Chema Alonso, José Palzón

Tactical Fingerprinting using metadata, hidden info and flown lost data using FOCA

2003 – MS Word bytes Tony Blair

Microsoft Word bytes Tony Blair in the butt

Home > Privacy > **Blair's Iraq Dossier**

Richard M. Smith (rms@computerbytesman.com)
June 30, 2003

Microsoft Word documents are notorious for containing private information in file headers which people would sometimes rather not share. The l
hard way.

Back in February 2003, 10 Downing Street published a dossier on Iraq's security and intelligence organizations. This dossier was cited by Colin
Dr. Glen Rangwala, a lecturer in politics at Cambridge University, quickly discovered that much of the material in the dossier was actually plagiar

You can read Dr. Rangwala's original analysis of the dossier from Feb. 5, 2003 at this URL:

http://www.casi.org.uk/discuss/2003/msg00457.html

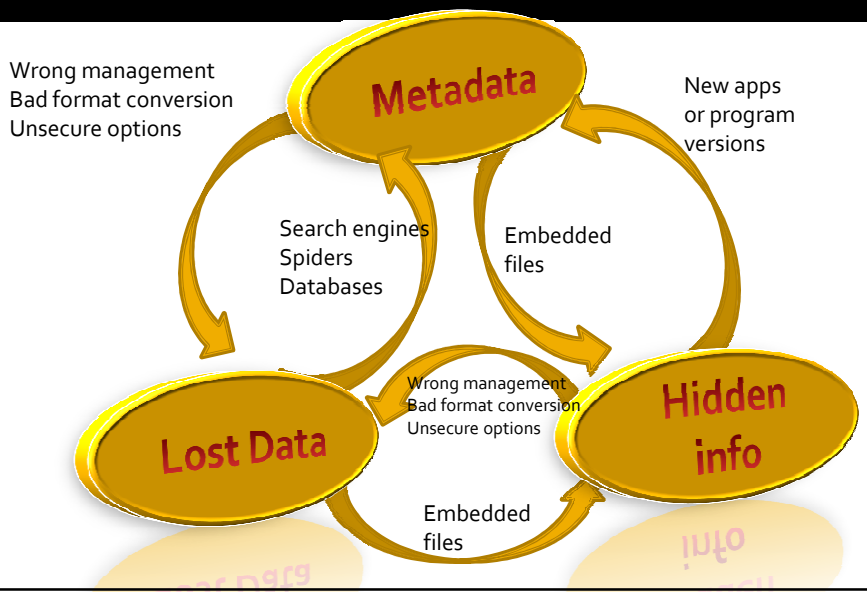
Blair's government made one additional mistake: they published the dossier as a Microsoft Word file on their Web site. When I first heard from I
had worked on the document. I downloaded the Word file containing the dossier from the 10 Downing Street Web site (http://www.number-10.

```
Rev. #1: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.asd"
Rev. #2: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.asd"
Rev. #3: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.asd"
Rev. #4: "JPratt" edited file "C:\TEMP\Iraq - security.doc"
Rev. #5: "JPratt" edited file "A:\Iraq - security.doc"
Rev. #6: "ablackshaw" edited file "C:\ABlackshaw\Iraq - security.doc"
Rev. #7: "ablackshaw" edited file "C:\ABlackshaw\A;Iraq - security.doc"
Rev. #8: "ablackshaw" edited file "A:\Iraq - security.doc"
Rev. #9: "MKhan" edited file "C:\TEMP\Iraq - security.doc"
Rev. #10: "MKhan" edited file "C:\WINNT\Profiles\mkhan\Desktop\Iraq.doc"
```

What kind of data can be found?

- Metadata:
  - Information stored to give information about the document.
    - For example: Creator, Organization, etc..
- Hidden information:
  - Information internally stored by programs and not editable.
    - For example: Template paths, Printers, db structure, etc…
- Lost data:
  - Information which is in documents due to human mistakes or negligence, because it was not intended to be there.
    - For example: Links to internal servers, data hidden by format, etc…

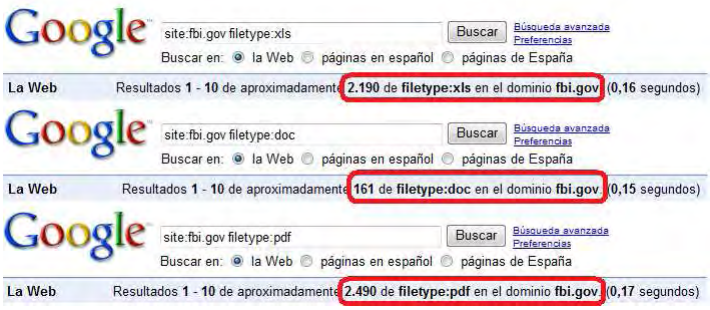
Metadata Lifecycle



So… are people aware of this?

- The answer is NOT.
- Almost nobody is cleaning documents.
- Companies publish thousand of documents without cleaning them before:
  - Metadata.
  - Hidden Info.
  - Lost data.

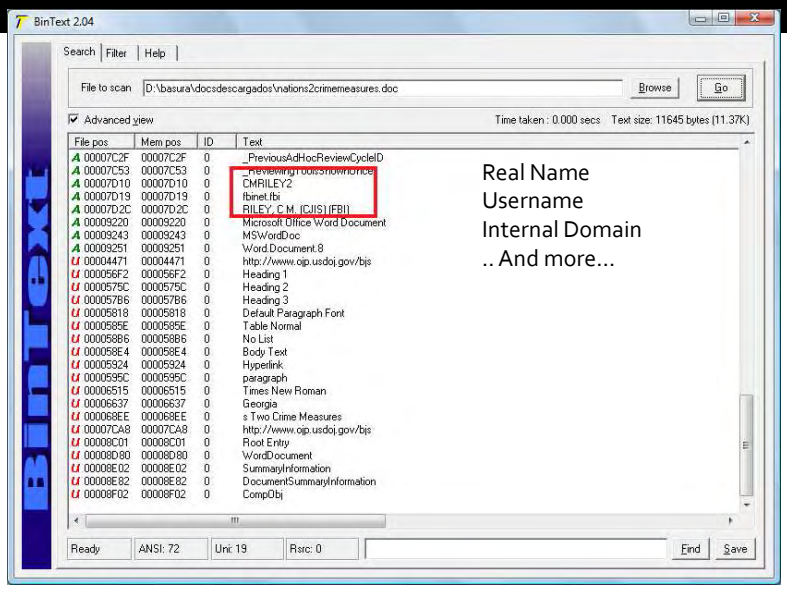
Sample: FBI.gov



Total: 4841 files

Demo: Looking for info in public FBI files

Are they clean?



Real Name
Username
Internal Domain
.. And more...

Sample: ec.europe.eu



Total: 896 files

Demo: Looking for info in public ec.europe.eu files

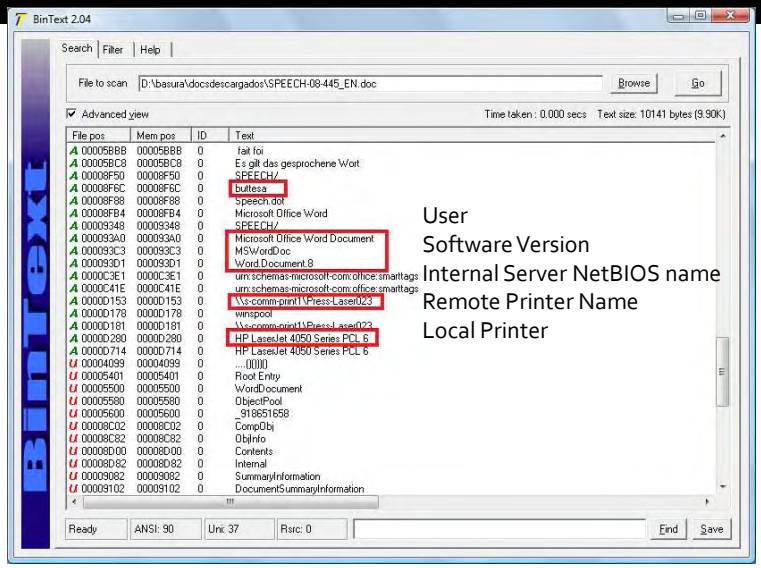


Are they clean?



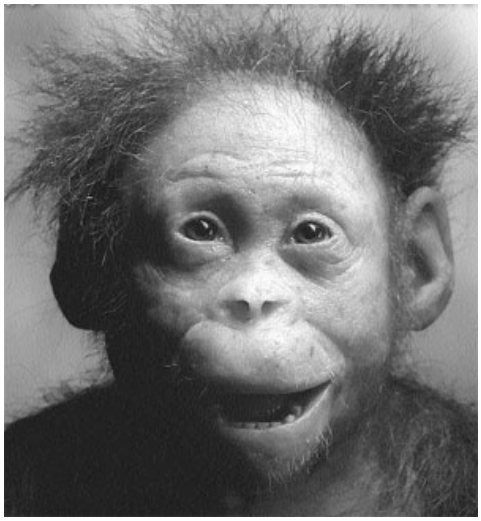
Total: 1075 files

Are they clean?



User
Software Version
Internal Server NetBIOS name
Remote Printer Name
Local Printer

How many files is my company publishing?

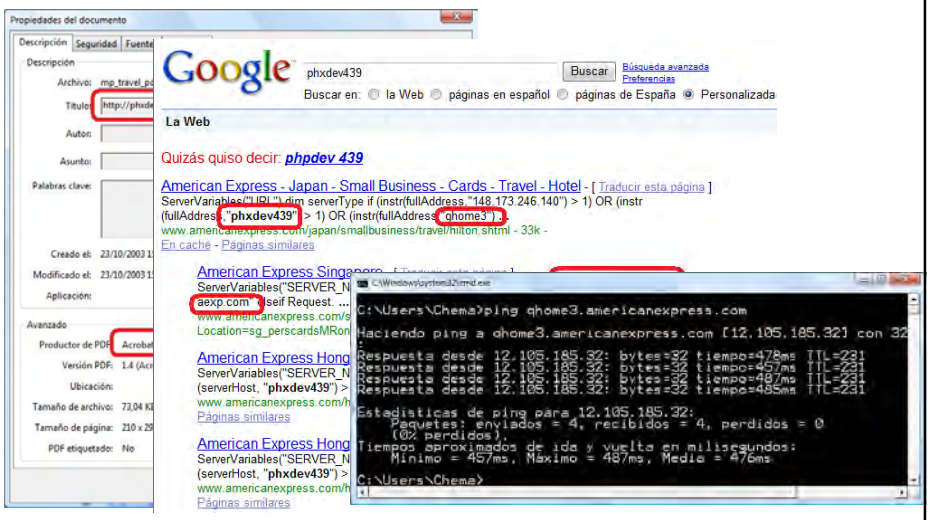


Not only Microsoft Office documents.

Sample: Printer info found in odf files returned by Google

```
$ grep Impresora resultados.txt | grep \\\\
10.odt  Impresora       \\srvprint\VISA28_329LC3
12.odt  Impresora       \\servidor\HP 2000C
19.odt  Impresora       \\10.177.1.126\EPSON EPL-6200 Advanced
23.odt  Impresora       \\1-215-01\hp LaserJet 1000
34.odt  Impresora       \\srv2.rm.oasi-servizi.it\RICOH roma corriodio nord
3.odt   Impresora       \\filserver\4100
63.odt  Impresora       \\gamma\HP Deskjet F300 series
27.ods  Impresora       \\Silegio\HP psc 2350 series
33.ods  Impresora       \\ABBY\BigGrayUglyThing
40.ods  Impresora       \\MUSTAFA\Samsung ML-1710 Series
46.ods  Impresora       \\http://10.0.0.42:631\Chaucer
80.ods  Impresora       \\192.168.250.1\HP OfficeJet G55
86.ods  Impresora       \\WLGPRINT1\L10 GA T632
89.ods  Impresora       \\B-THOMAS\hp LaserJet 1300 PCL 6
17.odp  Impresora       \\tatra\Minolta Di152 PCL6
22.odp  Impresora       \\10.177.1.126\EPSON EPL-6200 Advanced
4.odp   Impresora       \\Verra\LEMURIA
55.odp  Impresora       \\CHEGUEVARA\lp_serv
57.odp  Impresora       \\sm-ubur01-01.east.sun.com\bur116
69.odp  Impresora       \\ipp://iprint.innerweb.novell.com\prv-mktg-hp8000dn-2
86.odp  Impresora       \\EFOITDP02OS4\hp4100dtn
$
$ grep iMac resultados.txt
49.ods  Impresora       hp psc 2500 series @ Luisa's iMac
$
```

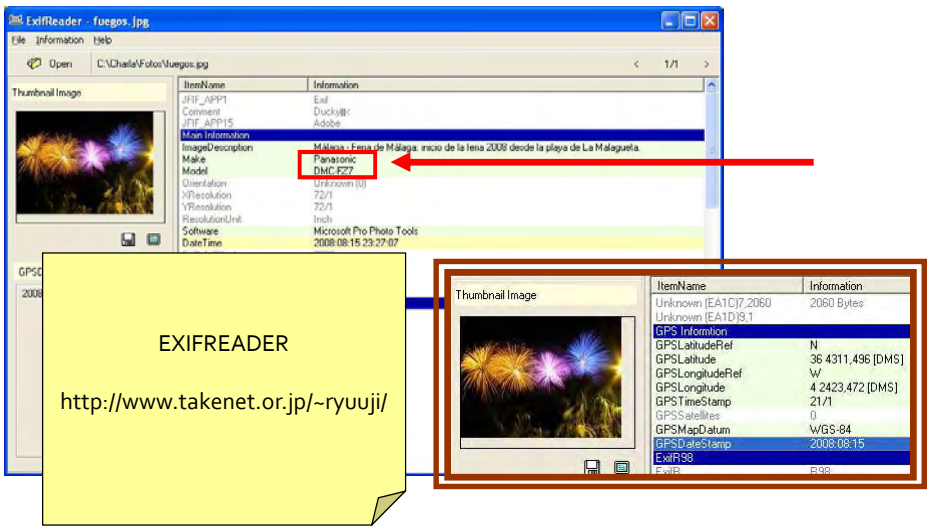
Sample: Info found in a PDF file



What files store Metadata, hidden info or lost data?

- Office documents:
  - Open Office documents.
  - MS Office documents.
  - PDF Documents.
    - XMP.
  - EPS Documents.
  - Graphic documents.
    - EXIFF.
    - XMP.
  - And almost everything….

Pictures with GPS info..



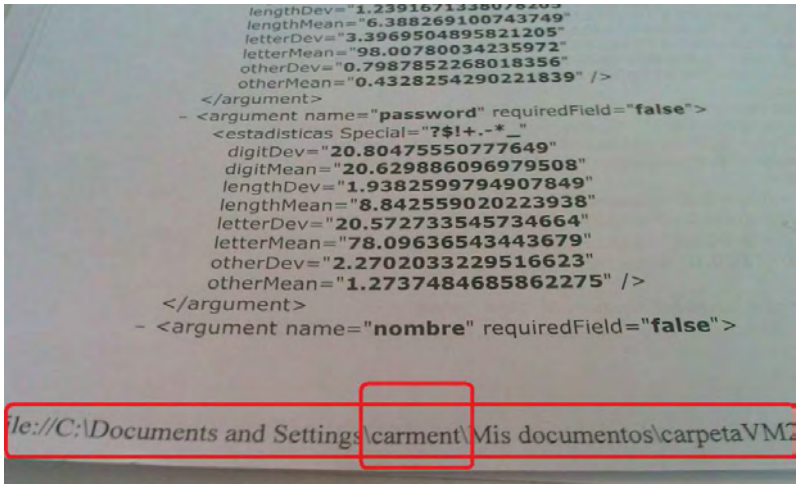
Demo: Looking for EXIF information in ODF file

Even Videos with users…

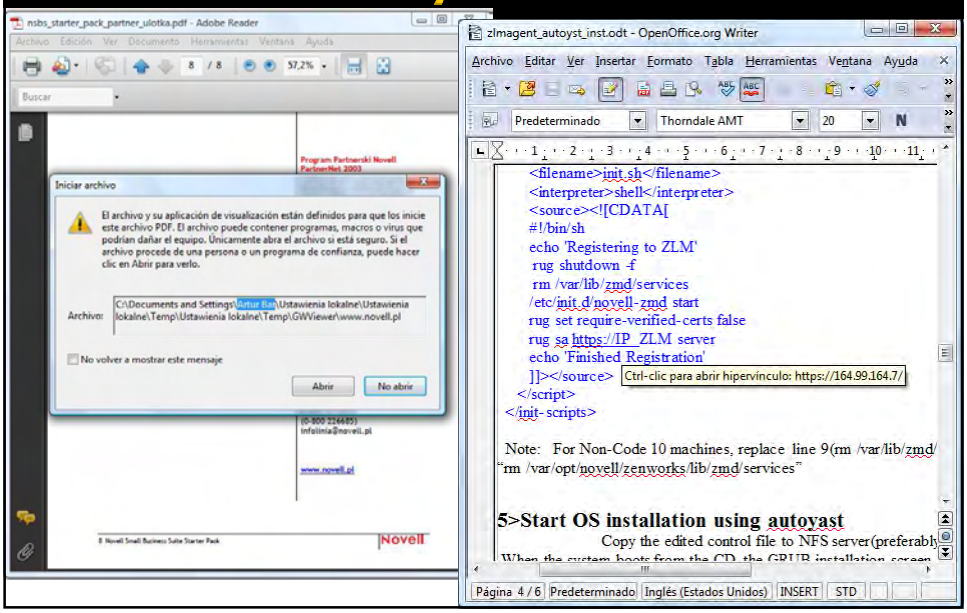


http://video.techrepublic.com.com/2422-14075_11-207247.html

And of course, printed txt



Lost data everywhere



Public server



What can be found? (I)

- Users:
  - Creators.
  - Modifiers .
  - Users in paths.
    - C:\Documents and settings\jfoo\myfile
    - /home/johnnyf
- History of use.
- Operating systems.
- Software versions.
- Paths.
  - Local and remote.
- Network info.
  - Shared Printers.
  - Shared Folders.
  - ACLS.

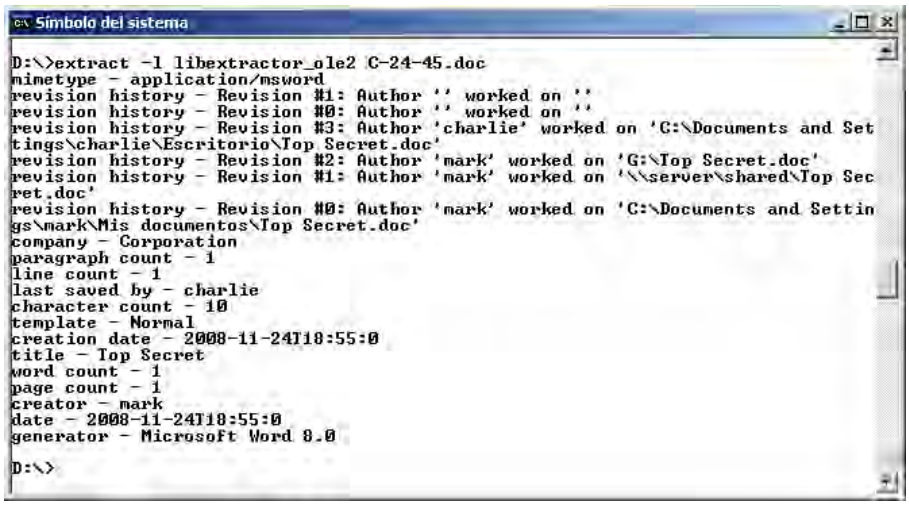
What can be found? (II)

- Printers.
  - Local and remote.
- Internal Servers.
  - NetBIOS Name.
  - Domain Name.
  - IP Address.
- Database structures.
  - Table names.
  - Colum names.
- Devices info.
  - Mobiles.
  - Photo cameras.
- Private Info.
  - Personal data.

How can metadata be extracted?

- Info is in the file in raw format:
  - Binary.
  - ASCII .
- Therefore Hex or ASCII editors can be used:
  - HexEdit.
  - Notepad++.
  - Bintext
- Special tools can be used:
  - Exif redaer
  - ExifTool
  - Libextractor.
  - Metagoofil.
  - ...
- ...or just open the file!

Tools: Libextractor



Tools: MetaGoofil

http://www[redacted]doc/47[redacted].doc

Local copy Open

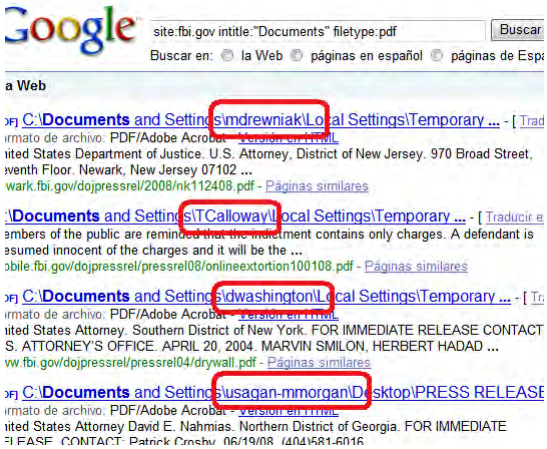
Important metadata:

```
mimetype - application/msword
revision history - Revision #9: Author 'Jennifer Romeo' worked on 'romeo HD:Documents:Articles & News:Articles_PM
revision history - Revision #8: Author 'jwlater' worked on '\\KGINTS\[redacted]hared\PMSEP\PMSEP 6\Conference Structur
revision history - Revision #7: Author 'John Newcomb' worked on 'C:\My Documents\Scientific Man\Current Programs\
revision history - Revision #6: Author 'Edutech User' worked on 'C:\My Documents\PMSEP Detailed Structure 8-9-02.
revision history - Revision #5: Author 'Edutech User' worked on 'C:\My Documents\PMSEP Detailed Structure 8-9-02.
revision history - Revision #4: Author 'Edutech User' worked on 'C:\My Documents\PMSEP Detailed Structure 8-7-02.
revision history - Revision #3: Author 'Edutech User' worked on 'C:\My Documents\PMSEP Detailed Structure 8-7-02.
revision history - Revision #2: Author 'Edutech User' worked on 'C:\WINDOWS\Temporary Internet Files\OLK4\PMSEP 5
revision history - Revision #1: Author 'Edutech User' worked on 'C:\Documents and Settings\krobinson\Application
revision history - Revision #0: Author 'Edutech User' worked on 'C:\Documents and Settings\krobinson\Application
language - U.S. English
paragraph count - 13
line count - 55
last saved by - Jennifer Romeo
character count - 6717
template - Normal
creation date - 2002-08-12T14:24:00Z
title - CONFERENCE STRUCTURE
word count - 1178
page count - 5
creator - John Newcomb
date - 2002-08-12T14:24:00Z
generator - Microsoft Word 9.0
```

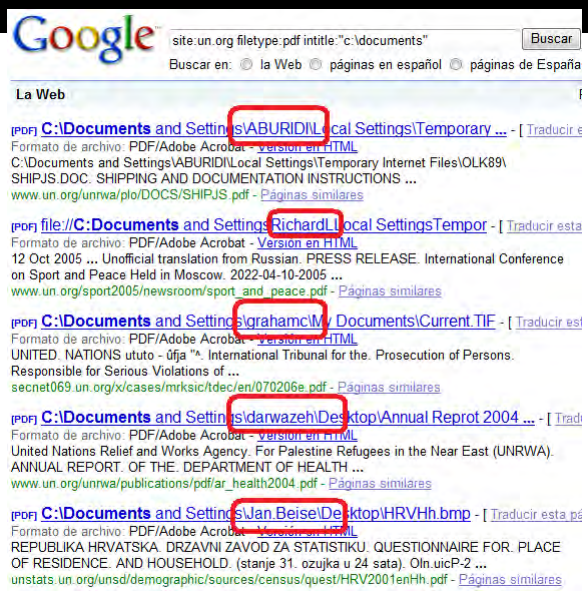
- http://www.edge-security.com/metagoofil.php

Yes, also Google….

Your FBI user



Your UN user



Your Scotland Yard user



Google site:police.uk intitle:"c:\documents and" filetype:pdf Buscar

Buscar en: la Web páginas en español páginas de España Per

La Web Resultados 1

[PDF] C:\Documents and Settings\Administrator\Desktop\2005-2006 ... - [Traducir esta pág

Formato de archivo: PDF/Adobe Acrobat - Versión en HTML

Page 1. ¬g s. tZ h. Á ª e g! ¤á.]. Á. 5. Ã. Ú" ã. "{ g. öZ r. äZ. Y. " gs. ™Z. C ì. Xt. 5. (latest good. Y. Zk. § b. »g. Á. ˆ ì. "Z k. @° ì{. F,. +Z. Y ...

www.derbyshire.police.uk/sei/s/3221/f9.pdf - Páginas similares

[PDF] C:\Documents and Settings\TnlCom\My Documents\Q05-267 Urdu[C5-0447 ... - [Tr

Formato de archivo: PDF/Adobe Acrobat - Versión en HTML

Page 1. 6, y. <. XZ. ¤/. W. \. Ã. Ë. `x. Æ !" g. } ¥ á]. 0845 113 5000. Ã y. <. X.) u ò». ² c. 999 u ò ßg. ‹ qw. Ô. 6, y. <. Xt. Ü y». ² j. Z. zgW. \. ÃZ ...

www.west-midlands.police.uk/PDF/contact-us/language/URDU.PDF - Páginas similares

[PDF] C:\Documents and Settings\p4510272\Local Settings\Temporary ... - [Traducir esta p

Formato de archivo: PDF/Adobe Acrobat

13-JAN-2000 04:2B FROM: TO:29S4P0304 p.m. Appendix A. Reference No: 165 - 06 - 077. Redacted copy of a receipt dated 17th October 2005, ...

www.surrey.police.uk/foi/Chief_Constables_Expenses_Appendix_A_077.pdf - Páginas similares

Your Carabinieri user

Google site:carabinieri.it filetype:pdf intitle:"Documents" Buscar

Buscar en: la Web páginas en español páginas de España

La Web

[PDF] C:\Documents and Settings\mazzolac\Documenti\Digitalizza1043.tif - [Traducir esta

Formato de archivo: PDF/Adobe Acrobat - Versión en HTML

TED-2007-61255-IT Pagina 1 di 2. Documento 2007-61255-IT. 13/03/2007 S50 Comunità europee - Forniture - Appalti aggiudicati. I. II. IV. V. VI. ...

www.carabinieri.it/NR/rdonlyres/4A795C84-5523-45FE-A1C3-5CDA4526350B/3057/FORAGGIO.pdf - Páginas similares

[PDF] C:\Documents and Settings\mazzolac\Documenti\Digitalizza1042.tif - [Traducir esta

Formato de archivo: PDF/Adobe Acrobat - Versión en HTML

TED-2007-81265-IT Pagina 1 di 3. Documento 2007-81265-iT. 05/04/2007 S67 Comu - Forniture - Appalti aggiudicati. I. 11. IV. V. VI. ...

www.carabinieri.it/NR/rdonlyres/829B5072-15C9-45CE-9485-662859109F08/3056/092INFORMATICO.pdf - Páginas similares



Your WhiteHouse user



Yes, we can!



Of course, your Novell user



Drawbacks

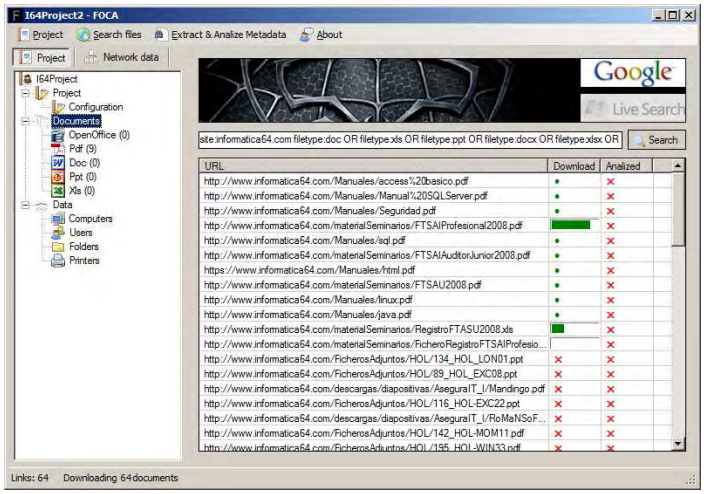
- These tools only extract metadata.
- Not looking for Hidden Info.
- Not looking for lost data.
- Not post-analysis.

Foca

- Fingerprinting Organizations with Collected Archives.
  - Search for documents
  - Automatic file downloading
  - Capable of extracting Metadata, hidden info and lost data.
  - Cluster information
  - Analyzes the info to fingerprint the network.

Demo: FOCA

FOCA: Gathering Files



FOCA Online

http://www.informatica64.com/FOCA

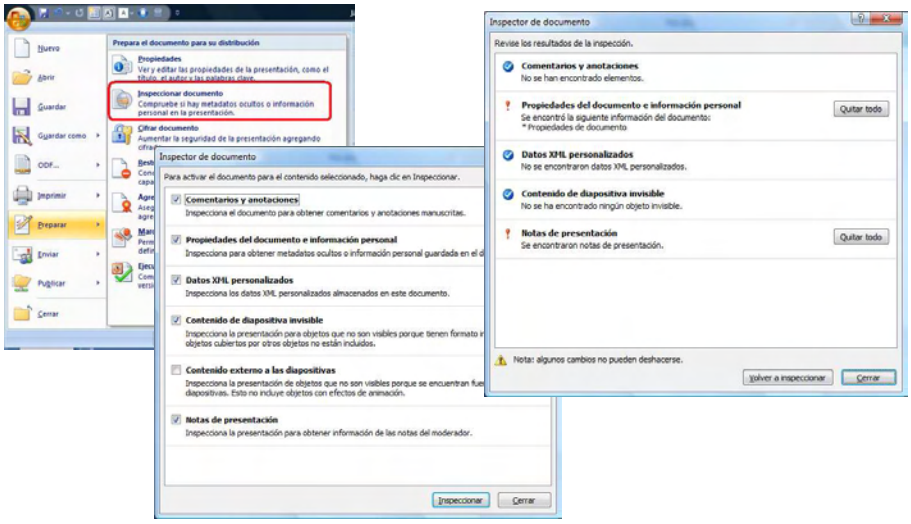


Solutions?

First: Clean all public documents



Clean your documents: MSOffice 2k7



Clean your documents: MSOffice 2k3 & XP

http://www.microsoft.com/downloads/details.aspx?displaylang=en&FamilyID=144e54ed-d43e-42ca-bc7b-5446d34e5360

Office 2003/XP Add-in: Remove Hidden Data

Brief Description

With this add-in you can permanently remove hidden data and collaboration data, such as change tracki PowerPoint files.

On This Page

- Quick Details
- System Requirements
- Related Resources
- Overview
- Instructions
- What Others Are Downloading

Download

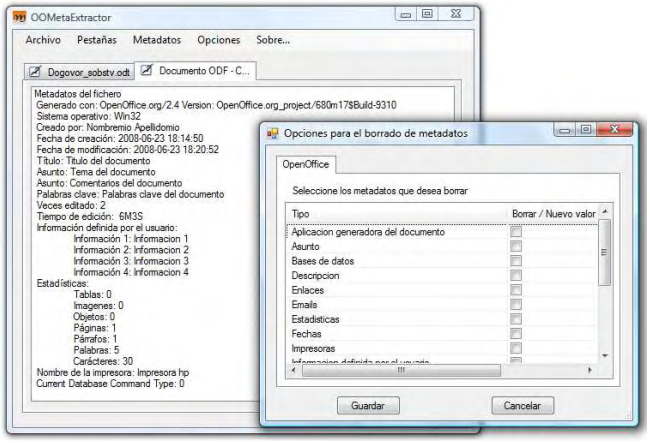
Quick Details

File Name:	rhdtool.exe
Version:	1.1
Date Published:	7/8/2008
Language:	English
Download Size:	260 KB
Estimated Download Time:	Dial-up (56K) 1 min
Change Language:	English Change

Demo: OpenOffice "Security" Options…

Cleaning documents

- OOMetaExtractor

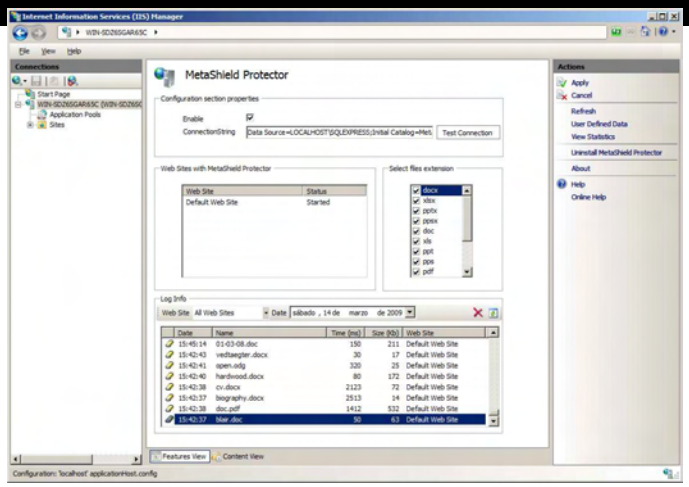


http://www.codeplex.org/oometaextractor

Are you safe relying on your users?



IIS MetaShield Protector

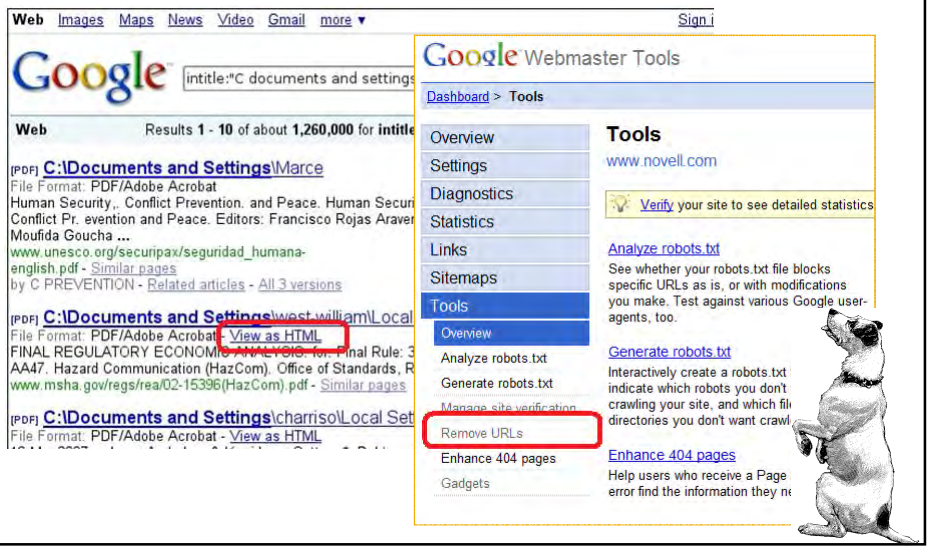


http://www.metashieldprotector.com

MetaShield Protector

Demo: Cleaning Documents

Second: Beg Google to delete all the cached files



Don´t trust your users!!!



Don´t complain about your job!!



Questions?

- Authors
  - Chema Alonso
    - chema@informatica64.com
  - Enrique Rando
    - Enrique.rando@juntadeandalucia.es
  - Alejandro Martín
    - amartin@informatica64.com
  - Francisco Oca
    - froca@informatica64.com
  - Antonio Guzmán
    - antonio.guzman@urjc.es

PS: This file also has metadata

