

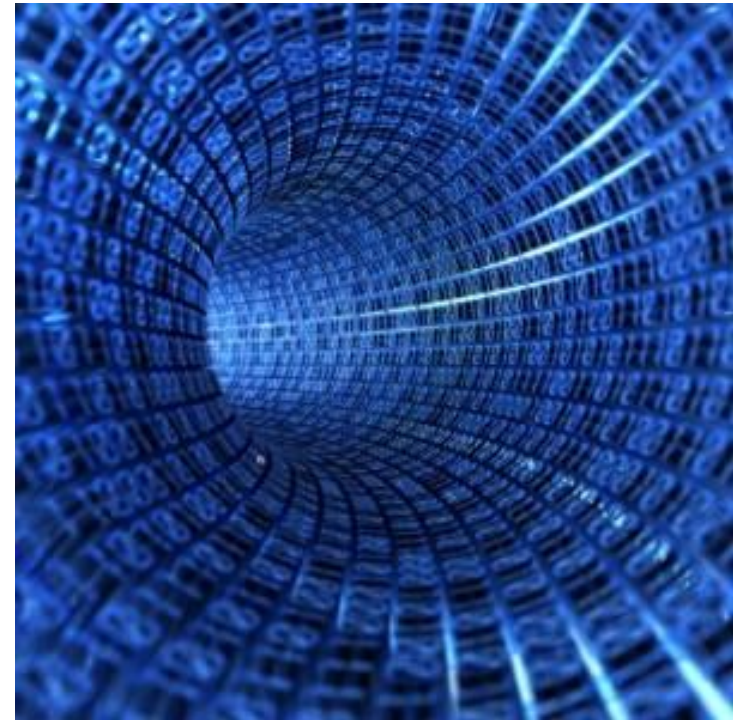


Discovery of Electronically Stored Information

ECBA conference

Tallinn

October 2012



Agenda

Introduction

eDiscovery investigation

- Collection
- Pre-processing
- Processing
- Review

Example projects

Key takeaways

Introduction

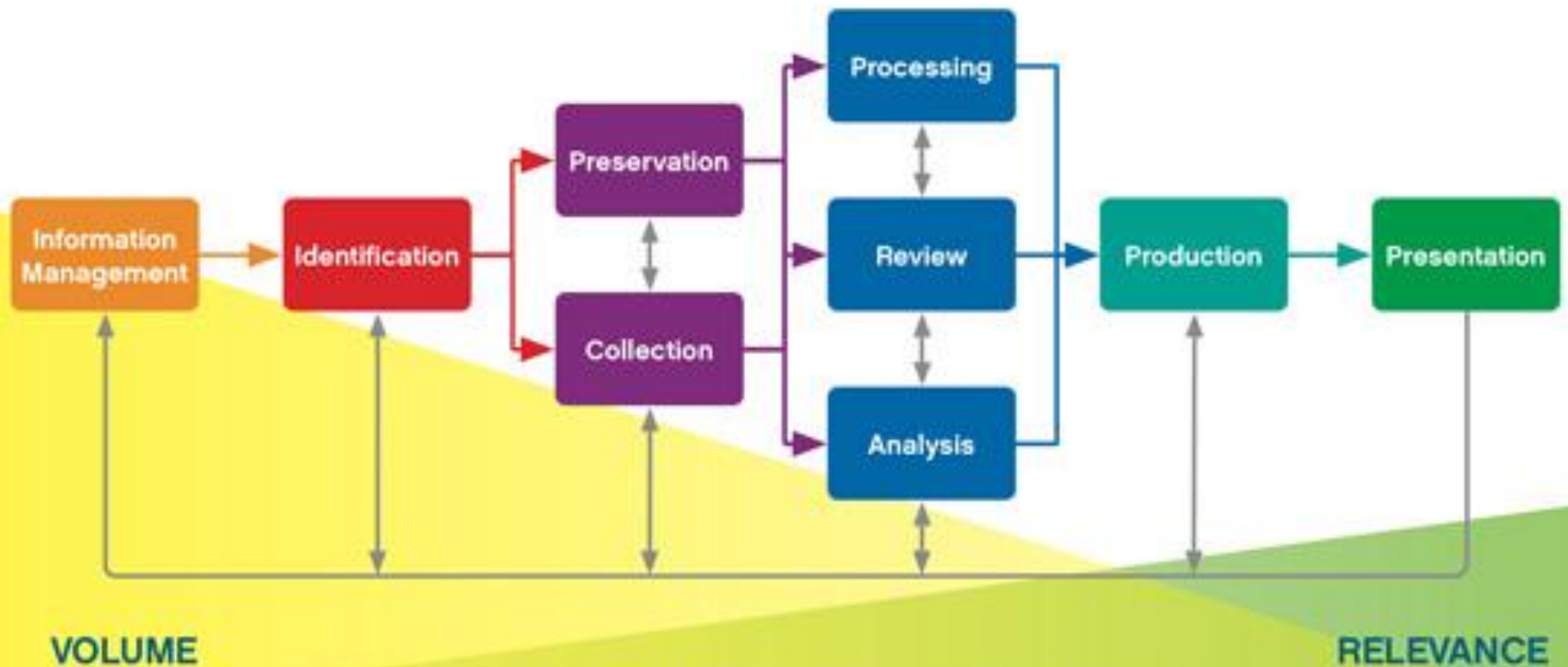
Digital world

- **More than 95% of the digital universe is unstructured data**
- **Enterprises have some liability for 80% of information in the digital universe at some point of its digital life**
- **90% of the data in the world today has been created in the last two years**

Source: IDC, IBM

eDiscovery reference model

Electronic Discovery Reference Model



Electronic Discovery Reference Model / © 2009 / v2.0 / edrm.net

eDiscovery investigation

eDiscovery investigation

- We will focus on the following phases:
 - Collection
 - Pre-processing
 - Processing
 - Review
- Common **drivers**
 - Regulatory – FCPA (Foreign Corrupt Practices Act)...
 - Fraud/incident investigations
- This presentation focuses on **unstructured data**

Collection phase

Basics of Collection phase

- **Collecting data in a forensically sound manner**
 - **Complete and accurate data acquisition**
 - **No changes to the original evidence**
 - **Documentation**
- Proven methodology and experienced staff essential
- Special software and hardware equipment needed
- Planning and logistics important, often work under time pressure
- Information identification – custodians, media, periods, priorities...

Storage types

- Custodian (user) computers – desktops, notebooks
- Server computers – server mailboxes, home folders, shared folders, e-rooms
- CDs, DVDs
- External hard disk drives
- Flash drives, memory cards
- Tape backups
- Hard copy documents
- Handheld devices
- Other (cloud...)

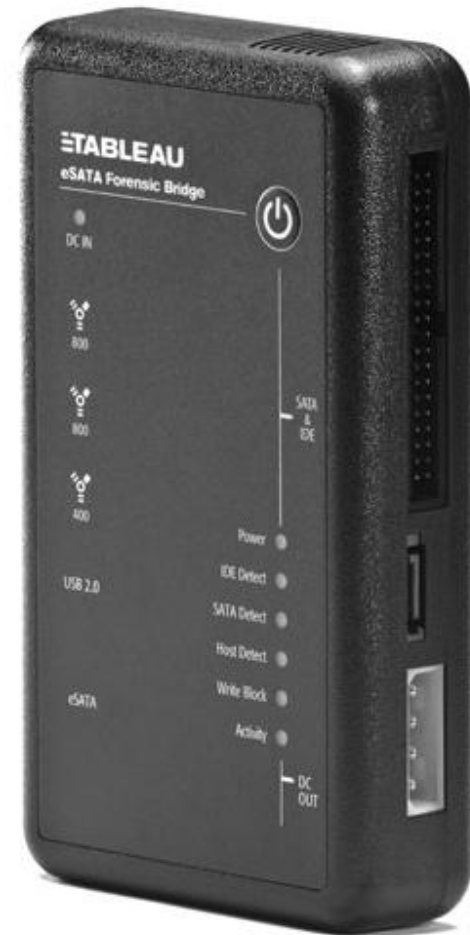
Example – HDD imaging

- Imaging a notebook hard disk drive using a hardware write-blocker and EnCase Forensic software suite
- Obtaining a bit-by-bit copy (exact replica) of entire hard disk drive data, including unallocated space
- Storing image files within an encrypted container



Collection methods

- Using **hardware forensic write-blockers**
- Using **standalone forensic duplicators**



Collection methods

- Using custom **live Linux distributions**
- Making a “**live image**” – using the subject computer, with OS booted – least desirable but sometimes only practical
 - RAID
 - Servers that cannot be turned off due to operational reasons
- Using other software tools
- Using third party service providers
 - Hard copy documents – scanning and OCRing
 - Tape backups

Handheld devices

- Special tools required



Legal issues

- Although there is a EU directive on data privacy protection the actual data privacy rules are different in EU member countries
- Legal consultations required
- Prior consent of custodians often required. This can lead to evidence being lost due to actions by custodians, but in practice the risk is mitigated by the following:
 - Custodians formally asked to retain all relevant data
 - Possibility to detect mass deletions and detect use of anti-forensics tools
 - Recovery of deleted items
 - Email communication distributed among multiple custodians
 - Personal backups

Pre-processing phase

Introduction

- **Objectives**
 - **Reduce volume of data – include only relevant data for further processing** (email, Office files...)
 - **Recover deleted data**
 - **Preserve the data and metadata**
- **Inclusive/exclusive approach**

Pre-processing steps

- **Recovering folders**
 - FAT partitions – searches unallocated space for signatures of a deleted folder, can rebuild files and folders that were within that deleted folder
 - NTFS partitions – searches unallocated space for MFT records and scans current MFT for files without parent folders, can recover the identified files
- **File signature analysis** – comparing file extensions with file headers
- **Expanding compound files** such as archives – based on file extensions and/or signatures
- **Filtering only relevant data** based on file extensions and/or signatures and other criteria
- **Exporting the resulting set of files**

Decryption

- Accessing corporate user computers with **full volume encryption** using appropriate credentials
- Encrypted files can be handled on an individual basis
 - Getting harder as the applications use stronger encryption methods (Office Word and Excel 97-2003 versus 2007)
 - Password is the weakest link
 - Distributed processing, GPU acceleration, use of indexed words

Digging deeper

- Recovery of deleted partitions
- Carving data from unallocated space
- Timeline analysis (e.g. mass deletion detection, detection of gaps in email)
- Exploring Windows artifacts
 - Identification of wiping software
 - Recycle bin files
 - Attached external storage device history
 - Web browsing history
 - Prefetch analysis
 - Recent document history, file browsing history
 - Encrypted files

Processing phase

Processing

- **The processing platform extracts the text and metadata from the input data set so these can be uploaded to the review platform**
- Some **features** of a processing platform
 - Reliable and extensive support of input data file formats
 - Extraction of metadata
 - Processing of attached/embedded data
 - Speed, scalability
 - De-duplication
 - Filtering (date ranges, search terms)
 - Limited recovery of deleted items (e.g. PST containers)

De-duplication and De-NIST

- Global or per custodian
- **Significantly reduces the volume of data for further processing** (40-60% typical) but relevant information can be lost (ie. Due to lost context from the storage path)
- De-NIST excludes tens of thousands known files (based on hash value, usually executables)

Processing platform – an example (NUIX)

The screenshot displays the Nuix Enterprise Workstation 3.6.0 interface. The main window is titled "Demo1 - Nuix Enterprise Workstation 3.6.0" and shows a search results view. The search criteria are "No date filter", "Jul 23, 1993", and "Mar 22, 2012". The results are displayed in a table with columns for Name and File Type. The selected item is "Karen Pope" (Microsoft Outlook Property...). The interface also includes a Document Navigator on the left, a Preview pane on the right, and a Review and Tag section at the bottom.

Document Navigator: Evidence (60536/60536 hits; 100.00%)
Demo1 (60,536)
Evidence 1 (55,593)
BPARKER.E01 (55,592)
Evidence 2 (4,927)
darron_c_giron_002.pst (4,926)
Evidence 3 (16)
TEST.pst (15)

Results: View by: Results | Immaterial Items: Hide | Deduplication: None

Name	File Type
Karen Pope	Microsoft Outlook Property...
NG Energy	Microsoft Outlook Property...
RE: Saturday	Microsoft Outlook Property...
UtiliCorp Gas Hedge	Microsoft Outlook Property...
RE: What's up	Microsoft Outlook Property...
NG Energy	Microsoft Outlook Property...
BB	Microsoft Outlook Property...
RE: Saturday	Microsoft Outlook Property...
办理减免税发★票13928472421	Microsoft Outlook Note
PDQ	Microsoft Outlook Property...
Kent Kopetzy	Microsoft Outlook Property...
Kent Kopetzy	Microsoft Outlook Property...
[Unnamed System File]	Microsoft Outlook Property...
Kent Kopetzy	Microsoft Outlook Property...
[Unnamed System File]	Microsoft Outlook Property...
PDQ	Microsoft Outlook Property...
PDQ	Microsoft Outlook Property...
Kot_9827.gif	Compuserve Graphic Inter...
Utilicorp	Microsoft Outlook Property...
Utilicorp	Microsoft Outlook Property...
10/24	Microsoft Outlook Property...
\$MFT	Unknown Binary File
Utilicorp	Microsoft Outlook Property...
Colorado	Microsoft Outlook Property...
10/24	Microsoft Outlook Property...
10/24	Microsoft Outlook Property...
办理减免税发★票13928472421	Microsoft Outlook Note

Displaying 23,854 items. (60,536 total > 36,682 immaterial items hidden)

Preview: Karen Pope | Path: Evidence 2 → darron_c_giron_002.pst → Orphaned Items | Duplicates: Exact (0) Near (2) | Similar items: High (12742) Medium (12742) Low (12742) | Text | Family (1) | Metadata | PDF | Native | Word List | Details: 3 lines

Metadata:
Created: May 12, 2009 2:17:30 AM | Author:
Last Modified: Apr 19, 2010 7:28:58 AM | Company:
Last Accessed:
Title:

Properties:
Date: Mon, 10 Sep 2001 21:56:25 +0200
File Created: Tue, 12 May 2009 02:17:30 +0200
File Modified: Mon, 19 Apr 2010 07:28:58 +0200
From: giron
Mapi-7201: 18
Mapi-7202: Tue, 12 May 2009 02:17:30 +0200
Mapi-7203: HHANQ3WRDOUT4T4HQNCWMVXZYADXJBACB
Mapi-7206: 0
Mapi-Access: 2
Mapi-Access-Level: 0
Mapi-Client-Submit-Time: Mon, 10 Sep 2001 21:56:25 +0200
Mapi-Conversation-Topic: Karen Pope
Mapi-Display-Bcc:
Mapi-Display-Cc:
Mapi-Display-To:
Mapi-Hasattach: false
Mapi-Importance: 0
Mapi-Internet-Cpid: 20127
Mapi-Message-Class: IPM.Note
Mapi-Message-Delivery-Time: Mon, 10 Sep 2001 21:56:25 +0200

Review and Tag: Karen Pope | Click here to configure Tags for this case.

Review phase

Review

- **The review platform makes information available for review, analysis and productions**
- Some **features** of a review platform
 - Document review – presentation, tagging...
 - Web-based access of multiple reviewers
 - Handling large volumes of data
 - Indexing, complex searches
 - Workflow capabilities
 - Productions
 - Object-level security
 - Advanced analytics (e.g. clustering, predictive coding)

Review platform – an example (Relativity)

311.000.000.0 Salt vs Pepper (demo and test) - Relativity - Windows Internet Explorer provided by Deloitte.

https://relativity.deloitte.nl/Relativity/List.aspx?AppID=1015632&ArtifactID=2685299&ArtifactTypeID=10

311.000.000.0 Salt vs Pepper (demo and test) - Rel...

311.000.000.0 Salt vs Pepper (demo and test)

Hi, Johan

Documents Admin Options Batches Reports Custodian Tracking User Status Analytics OCR Pivot Profiles

Folders

- 311.000.000.0 Salt vs Pepper (demo)
 - Admin Staging
 - Custodians
 - File Servers
 - Foreign Documents
 - Transcripts

Documents - Threading In This Folder & Subfolders

0 Selected Item(s) [Reset Column Sizes](#) | [Show Filters](#) | Clear All | Items 1 - 200 (of 1.000)

	<input type="checkbox"/>	Control Number	Sent Date	Email Author	Email To	Subject	Responsiveness	Issues	Privilege
1	<input type="checkbox"/>	AS000001	4-5-2009 15:45	asieja@kcura.com	nrobertson@kcura.com	kCura Relativity			
2	<input type="checkbox"/>	AS000002	10-1-2008 4:52	asieja@kcura.com		Relativity Review Stats.xls			
3	<input type="checkbox"/>	AS000003	10-1-2008 4:52	asieja@kcura.com		client_presentation.ppt			
4	<input type="checkbox"/>	AS000004	10-1-2008 4:52	asieja@kcura.com		맥주제조과정.doc			
5	<input type="checkbox"/>	AS000005	10-1-2008 4:52	asieja@kcura.com		relativity_pilot_agenda.doc			
6	<input type="checkbox"/>	AS000006	10-1-2008 4:52	asieja@kcura.com		big_video.wav			
7	<input type="checkbox"/>	EN000001	13-12-2000 8:35	messenger@ecm.bloomberg		Bloomberg Power Lines Report			
8	<input type="checkbox"/>	EN000002	13-12-2000 18:41	Multex Investor Network	pallen@enron.com	December 14, 2000 - Bear Stearns' predictions for telecom in Latin America			
9	<input type="checkbox"/>	EN000003	9-10-2000 7:16	phillip.allen@enron.com	keith.holst@enron.com	Consolidated positions: Issues & To Do list			
10	<input type="checkbox"/>	EN000004	9-10-2000 7:00	phillip.allen@enron.com	keith.holst@enron.com	Consolidated positions: Issues & To Do list			
11	<input type="checkbox"/>	EN000005	5-10-2000 6:26	phillip.allen@enron.com	david.delainey@enron.com				
12	<input type="checkbox"/>	EN000006	5-10-2000 5:55	phillip.allen@enron.com	paula.harris@enron.com	Re: 2001 Margin Plan			
13	<input type="checkbox"/>	EN000007	4-10-2000 9:23	phillip.allen@enron.com	ina.rangel@enron.com	Var, Reporting and Resources Meeting			
14	<input type="checkbox"/>	EN000008	3-10-2000 9:30	phillip.allen@enron.com	pallen70@hotmail.com	Westgate			
15	<input type="checkbox"/>	EN000009	3-10-2000 9:15	phillip.allen@enron.com	ina.rangel@enron.com	Meeting re: Storage Strategies in the West			
16	<input type="checkbox"/>	EN000010	3-10-2000 9:13	phillip.allen@enron.com	bs_stone@yahoo.com				

Checked Edit Go

Viewing the First 1.000 of 74.137 items in sets of 200 per page

Trusted sites | Protected Mode: Off 100%

Review platform – an example (Relativity)

311.000.000.0 Salt vs Pepper (demo and test) - Relativity - Windows Internet Explorer provided by Deloitte.

https://relativity.deloitte.nl/Relativity/Case/Document/Review.aspx?AppID=1015632&ArtifactID=2893762&profilerMode=View&ArtifactTypeID=10

Convert Select

AS000001 Document 1 of 1000

Viewer Native Image Extracted Text Acme Production Delete Images Edit Coding

100% Draft

[To: Nick Robertson (nrobertson@kcura.com)]
[From: Andrew H. Sieja]
[Sent: Mon 04-05-2009 15:45:48]
[Subject: kCura Relativity]
[big_video.way](#)
[Relativity Review Stats.xls](#)
[client_presentation.ppt](#)
[korean.doc](#)
[relativity_pilot_agenda.doc](#)

Hi Nick-

I hope you are enjoying this demonstration of Relativity. During the demonstration we are going to cover a lot of material so please stay awake! If you have any questions, don't hesitate to stop me along the way, but if you sit tight, I'll probably cover it at some point during the demonstration.

We are about to go through some attachments that showcase how Relativity displays native documents inside our proprietary viewer. It will include an Excel spreadsheet, a Word document, a PowerPoint presentation, a Word document in Korean, a document with tracked changes, and finally a 10 MB video file.

I just want to note that we are accessing Relativity deployed at our datacenter clear across the internet. You will notice that the documents still come up pretty quickly.

Enjoy the rest of the demo...

Document Details

Control Number: AS000001

Custodians: Sieja, Andrew

Coding

Responsiveness:

Privilege:

Confidentiality:

Issues:

Reviewer Comments

This is a Hot Doc.

Redaction Log Comments

this document redacted because of x, y, z

Attorney 865,38 kb/s Edit

Search terms

- **Recall and precision – trade-off**
 - Recall – number of relevant documents retrieved by a search / total number of existing relevant documents
 - Precision – number of relevant documents retrieved by a search / total number of documents retrieved by that search
- **Early case assessment (ECA)**
 - Test the search terms for number of hits and relevancy of hits (on a sample), iteratively refine the search terms
- **Keep updating the search terms** based on the outcomes of the review and new facts identified during the engagement

Search terms

- Some terms are likely to generate **excessive number of search hits**
 - General terms such as “payment” or “invoice”, employee names, words that appear in email signatures, homonyms (words that have the same spelling but different meaning)...
 - Depending on the client needs, terms with large number of hits can be acceptable
- Account for **stemming** – either replace the variable part of word with a wildcard or list all the forms
- Include **misspellings**
- Search for terms with and without **accents** (emails)
- Include search terms in multiple **languages** when appropriate
- Use thesaurus – add **synonyms**
- Use **Boolean** and **other operators** (proximity searching, exact phrases)

Technology assisted review

- **Near de-duplication**
 - Groups content similar documents together
- **Email threading**
 - Entire thread visible, enables focusing only on unique or most inclusive items
- **Unsupervised machine learning – Clustering**
 - Creates clusters based upon patterns of words with relative weight
- **Supervised machine learning – Predictive coding**
 - The machine will “learn” how to score documents for relevance based on past coding made by humans
 - The machine then sends iterative statistically generated samples to the reviewer, improving the accuracy of relevance
 - Once the desired level of accuracy is achieved, the machine training is complete. The remaining population can be scored via the machine

Example projects

Project 1 – France, a pharmaceutical company

- Allegations: FCPA (Foreign Corrupt Practices Act) violations
- Joint effort of Deloitte Netherlands, Czech Republic and US, client based in US
- Extensive data collection at short notice
 - Five countries, 30 custodians
 - More than 8 TB raw data collected
- Setup of standalone processing and review platform
 - Custom NUIX and Relativity setup on the client's premises – “mobile” solution ready within several days

Project 1 – France, a pharmaceutical company

- Processing large volumes of data
 - Pre-processing reduced the amount of data for indexing by a factor of 10
 - Further reductions in data uploaded into the review platform by applying global de-duplication (as per the client's request) and individual date ranges per custodian

Project 1 – France, a pharmaceutical company

- Challenges of review
 - Client adding more reviewers at short notice to meet aggressive deadlines
 - Constant requests for review customizations brought surprising level of complexity
 - No ECA performed due to the time pressure – search terms could have been refined earlier
 - Used industry standard machine translation software, still some language pairs worked poorly (ie Bulgarian vs. English) – client added more reviewers
 - “Legal can’t see how complex can data processing be” – we are here to meet the client’s expectations

Project 2 – Germany, an industrial conglomerate

- Allegations: FCPA violations
- Largest forensic investigation ever completed by that time
- Processed the laptops, mailboxes, PDAs, home directories and external media of 5,000 employees on a global scale
- Several other work streams running in parallel
- We helped the client to reduce the initial assumption of \$6 billion fine to \$1 billion finally paid

Key takeaways

Key takeaways

- **Forensically sound data collection, processing and review support require proven methodology, specialist knowledge and tools**
- Pre-process the data by applying a well tested procedure – get more evidence later
- Perform Early Case Assessment whenever possible, refine search terms iteratively
- Continuous and thorough quality control is essential

Thank you!

Jan Balatka

Senior manager

Forensic & Dispute services – Analytic & Forensic Technology

Deloitte Advisory s.r.o., Karolinska 654/2, 18600 Praha 8, Czech Republic

mobile: +420 731 450 902, email: jbalatka@deloittece.com

Disclaimer

This presentation contains general information only and is based on the experiences and research of Deloitte and its practitioners. Deloitte is not, by means of this presentation, rendering business, financial, legal, investment, or other professional advice or services. This presentation is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte, its affiliates, and related entities shall not be responsible for any loss sustained by any person who relies on this presentation.

Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/cz/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms.

© 2012 Deloitte Czech Republic