user is given the option of saving in the new XML format that Microsoft uses in hopes of bringing it closer to a more open standards format. It is not the same as the OpenDocument XML format; but nevertheless, it stores information in a much more accessible manner than before. After Office 2003, the XML format will be the standard file format rather than just an option.

When saving documents in the new XML format, the user or investigator can access the metadata by viewing the source XML file with a text editor. The following is not meant to be an exclusive list of the information stored, but merely an example of how the metadata is stored near the beginning of a test Microsoft Office 2003 Word XML file:

```
<o:DocumentProperties>
  <o:Title>This is a test</o:Title>
  <o:Author>JJ</o:Author>
  <o:LastAuthor>JJ</o:LastAuthor>
  <o:Revision>2</o:Revision>
  <o:TotalTime>0</o:TotalTime>
  <o:Created>2006-02-26T02:06:00Z</o:Created>
  <o:LastSaved>2006-02-
26T02:06:00Z</o:LastSaved>
  <o:Pages>1</o:Pages>
  <o:Words>9</o:Words>
  <o:Characters>55</o:Characters>
  <o:Company>Self</o:Company>
  <o:Lines>1</o:Lines>
  <o:Paragraphs>1</o:Paragraphs>
  <o:CharactersWithSpaces>63</o:CharactersWith
Spaces>
  <o:Version>11.6568</o:Version>
</o:DocumentProperties>
```

As shown above, plenty of information is readily available to anyone who can view this document as a plain text file or with an XML editor. Microsoft begins the metadata section with the `<o:DocumentProperties>` and continues with fairly self explanatory tags describing the document, as well as the authors, and company information. Needless to say, this can be helpful to the investigator, as well as the normal user, because the file contents are apparent as opposed to the binary format discussed previously.

### 7.3 Microsoft's Rich Text Format (RTF)

Microsoft's RTF documents have been in existence for many years. Before documents were easily transferable and recognized by the multitude of document creation and editing programs, this format facilitated cross-platform, cross-application document sharing. As such, many users are unaware of the information that is stored with these documents. One might believe that because this is a simplified text format, that much of the metadata information is not stored with the document as with Word's .doc files. However, this is not the case.

Microsoft stores much of the same type of information in RTF files [25] as it does with the normal Word files. A sample format is provided below and was extracted from an RTF file originally created in Microsoft Word 2003 and then saved as a RTF file:

```
{\info
  {\title Document Title}
  {\author Author Name}
  {\operator Userid}
  {\creatim\yr2006\mo4\dy2\hr7\min37}
  {\revtim\yr2006\mo4\dy2\hr7\min37}
  {\version2}
  {\vedmins0}
```

```

{\nofpages5}
{\nofwords709}
{\nofchars4042}
{\*\company Company Name }
{\nofcharsws4742}
{\vern24579}
}

```

What is interesting about the metadata stored in RTF document is that it still keeps track of revisions and version numbers. The tags used by the RTF files are self-explanatory and are saved with the RTF document automatically. If a user creates a RTF document, or saves a Word document as an RTF file, the metadata contained within the file will then also be contained within the RTF document file.

## 7.4 OpenOffice.org XML

OpenOffice.org takes a unique approach in the creation and storage of information in its native XML format. Each time a document, spreadsheet, or presentation is created, a series of files are created and stored in one archive. Inside of this archive are several different files that contain all the information the application requires.

For example, when an OpenOffice.org text document is created, the following files are stored inside of the .sxw file: content.xml, meta.xml, settings.xml, and styles.xml. There are a few more files stored, but they are not relevant to this discussion. In order to view the contents of these files, the files must be extracted out of the .sxw file, much like a zip file.

Once the files are extracted, it is evident that the meta.xml file is the most likely candidate to contain the metadata associated with the document. The following is not meant to be an exhaustive list of the information stored, but merely an example of how the metadata is stored near the beginning of a test OpenOffice.org Writer document:

```

<office:meta>
  <dc:title>Title</dc:title>
  <dc:description>Comment</dc:description>
  <dc:language>en-US</dc:language>
  <meta:initial-creator>Name</meta:initial-creator>
  <meta:creation-date>X</meta:creation-date>
  <dc:creator>Name</dc:creator>
  <meta:generator>NeoOffice</meta:generator>
  <dc:creator>Name</dc:creator>
  <meta:keywords>
    <meta:keyword>First</meta:keyword>
    <meta:keyword>Second</meta:keyword>
  </meta:keywords>
  <dc:date>X</dc:date>
  <dc:subject>Subject</dc:subject>
  <meta:printed-by>Name</meta:printed-by>
  <meta:print-date>X</meta:print-date>
  <meta:duration-time>X</meta:editing-duration>
  <meta:editing-cycles>4</meta:editing-cycles>
  <meta:editing-duration>X</meta:editing-duration>
</office:meta>

```

This is just a small example of the metadata stored with OpenOffice.org documents. The important point to note about metadata in OpenOffice.org's documents, presentations, and spreadsheets, is the ease of access afforded by storing all of the information in the meta.xml file. This provides ready access by users and investigators to the wealth of information that can be stored with the OpenOffice.org office products.

## 7.5 Adobe's PDF

As mentioned previously, Adobe's PDF document formation has gained in popularity over the past several years because of

the cross-platform capabilities it affords. Each time a PDF document is created; the source document's metadata is copied as well as incorporated into Adobe's format for metadata storage, unless this option is specifically disabled by the user. A sample format, non-XMP, is provided below and obtained from the *PDF Reference Manual version 1.6* [22]:

```

1 0 obj
  << /Title (PostScript Language Reference)
    /Author (Adobe Systems Incorporated)
    /Creator (Adobe Framemaker)
    /Producer (Acrobat Distiller)
    /CreationDate (D:19970915110347-08'00')
    /ModDate (D:19990209153925-08'00')
  >>
endobj

```

Given that the XMP specification is very extensive and has a multitude of options, the reader is referred to the specification document [26] for further information. Needless to say, storing the metadata in a type of XML format will aid tremendously in the document forensics arena.

## 8. DETECTION TOOLS

Metadata detection tools are prevalent, given the number programs that were returned after a search at Google. A few freeware and shareware programs were chosen randomly from the results of my search to provide a discussion on the functionality of the available tools. In addition, one program was obtained from a guest speaker at a conference as a special circumstance for research purposes.

The tests conducted in this section are not scientific and do not examine a large, diversified set of documents. The twenty documents tested consisted of an assortment of: Microsoft Word documents, PowerPoint presentations, and Excel spreadsheets (pre-XML and new XML formats); OpenOffice.org Writer documents, Impress presentations, and Calc spreadsheets; and PDF documents.

Eighty-five percent of the documents were chosen randomly from those discovered through Google searches, and at the onset of the tests, were not known to contain any metadata. The remaining fifteen percent of the documents were chosen from the author's own document library as they were known to contain much of the types of metadata discussed in this document, therefore highlighting the capabilities/limitations of the detection tools discussed in this section.

### 8.1 libextractor and extract

This open source command-line tool [27] was discovered through one of many searches for metadata extraction tools. *Libextractor* is the actual library that is used to extract metadata from several different document formats and the *extract* tool is the actual program executed on the command-line to perform the extraction. Both the library and program are available for Linux and Windows platforms; free; and redistributable under the GNU general public license.

From the library's home page [27], it supports the following formats: HTML, PDF, PS, OLE2 (DOC, XLS, PPT), OpenOffice (sxw), StarOffice (sdw), DVI, MAN, MP3 (ID3v1 and ID3v2), OGG, WAV, EXIV2, JPEG, GIF, PNG, TIFF, DEB, RPM, TAR(.GZ), ZIP, ELF, REAL, RIFF (AVI), MPEG, QT and ASF. Even though this utility has the capability to detect metadata in many different file formats, which can undoubtedly prove beneficial to the forensic examiner, this section will only explore its capabilities against the PDF, OLE2, and OpenOffice file formats.

The installation of both the library and program were

straightforward. The installation machine was running Ubuntu Linux which offered the capability to simply use the *apt-get* utility to install utilities. Likewise, running the program was simple and yielded immediate results.

To run the program, one only needs to issue the command followed by the filename to be examined:

```
$ extract filename
```

Using the tool in this manner will simply display the results to the terminal screen in an easy to read format. It does have additional options, one of which is the ability to change the language settings (i.e. search a German instead of U.S. English document). Another useful option is the ability to output the results of the metadata discovery in BiBTeX format in case the examiner or author used these types of entries to assist in documentation.

During the conduct of this test, options were added to the command line to cause the program to be verbose (*-v*) and print the filename in the results (*-f*). This program was able to operate and report basic metadata results on each of the file types outlined previously. The basic types of metadata reported by the program on each file include the following:

- Operating System (On which the file was created)
- Size
- Organization
- Modification Date
- Creation Date
- Page Count
- Software (On which the file was created)
- Version (of the file)
- Format (Template used to create the file)
- Author
- Title
- Keywords
- Filename

Not all of the files tested reported all of the previous metadata, but the program picked up as much as each file contained, with regards to the basic information above. This program did not report metadata such as comments, last ten authors, or track changes information, as it is not designed to do so. The user can issue the *-v* option to the program to view all the types of metadata on which the program can report regarding the many different supported file types.

This program is only designed to report basic metadata information associated with a file. Given its design, it performed well and worked as expected. However, to discover the data that may be more incriminating to a user and beneficial to a forensic investigator, it is inadequate.

## 8.2 pdfinfo

*Pdftinfo* is a command-line tool provided in the open source *xpdf* [28] package for different versions of Linux and UNIX. The purpose of the *pdftinfo* program is to print out metadata information that is stored in PDF documents. The program has a base functionality that is invoked with the following command:

```
$ pdftinfo filename
```

More information regarding the stored metadata can be returned in XML format by supplying it with an additional option (*-meta*). According to the *pdftinfo* man page, the program is designed to report the following metadata information that may be contained within a tested file:

- Title

- Subject
- Keywords
- Author
- Creator
- Producer
- Creation date
- Modification date
- Tagged (yes/no)
- Page count
- Encrypted flag (yes/no)
- Print and copy permissions (if encrypted)
- Page size
- File size
- Linearized (yes/no)
- PDF version

The program worked as described and as expected reporting all basic metadata present in the document. However, this program, like the *extract* program discussed previously, was not able to report on embedded comments in a PDF file. Of the PDF documents tested, a good amount of information was reported, none of which provided a tremendous amount of useful data except for the type of program used to create the document as well as dates created and modified. These dates could be useful by an investigator to create a timeline of activity leading an investigator to other information or supporting hypotheses.

## 8.3 Metadata Analyzer

Metadata Analyzer [29] is a freeware application for use on Windows ME/NT/2000/XP/2003. It has only one purpose and that is to report on the metadata present in Microsoft Word documents, PowerPoint presentations, and Excel spreadsheets. The program is small, easy to install, and operates as a stand-alone program. When launched, the program offers a user interface that initially detects and reports the built-in Office properties: User Name, User Initials, and Company.

The user has the capability to point the program to a file for analysis or the user can simply drag the respective Word, PowerPoint, or Excel document icon onto the program. Once that action is performed, the user must press the Analyze button to begin metadata detection. The program offers an easy to read report but no capability to save, which is not conducive to documentation. The following document properties are reported:

- Title
- Creation Time
- Last Edited
- Last Print Date
- Author
- Last Author
- Company
- Revision Number
- Total Editing Time

Furthermore, the program reports any custom properties that may also be present in the document metadata. This program is easy to use but like the other programs tested and discussed previously, lacks the capability to report more advanced metadata information such as embedded comments and revisions from Track Changes.

## 8.4 ezClean

ezClean [30] is a commercial metadata detection and removal tool for use on Microsoft Office products. This product works on all Windows operating systems after and including

Windows 95; however, the program incorporates itself into Microsoft Office and therefore requires Office 2000 or later. KKL Software [31] sells the program for a nominal fee but offers a 45 day free trial, which was used to perform document testing.

Program installation was easy after which the program is accessible through a single button on the Word, Excel, or PowerPoint toolbar. While a document is open in the respective Office program, the user simply has to click the button, which causes the software to automatically check for metadata.

Even though the major purpose of the program is to remove metadata saved with the document, it provides a straightforward user interface and reporting tool before the user has the choice to remove the data. The initial check opens an interface that presents basic options for viewing the metadata; however, the report function creates a web page that provides complete access to the detected information.

The existence of metadata is reported in several different sections. A sample of the type of metadata reported under the respective section is listed below:

- Built-in Document Properties
  - Title
  - Subject
  - Author
  - Company
- Document Statistics
  - Created (Date)
  - Modified (Date)
  - Printed (Date)
  - Last saved by
  - Revision number
- Custom Properties
  - \_EmailSubject
  - \_AuthorEmail
  - \_AuthorEmailDisplayName
- Macros
- Attached Template
- Comments
- Embedded OLE Objects
- Author and Filename History
- Undo History
- Fast Saves
- Track Changes Feature (ON/OFF)
- Document Revisions (from Track Changes)
  - Author (Userid)
  - Type (i.e. Inserted)
  - Text (i.e. “added text”)

The Custom Properties section listed above was different and specific for each of the documents containing this type of information. However, it was a valuable piece of information as it was easy to see who originally authored the document and who sent it out through e-mail at one time or another.

For instance, during the test of the document that returned the above custom properties, it was discovered that the document author was different than the person that e-mailed the document. While this may not be a major discovery in many cases, a forensic investigator could perform a link analysis and attempt to follow the personnel with access to the document. In another document, which in fact was a 2005 first quarter financial report from a major wireless phone company obtained from Google, the person who e-mailed the document as well as the web-based application that was used to create the report was discovered from the custom properties

[32].

In addition, the program was robust at not only reporting that the Track Changes feature was on or off, but more importantly, it reported all of the information present with respect to all the changes that were made as well as the userid of the revisioner, the type of action taken, and the text changed, deleted, or added to the document.

## 8.5 Document Detective

Document Detective [33] is a commercial product designed by SRS Technologies and is catered to government use. A trial version for research purposes was obtained after attending a conference presentation by its author. It is a Windows based program that is designed to work on Microsoft Windows 2000 and XP. It also requires that the user have Microsoft Office XP or 2003 installed on the program used for testing. According to the program’s documentation [34], it works on: Word documents; Excel spreadsheets; PowerPoint presentations; Rich Text Files; HTML files; web archives (MHT & MHTML); XML files; text files; and PDF documents.

As mentioned before, an evaluation copy had to be specifically requested from the program’s author, so it was not as easily obtainable as the other programs tested. The installation of the program was straightforward, after which the program can either be launched from within Word, PowerPoint, and Excel, or as a separate program. All documents were tested against this program, which provided a wealth of information regarding the metadata contained within each of the documents.

This program is not only designed to detect metadata but scrub documents for defined keywords. As such, the first screen that is presented allows the user to select from a predefined list of keyword categories, such as “Corporate Releasable Information,” that will then allow the user to determine what set of keywords he/she would like the program to flag if present. After selection, the program scans the document and presents all the metadata information and keywords found. The discovered data is presented in a series of categories.

A sample of the metadata categories and sub-categories presented follows:

- Document
  - Path
  - Filename
- Built in Document Properties
  - Title
  - Subject
  - Author
  - Keywords
  - Comments
  - Template
  - Last Author
  - Creation Date
  - Last Save Time
  - Company
- Comments
  - Date
  - Author
  - Initial
  - Range Text
  - Reference Text
  - Scope Text
  - Show Tip

- Revisions
  - Author
  - Date
  - Format Description
  - Range Text
  - Type

This program was by far the most thorough program at reporting metadata information tested to this point. It reported far more detailed information than the aforementioned programs; however, the interface was much more difficult to navigate than all the other tools in order to view the desired information. This is the first version of this program so it is expected that it will undergo some major improvements as time goes on.

One of the benefits of this program is that it worked on the full range of Microsoft Office products, to include the new Office XML format, and PDF documents. However, it was not able to work on the OpenOffice.org range of files even though they are by default saved in an XML format.

## 8.6 TRACE! by Workshare

TRACE! [35] proved to be one of the more robust tools. It is free software provided to scan Microsoft Office documents (Word, PowerPoint, and Excel) and runs on English versions of Microsoft Windows 2000 and XP. The purpose of the program is to scan documents for metadata and extensively report its findings.

TRACE! offers the user several different methods to scan supported files such as: scanning the active document currently being edited; selecting a file or several files located on the hard drive; scanning e-mail attachments located in the user's Microsoft Outlook local mail folder(s); and specifying a URL, which will cause the program to search a specified domain for Word, Excel, and PowerPoint files and evaluate the metadata in the discovered documents.

TRACE! monitors and reports risk levels of the document, spreadsheet or presentation. The risk levels used by this program are high, medium, and low. A determination of the level of risk posed by the document depends on the types of metadata present in the document. The details of what qualifies as high, medium, and low risk are found in the user's guide [36] and those elements dealing directly with metadata are included below:

High Risk:

- Identity Information
- Track Changes
- Comments
- Hidden Text
- Speaker Notes (PowerPoint)
- Versions
- Last Ten Authors with UNC Paths
- Hidden Slides (PowerPoint)
- Automatic Word Versioning

Medium Risk:

- Custom Properties
- Macros
- Document Reviews

Low Risk:

- Last Ten Authors w/o UNC paths
- Document Statistics
- Built In Properties
- Attached Template

This program was the most notable in the range of information reported as well as the ways that the program could be used. For instance, after running the program against all of the test documents and obtaining the expected information, the jmu.edu domain was scanned for Word documents. TRACE! has a built-in capability to scan entire domains, by simply choosing that option from the main user interface. The scan on the jmu.edu domain took a long time, scanned over 500 discovered documents, and reported the risk level of many of those documents.

From the documents scanned in the test set as well as the domain, TRACE! reported its results in an easy to read and save format. Each document was broken down into the risk categories listed above with the respective information that led to the assessed risk category. The information was presented in HTML format which allowed for excellent documentation and highlighted all information that should have been reported.

During the test that was conducted against the jmu.edu domain, I was able to discover one specific document that reported the last ten authors, which included their userid. In addition, the network path was reported on which the document was last saved (\DATA7\OSP\COMMON\00nwsltr). With this information, the investigator now has the network directory that the userid has privileges on which to save. The \_PID\_GUID associated with this document was also reported, which could lead to discovery of the exact machine on which the specific user created the document.

## 8.7 Tool and Test Results

Overall, the programs tested provided at least the basic metadata present in each document. The Linux tools tested, as well as Metadata Analyzer, were by far the least capable of the group of tools. They provided minimal information and only one, *libextractor*, was able to work on more than one set of document types.

ezClean, Document Detective, and TRACE! provided a wealth of information about the metadata they were designed to detect and report, which makes them more useful in forensic scenarios. ezClean incorporated itself in the Microsoft Office programs and was very easy to use.

Document Detective was a little more difficult because it installs itself as a macro inside of Microsoft Office programs. If the user has the macro security level set to high, the program will not run; if set to medium, the user has to press several buttons to allow the program to run. However, the ability of Document Detective to work on a wider range of documents makes it more useful in a dynamic forensic environment.

TRACE! is the best program with regards to usability and interface, but was limited by the amount of documents on which it was designed to work. Its user interface, detection and reporting capabilities were richer than the other programs tested. If an investigator is only required to examine Microsoft Office documents, TRACE! should be considered for addition in the forensic analysis toolkit.

Although only briefly touched upon during the discussion of the various tools, much information was discovered during the conduct of the investigation. Usernames were discovered in several documents that can be used to ascertain how a user logs onto a network and proved even more valuable when network storage paths were discovered. The combination of usernames and storage paths allows the investigator to get a

brief insight into how user logins are assigned and the associated network layout.

In addition, the documents having document revisions present through the use of Track Changes, as well as comments, allowed the discovery of how many different people viewed and changed the document. Discoveries such as this proved useful in examination of the Dodgy Dossier and will continue to prove valuable in future examinations.

## 9. CONCLUSION

Metadata is data that describes other data. As examined in this paper, metadata can lead to the discovery of a plethora of information. Examination of document metadata can lead to the discovery of information such as: document author names; names of contributors as well as their recommended changes and comments; network storage path locations; userids of the document author; as well as computer specific information such as the GUID.

Several different real-world scenarios were presented demonstrating how metadata was exploited to reveal hidden information. In some cases, it was admitted as evidence to refute or support claims. In others, it resulted in an embarrassing situation for companies and governments alike.

This paper discussed the types of metadata information stored in documents, spreadsheets, and presentations created in Microsoft Office and OpenOffice.org applications, PDF document, and provided examples of their respective storage formats. As shown in this paper, a move towards XML file formats has provided investigators easier access to this type of information.

Finally, this paper tested and presented results of several tools available to detect and document the presence of metadata. *Libextractor*, *pdfinfo*, and Metadata Analyzer provide limited functionality, while *ezClean*, Document Detective, and TRACE! demonstrated greater detection and reporting capabilities. Many of the tools tested are alone not sufficient for an investigator; however, a combination of those mentioned can report the presence of all the different types of metadata present in the documents addressed during this paper.

Metadata is useful information that can benefit users as the trend to participate in virtual collaboration grows. Understanding metadata and the implementation of proactive control measures are essential to prevent the leakage of sensitive information. Until users become more conscious of the hidden information contained in and disseminated with their documents, forensic investigators can and will exploit document metadata during the conduct of an investigation.

## 10. REFERENCES

[1] U.S. Department of Justice, *Forensic Examination of Digital Evidence: A Guide for Law Enforcement*. NIJ Special Report. April 2004. Retrieved 1 March 2006 from <http://www.ncjrs.org/pdffiles1/nij/199408.pdf>.

[2] Howell, B. "Digital Forensics: Sleuthing on Hard Drives and Networks," *The Vermont Bar Journal*. Fall 2005. Retrieved 2 March 2006 from [http://www.strozllc.com/docs/pdf/BHowell\\_DigitalForensics.pdf](http://www.strozllc.com/docs/pdf/BHowell_DigitalForensics.pdf).

[3] Ball, C. "Make Friends with Metadata," *LegalTechnology*. 26 January 2006. Retrieved 1 March 2006 from <http://www.law.com/jsp/ltm/pubArticleLTN.jsp?id=1138183510640>.

[4] Zeller, T. Jr. "Beware Your Trail of Digital Fingerprints,"

*New York Times Online*, 7 November 2005. Retrieved 28 February 2006 from <http://www.nytimes.com/2005/11/07/business/07link.html?pagewanted=1&ei=5090&en=98e8af679a0797f4&ex=1289019600&partner=rssuserland&emc=rss>.

[5] Byers, S. "Information Leakage Caused by Hidden Data in Published Documents," *IEEE Security and Privacy Magazine*. March/April 2004.

[6] "Dangers of Document Metadata," Retrieved 26 February 2006 from [http://www.metadatarisk.org/document\\_security/dangers\\_of\\_docmetadata\\_overview.htm](http://www.metadatarisk.org/document_security/dangers_of_docmetadata_overview.htm).

[7] Kernan, D. *Hidden Data in Electronic Documents*. GIAC GSEC Practical, 5 July 2004. Retrieved 22 February 2006 from <http://www.sans.org/rr/whitepapers/privacy/1455.php>.

[8] Microsoft Knowledgebase Article 222180, "How and why unique identifiers are created in Office documents," Revision 1.0, 5 August 2004. Retrieved 30 March 2006 from <http://support.microsoft.com/kb/222180/>.

[9] Leach, P., Mealling, M. and Salz, R. *Request for Comments (RFC) 4122: A Universally Unique Identifier (UUID) URN Namespace*, July 2005. Retrieved 15 May 2006 from <http://www.ietf.org/rfc/rfc4122.txt>.

[10] Reiter, L. and Louderback, J. "Melissa trail leads to 'ex' virus writer," *ZDNet Online*. 29 March 1999. Retrieved 30 March 2006 from [http://news.zdnet.com/2100-9595\\_22-514175.html](http://news.zdnet.com/2100-9595_22-514175.html).

[11] Ewalt, D. "When Words Come Back from the Dead," 13 December 2005. Retrieved 23 February 2006 from [http://www.forbes.com/2005/12/13/microsoft-word-merck\\_cx\\_de\\_1214word.html?partner=yahootix](http://www.forbes.com/2005/12/13/microsoft-word-merck_cx_de_1214word.html?partner=yahootix).

[12] Bone, J. and Blanford, N. "UN office doctored report on murder of Hairy," *Times Online*. 22 October 2005. Retrieved 28 February 2006 from <http://www.timesonline.co.uk/article/0,,251-1837848,00.html>.

[13] Karp, D. "Revealing Codes," *PC Magazine*. 8 June 2004. Retrieved 28 February 2006 from <http://www.pcmag.com/article2/0,1759,1585411,00.asp>.

[14] Loney, M. "'Dodgy-dossier syndrome' rife in the workplace," *ZDNet UK*. 14 November 2003. Retrieved 28 February 2006 from <http://news.zdnet.co.uk/business/management/0,39020654,39117905,00.htm>.

[15] Lemos, R. "Melissa creator may be uncovered," *ZDNet Online*. 29 March 1999. Retrieved 30 March 2006 from [http://news.zdnet.com/2100-9595\\_22-514170.html](http://news.zdnet.com/2100-9595_22-514170.html).

[16] Bezroukov, N. "Melissa worm/virus – a worm parasiting on MS Office 97 architectural problems and MS Word users' ignorance," Retrieved 30 March 2006 from [http://www.softpanorama.org/Antivirus/AV\\_Secrets/Vgallery/melissa.shtml](http://www.softpanorama.org/Antivirus/AV_Secrets/Vgallery/melissa.shtml).

[17] Microsoft Knowledgebase Article 223396, "How to minimize metadata in Office documents," Revision 3.2, 28 January 2005. Retrieved 20 February 2006 from <http://support.microsoft.com/kb/223396>.

[18] Microsoft Knowledgebase Article 290945, "How to minimize metadata in Word 2002," Revision 2.6, 6 January 2006. Retrieved 30 March 2006 from <http://support.microsoft.com/kb/290945/>.

- [19] Microsoft Knowledgebase Article 223789, "How to minimize metadata in Microsoft Excel Workbooks," Revision 5.0, 11 February 2005. Retrieved 30 March 2006 from <http://support.microsoft.com/kb/223789/EN-US/>.
- [20] Microsoft Knowledgebase Article 314880, "How to minimize the amount of metadata in PowerPoint 2002 presentations," Revision 1.0, 10 September 2004. Retrieved 30 March 2006 from <http://support.microsoft.com/kb/314880/EN-US/>.
- [21] Sun Microsystems, *OpenOffice.org XML File Format 1.0 Technical Reference Manual*, Version 2, December 2002. Retrieved 20 February 2006 from [http://xml.openoffice.org/xml\\_specification.pdf](http://xml.openoffice.org/xml_specification.pdf).
- [22] Adobe Systems Incorporated, *PDF Reference- Adobe Portable Document Format*, 5<sup>th</sup> ed, Version 1.6, 14 November 2004. Retrieved 30 March 2006 from <http://partners.adobe.com/public/developer/en/pdf/PDFReference16.pdf>.
- [23] Adobe Systems Incorporated, "Extensible Metadata Platform (XMP) homepage," Retrieved 30 March 2006 from <http://www.adobe.com/products/xmp/main.html>.
- [24] The Programmer's File Format Collection. <http://www.wotsit.org/>
- [25] Microsoft Download Center, "Word 2003: Rich Text Format (RTF) Specification, Version 1.8," 20 April 2004. Retrieved 30 March 2006 from <http://www.microsoft.com/downloads/details.aspx?FamilyID=ac57de32-17f0-4b46-9e4e-467ef9bc5540&displaylang=en>.
- [26] Adobe Systems Incorporated, *XMP Specification*, 2004. Retrieved 30 March 2006 from <http://www.adobe.com/products/xmp/pdfs/xmpspec.pdf>.
- [27] libextractor - A Simple Library for Keyword Extraction. <http://gnunet.org/libextractor/>.
- [28] Xpdf: A PDF Viewer for Linux. <http://www.foolabs.com/xpdf/home.html>.
- [29] Metadata Analyzer. <http://www.smartpctools.com/metadata/>.
- [30] ezClean Home Page. <http://www.kklsoftware.com/products/ezClean/details.asp>.
- [31] KKL Software Home Page. <http://www.kklsoftware.com/index.asp>.
- [32] Document located at [http://media.corporate-ir.net/media\\_files/irol/95/95539/reports/cingular\\_proforma.xls](http://media.corporate-ir.net/media_files/irol/95/95539/reports/cingular_proforma.xls)
- [33] Document Detective Home Page. <http://www.stg.srs.com/eds/docdet/>.
- [34] Document Detective Capabilities Page. <http://www.stg.srs.com/eds/docdet/capabilities.htm>.
- [35] TRACE! by Workshare. <http://www.workshare.com/products/trace/>.
- [36] TRACE! by Workshare: User's Guide. [http://www.workshare.com/downloads/products/trace/build\\_8256/TRACE!%20by%20Workshare%20Quick%20Reference%20Guide%202.0.pdf](http://www.workshare.com/downloads/products/trace/build_8256/TRACE!%20by%20Workshare%20Quick%20Reference%20Guide%202.0.pdf).