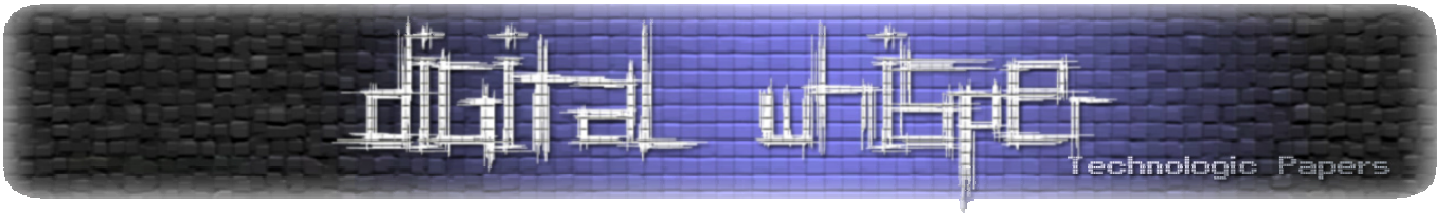


Meta Data – פירורי מידע לאויב שלך

מאת אפיק קסטיאל (cp77fk4r)

"המונח Meta Data הוא אחד מהמונחים החשובים ביותר בעולם ניהול המידע. שימוש נכון ב-Meta Data יכול לשפר משמעותית את יכולת הנגישות למידע, ואת אפקטיביות השימוש באותו מידע. המשמעות המילולית של המונח היא "מידע על מידע", כלומר שדות מידע שמוסיפים (אוטומטית או ידנית) למידע. יש שתי מטרות עיקריות ל-Meta Data:

- להוסיף אינפורמציה שאינה כתובה בתוכן עצמו, ואשר עשויה לאפשר להבין את הכתוב טוב יותר, או לעזור לנו להתייחס לכתוב בהתאם. לדוגמא – שדה Meta Data סטנדרטי הוא התאריך. במקרים רבים התאריך עוזר לנו להבין אם הכתוב רלוונטי עבורנו או לא. שדה אחר הוא שם הכותב, שמאפשר לנו לשפוט לגבי מהימנות התוכן (אם אנחנו מכירים את הכותב). גם שדות נוספים כגון – אירוע שבמסגרתו התוכן נכתב, או יחידה ארגונית – עוזרים לנו להבין מניעים/זוויות מבט והיבטים נוספים, שאינם באים בהכרח לידי ביטוי בתוכן עצמו.

- לסייע באחזור המידע. שדות Meta Data רבים מהווים כלי מרכזי באחזור נוח ומוצלח של תכנים. היכולת לתייג תוכן עפ"י מספר מאפיינים, מאפשר לנו לאחזר אותו ע"י שימוש באחד או יותר מאותם מאפיינים. חיפוש עפ"י מספר מאפיינים מאפשר לנו לצמצם את החיפוש ולמקד אותו בדיוק בתכנים אותם אנו מחפשים. כאשר בונים פתרון ניהול תכנים, במקרים רבים ההצלחה שלו תלויה בבחירה נכונה של שדות ה-Meta Data ובהטמעה מוצלחת של השימוש בהם."

(נכתב במקור ע"י יאיר דמבינסקי, סמנכ"ל פרוייקטים Byon IT Solutions)

כמו שאפשר לראות Meta Data משתמש בעיקר להגדרת/איחזור יעיל של המידע-שזה דבר מצוין, אך כמו לכל דבר-גם לנושא הזה יש חסרונות.

הסיכון

מספר רב של תוכנות מוסיפות Meta Data לקבצי הפלט שלהם, אם מדובר במסמכי Doc של Word או במסמכי PDF של Adobe, ואם מדובר בקבצי תמונות של Photoshop או GIMP וגם אם מדובר בקבצים בינאריים שקומפלו ע"י מהדר כזה או אחר.

כאשר כותב מאמר מעוניין להוסיף Meta Data למאמר שלו בכדי לשפר את יכולות האיחזור של מנוע חיפוש כזה או אחר לגבי אותו המאמר זה מצוין, אבל בהרבה מאוד מקרים תוכנות שונות מצרפות למאמר Meta Data שיכול לעזור לגורמים עויינים ללמוד נתונים "אישיים" שיכולים לעזור להם או למקד אותם בעת פריצה לאותו מחשב או לרשת האירגון שבה נכתב המאמר.

הרבה אירגונים גדולים לא מפנים מספיק תשומת-לב לנושא, ולמרות שלרוב לא נחשפות ssמאות או פרטים אישיים של משתמשים, עדיין מדובר במידע יעיל מאוד כאשר מדובר בניתוח מבנה האירגון או הרשת הפנימית של אותו האריגון.

תוכנות שונות ודרכי ייצוא שונות משנות את ה-Meta Data שנכנס לקובץ המיוצא, אך אפשר למצוא שקבצים שונים מכילים Meta Data כגון:

- שם המחשב (ולפעמים גם ה-IP הפנימי של אותו המחשב) עליו נכתב הקובץ.
- שם המשתמש (ולפעמים גם שם הדומיין) שבחשבונו נכתב הקובץ.
- שם האפליקציה (ולרב גם גרסתה) שייצאה את הקובץ.
- סוג מערכת ההפעלה וגירסתה/הפצתה שעליה כתבו את הקובץ.
- מיקום הקובץ שכותב הקובץ בחר בעת שמירת הקובץ.
- כתובת דוא"ל של המשתמש שכתב את הקובץ.
- שמות/כתובות שרתים ברשת הפנימית המבצעים ניהול של הקבצים.

- שם/סוג/גרסאת המדפסת שבה השתמשו בכדי להדפיס את הקובץ.
- כמות הזמן שבו נכתב הקובץ.

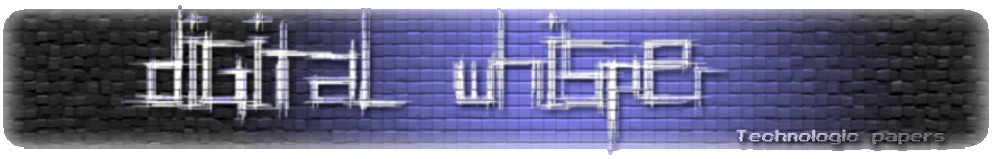
כאשר מדובר ב-Meta Data שאנו בחרנו להכניס לקובץ, כגון שמו של הכותב, התאריך או המקום בו נכתב המאמר, אין מדובר בסיכון. אך במידה מדובר ב-Meta Data שמוזרק לקובץ באופן אוטומטי ע"י האפליקציה שבעזרתה כתבנו את הקובץ ואין לנו שליטה עליו (ולרב אנו בכלל לא מודעים שהמידע הוסף), מדובר פה בסיכון גדול מאוד שיכול לא רק לסכן את עמדת הקצה שממנה ביצענו את כתיבת הקובץ, אלא את כלל האירגון.

Meta Data in Microsoft Office

אחת מחבילות התוכנה היותר ידועות לשמצה בנושא השימוש ב-Meta Data "רגיש" היא חבילת ה-Office של חברת Microsoft.

לפי מיקרוסופט, ה-Meta Data העלול להשמר עם כל יצירת/שמירת קובץ הוא:

- Your name
- Your initials
- Your company or organization name
- The name of your computer
- The name of the network server or hard disk where you saved the document
- Other file properties and summary information
- Non-visible portions of embedded OLE objects
- The names of previous document authors
- Document revisions
- Document versions
- Template information
- Hidden text or cells
- Personalized views
- Comments



לצורך ההמחשה, במסגרת כתיבת המאמר ביצעתי חיפוש לקבצי Doc בהאתר Microsoft.com בעזרת גוגל:

Site:www.Microsoft.com FileType:doc

הגעתי לקובץ הבא: http://www.microsoft.com/rus/docs/mpa/eng/mpa.doc

בקובץ הנ"ל ישנו מאמר מעניין מאוד (כן בטח) של Microsoft שכותרתו הוא:

"Microsoft Product Activation"

במידה ואנתח את תוכנו של הקובץ באמצעות התוכנות הנכונות (לצורך המאמר אשתמש בתוכנה FOCA 3 RC שלאחרונה עברה שיפור משמעותי) אוכל לדלות נתונים מעניינים כגון:

- משתמשים שונים שהיו חלק מעריכת הקובץ:

Users	
Username	Leo
Username	Microsoft
Username	SergeyA
Username	Administrator

- מיקומים שונים שבוצע שימוש בהם על המחשב שממנו כתבו את הקובץ:

History	
Author	Leo
Path	G:\All Documents\Work\Design.ru\4 Ginsburg\Microsoft MPA\ENG_MPA.doc
Author	Leo
Path	G:\All Documents\Work\Design.ru\4 Ginsburg\Microsoft MPA\ENG_MPA.doc
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd
Author	Leo
Path	G:\All Documents\Work\Design.ru\4 Ginsburg\Microsoft MPA\ENG_MPA.doc
Author	Leo
Path	G:\Documents and Settings\Administrator\Application Data\Microsoft\Word\AutoRecovery save of ENG_MPA.asd

- כתובת המייל של אחד המשתמשים:

Emails	
Email	sergeya@MICROSOFT.com

- האפליקציה שבאמצעותה נכתב הקובץ, החברה הרשומה ובנוסף-מערכת ההפעלה עליה רצה אותה אפליקציה בעת כתיבת הקובץ:

Other Metadata	
Application	Microsoft Office 2000
Operating system	Windows Server 2000
Company	Microsoft Corp.

איך הנתונים האלה יכולים לעזור לנו בעת פריצה לאירגון? לדוגמא, אני יכול לדעת שקיים סיכוי רב שאם אני אשלח מייל לכתובת: sergeya@MICROSOFT.com

המשתמש שיפתח את המאמר יהיה משתמש עם הרשאות Administrator.

דוגמא נוספת היא הקובץ הבא: (שגם אותו מצאתי באותה הדרך)

http://www.microsoft.com/korea/windows2000/docs/appsvcs_ko.doc

מאמר מאוד מעניין ומומלץ בקוראנית. ניתוח של הקובץ הנ"ל באותה התוכנה מאפשר לנו לדלות מידע נוסף על האירגון:

- משתמשים:

Users	
Username	KwonGee Kwak
Username	BED Web Team
Username	Noelle G. Knapp
Username	Microsoft
Username	박양우
Username	a-noelk

- מיקומים:

Folders	
Folder	C:\WINNT\Profiles\a-noelk\Application Data\Microsoft\Word\
Folder	C:\WINNT\Profiles\a-noelk\Desktop\
Folder	E:\Serapis_Documents\Wtc\Ms\0127\작업 파일\target\appsvcs\
Folder	C:\WINDOWS\바탕 화면\
Folder	\\SERVER3\WORK8\InsungInfo\Projects\WLIS907(MSCOM_W2K RTM)\Work\000121_RTM\DOC\target\appsvcs\
Folder	C:\WINDOWS\TEMP\

(אפשר לשים לב שבמקרה הנ"ל ה-Meta Data כולל כתובת המשויכת למשאב חיצוני ברשת הפנימית באירגון בשם: \\SERVER3)

- אפליקציה ומערכת הפעלה:

Other Metadata	
Application	Microsoft Office 97
Operating system	Windows 98

איך המידע הזה עוזר לנו? אנחנו יכולים לדעת כי לאחד המשתמשים שהשתתף בכתיבת המאמר יש גישה למשאב רשת נוסף: SERVER3. יש סיכוי שבמידה ונצליח להשיג גישה לחשבונו של המשתמש הנכון-נקבל גם גישה לאותו משאב חיצוני. המשאב הזה אולי לא מעניין אותנו כל כך, לעומת זאת מניתוח של הקובץ הבא:

http://www.microsoft.com/colombia/ftpfiles/premier.doc

אנחנו יכולים להבין כי למשתמש הבא:

Users	
Username	IAB

יש גישה לשרת הבא:

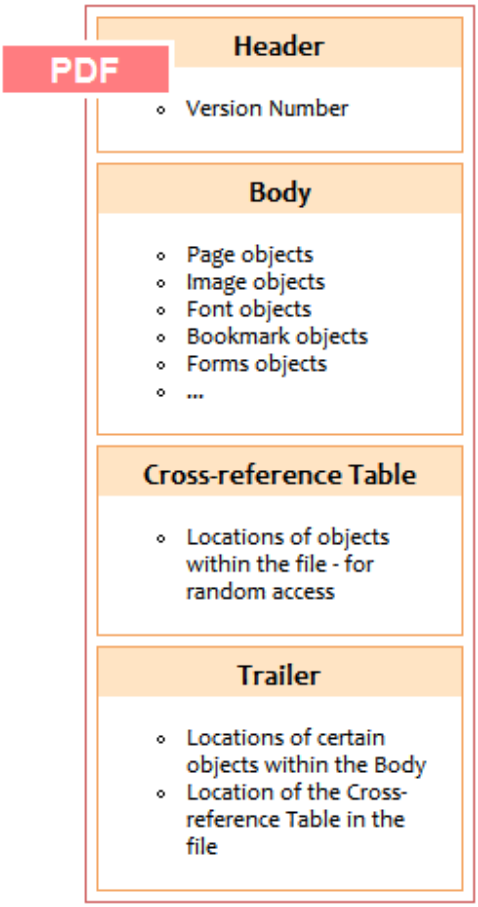
Folders	
Folder	\\MS_PRODUCCION\MSINETPUB\MS-Structure\Colombia\ftpfiles\

ועל-פי שמו אפשר להניח כי מדובר במשאב מעניין יותר.

שימו לב שאם נפתח את קבצי ה-Doc הללו בעזרת Word לא נוכל לראות שום חלק מה-Meta Data שראינו בקובץ עצמו, מפני שהוא מאוחסן מחוץ **לתחום** שבו קוראת Word.

גם בקבצי ה-PDF ישנו Meta Data רב המתווסף באופן אוטומטי וכברירת מחדל לקובץ בעת יצירתו, וגם פה מדובר במידע שיכול לחשוף פרטים רגישים על יוצרו של הקובץ.

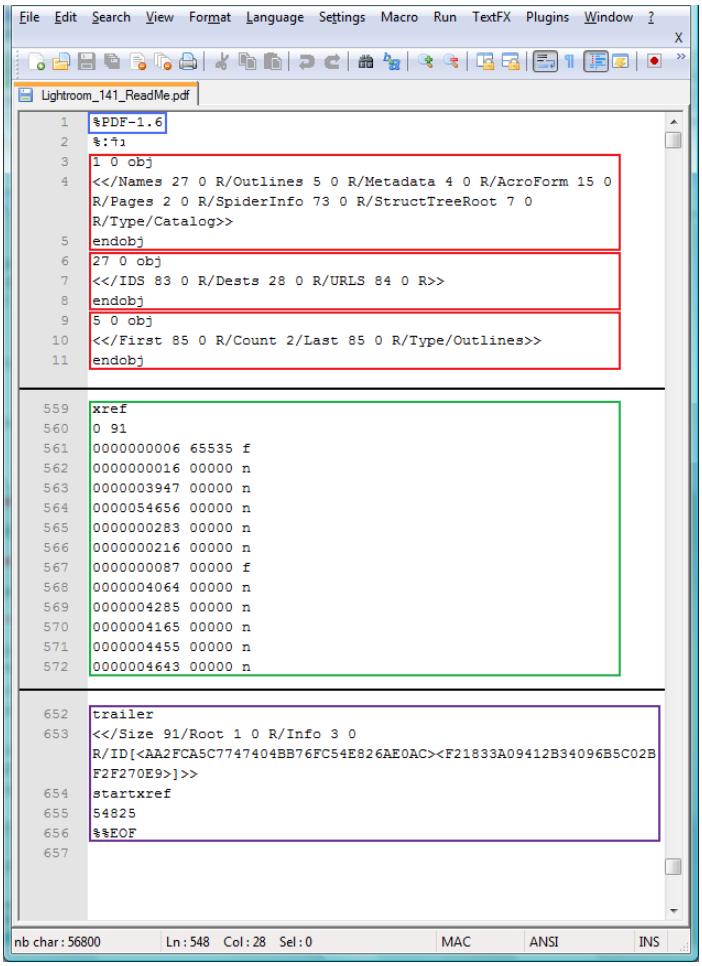
קבצי PDF קצת יותר מסודרים בכל הנושא של ה-Meta Data אך עדיין חושפים פרטים היכולים לעזור לתוקף להבין דבר או שניים על האירגון שלנו. קובץ PDF סטנדרטי בנוי באפן הבא:



(התמונה נלקחה מאתר הבית של gnostice.com)

- בתחילת הקובץ נשמרת גרסאתו (בכדי שהמפענחים ידעו באיזה אופן להציג אותו במקרים של שינוי עתידי במבנה הקובץ)
- לאחר מכן-מגיע גוף הקובץ, בגוף נשמרת רשימת האובייקטים המגדירים את הקובץ ומאפייניו.
- בחלק השלישי של הקובץ מופיעה טבלת ה-Cross-Reference (במבנה הקובץ היא מצויינת כ-xref) הטבלה תכלול את הכתובות של כל אובייקט.
- בחלק האחרון יופיעו נתוני מבנה הקובץ, כמו למשל-כתובתה של טבלת ה-xref או איזה אובייקט הוא האובייקט הראשי.

בכדי לפשט את הנושא, לקחתי קובץ PDF וחילקתי אותו לפני מבנה הנתונים שלו, שימו לב:



- **כחול** – Header.
- **אדום** – Objects List.
- **ירוק** – Cross Reference Table (xref).
- **סגול** – Trailer.

אז איפה בעצם נשמר ה-Meta Data בקובץ ה-PDF?

ה-Meta Data נשמר ב-Body, כחלק מנתוני האובייקטים הבלתי נראים של הקובץ (לא האובייקטים האחראים על מאפייניו-כגון סוג הפונט, צבע, גודל וכו', אלה האובייקטים האחראים על מאפייני הקובץ עצמו).

לאחרונה הנושא צבר תאוצה בעקבות הוויכוח/דיון שהתעורר ב-BugTraq בכל נושא ה-Meta Data בקבצי PDF לאחר ש-Inferno (הבחור מ-SecureThoughts.com) פרסם פוסט בשם:

"Millions of PDF invisibly embedded with your internal disk paths" בפוסט הוא מסביר על בעיה שהוא גילה במנגנון ייצוא הקבצים בפורמט PDF הקיים בדפדפן Internet Explorer. החשיפה היא, שבתהליך הייצוא, הדפדפן מוסיף כ-Meta Data לקובץ ה-PDF את המיקום (Path) של הקובץ על המחשב המקומי, והוא מסביר את התהליך:

1. Pick a .HTM or .HTML or .MHT file on your local computer.

2. Open this file in IE and click Ctrl-P.

(OR Right-click the file in explorer and select PRINT from context menu.)

4. Select any PDF writer as Printer such as Adobe PDF / CutePDF / PrimoPDF / etc.

5. Click Print. When the PDF writer asks for a filename, provide any name.

6. Open the generated pdf in notepad, and search for “file://” without quotes

בכדי להראות עד כמה נרחב השימוש בפונקציה, הוא מציע לבצע את החיפוש:

filetype:pdf file c (htm OR html OR mhtml)

במנוע החיפוש גוגל או במנוע החיפוש בינג.

לדוגמא ל- Meta Data "עוין" בקבצי PDF, ביצעתי את החיפוש הבא:

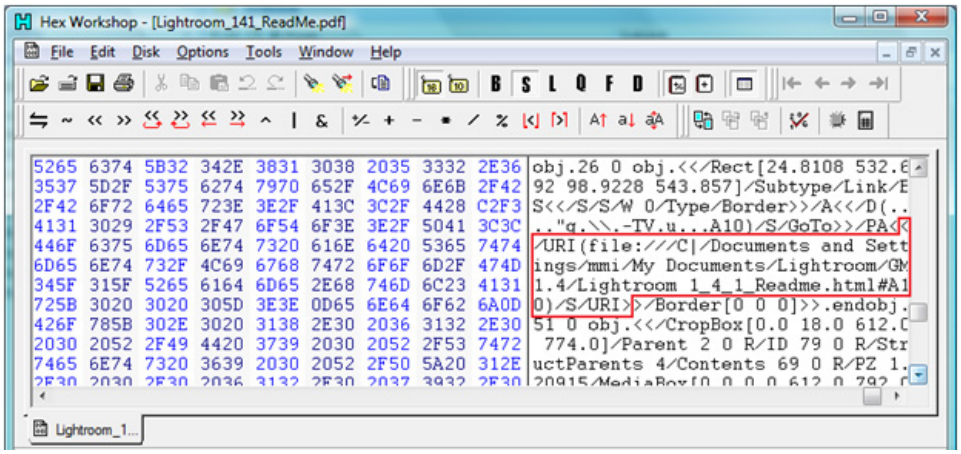
site:www.adobe.com filetype:pdf file c (htm OR html OR mhtml)

התוצאה הראשונה שיצאה היא:

http://www.adobe.com/special/photoshop/Lightroom_141_ReadMe.pdf

(קובץ ה-Read-Me של Photoshop Lightroom)

אם נפתח את הקובץ בעזרת Hex Editor כגון Hex Workshop ולא בעזרת תוכנת PDF Reader כגון Acrobat Reader, נוכל לראות את הדבר הבא:



שימו לב שאפשר לראות כי אכן התווסף ה-Path שבו נשמר הקובץ בעת שמירתו כ-Meta Data, בנוסף שימו לב שלפי ה-Path נוכל לראות את שמו של המשתמש שבו השתמש כותב המאמר.

שימו לב כי לא משנה באיזו תוכנת PDF Writer השתמשו בכדי ליצור את הקובץ, הבעיה אינה נמצאת שם אלה נמצאת במנגנון הייצוא של Internet Explorer.

האם הסיכון ממשי?

ניתוח של קובץ יחיד אומנם לא ייתן לנו מידע יעיל אשר יכול לעזור בפריצה לאירגון מסויים, אך במידה וננתח כמות גדולה של קבצים מאותו אירגון, נוכל לבצע מיפוי (חלקי אך משמעותי) של משתמשיו, כתובות המיילים שלהם, ונגישותם למשאבי הרשת השונים.

לדוגמא, חיפוש בגוגל תחת המחרוזות הבאות:

Site:www.Microsoft.com FileType:doc

Site:www.Microsoft.com FileType:docx

Site:www.Microsoft.com FileType:ppt

Site:www.Microsoft.com FileType:pptx

Site:www.Microsoft.com FileType:xls

Site:www.Microsoft.com FileType:xlsx

Site:www.Microsoft.com FileType:pdf

הביאו ביחד יותר מ-2600 קבצים שניתן לדלות מהם מידע על הרשת הפנימית של האירגון היושב תחת הדומיין: www.Microsoft.com

לנתח כל קובץ באופן ידני אכן בלתי אפשרי כשמדובר במספר קבצים מספיק גדול בכדי לתת לנו תמונה ולו קטנה יחסית לגבי אירגון מסויים. חבר'ה רציניים מ-informatica64 כתבו כלי בשם FOCA RC3 (בקרוב יוצאת הגרסא FOCA 2) שמקבל שאילתה כגון:

Site:www.Domain.com FileType:doc

ויודע לאתר את כל התוצאות המופיעות במנועי החיפוש גוגל ו-בינג, להוריד אותם ולבצע מהם ניתוח של הרשת הפנימית של אותו אירגון הכולל בין השאר:

- שמות משתמשים באירגון.
- כתובות מייל המשמשים את המשתמשים באירגון.
- הרשאות למשאבי רשת מרוחקים (Remote Users/Folders)

- מיקומים של תיקיות על מחשבים/שרתים ספציפיים באירגון.
- מיקומי מדפסות מקומיות/מרוחקות באירגון.
- גרסאות של תוכנות המותקנות על מחשבים ספציפיים באירגון.
- שמות שרתים באירגון.

התוצאות מדהימות-שווה לנסות.

דרכי התגוננות

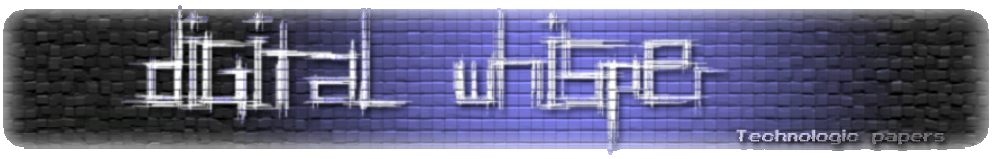
כאשר מדובר באירגון גדול קשה להתגונן מפני הסכנה הזאת, אך במידה ומדובר בקובץ ספציפי עליו רוצים להגן קיימות שתי אפשרויות:

- להשתמש בכלי ייעודי לנושא.
- לדרוס את המידע באופן ידני.

הדרך הראשונה פשוטה ונוחה יותר בדרך כלל, אך אי-אפשר להשתמש בה בכל סוגי הקבצים מפני שלא לכל סוגי הקבצים קיימים כלי "Metadata scrubber". במידה ומדובר בקבצי Office כגון Doc, Ppt, Xls וכו' ניתן להשתמש בתוכנה Remove Hidden Data מבית מיקרוסופט, או בתוכנה Doc Scrubber למסמכי Word.

במידה ואנו משתמשים בקבצים שאין להם תוכנה ייעודית- או שנעדיף לבצע זאת ידנית בכדי לחסל כל סיכוי ל-Meta Data פשוט נפתח את הקובץ באמצעות כל כלי Hex Editor וכך נוכל לערוך את כל ה-Meta Data הקיים בקובץ למה שנרצה.

כאשר מדובר בארגון, נהלי חברה מסודרים יכולים לצמצם את הבעיה. למשל – נוהל עבור אחראי האתר המחייב לבצע דריסת meta data בכל קובץ Word העולה לאתר התוכן של האירגון.



סיכום

השימוש ב-Meta Data חשוב מאוד ובמקרים מסויימים הוא יכול ליעל את אירגון המידע והנגישות אליו באופן משמעותי, אך חשוב מאוד לזכור כיצד נכון להשתמש בו, לא בכל המקרים ולא בכל סוג של Meta Data הדבר נכון. אנו חייבים תמיד לדעת איזה מידע אנחנו משתפים עם שאר האינטרנט והאם המידע הזה יוכל לשמש כנגדנו באחד הימים.