Deloitte.

Computer Forensics and Electronic Discovery

Dr. Bruce V. Hartley, CISSP, Deloitte Financial Advisory Services LLP

August 9, 2006

Agenda

- Introduction
- What is Electronic Discovery?
- Why Electronic Discovery?
- An Overview of the Electronic Discovery Process
- Where Does Computer Forensics Fit In?
- Some Case Studies…Smoking Guns
- New Technologies and Tools for Analysis of Data
- Conclusions and Recommendations
- Questions and Answers

Introduction

- The American Bar Association Digital Evidence Project and National Law Journal Report:
  - Over 30 Billion emails are sent daily
  - Over 90% of ALL information is now electronic
  - 70% of electronic information has never been printed
  - One in five US companies' employees email has been subpoenaed
  - Typical Fortune 500 company has 125 on-going cases with at least 75% requiring electronic discovery
  - Estimated that US companies spent $1.2B in outside e-discovery services in 2005 and $1.9B in 2006
  - 62% of companies surveyed doubt they can show their electronic records are accurate and reliable
  - Estimated that US companies will spend $4.6B internally just to analyze email traffic!

What is “Electronic Discovery”

The identification, location, preservation, retrieval, review and production of electronic documents and information in regulatory, civil and criminal environments.

. . . and the policies and procedures for information management, information technology, records retention, and corporate compliance to support the process and minimize its risks and costs.

Why Electronic Discovery?

- Required by Law – Federal Rule 26(a)
- Case Law
- Significant Sanctions (Fines) for Not Producing!
- This is serious stuff!
  - Judges no longer content to accept excuses
  - More and more attorneys and judges are very computer literate
  - Expectation is that companies manage and maintain all their data, paper and electronic
  - No excuses for non-production if demanded…

The Law

- Federal Rule of Civil Procedure 26(a)(1)(B) requires a party to provide to other parties "a copy of, or a description by category and location of, all documents, data compilations, and tangible things that are in the possession, custody, or control of the party and the disclosing party may use to support its claims or defenses . . .

Case Law

- Linnen v. A.H. Robins Co., 1999 WL 462015 (Mass. Super. June 16, 1999). “A **discovery** request aimed at the production of records retained in some **electronic** form is no different in principle, from a request for documents contained in any office file cabinet.”

Case Law

- Anti-Monopoly, Inc. v. Hasbro, Inc., 94 CIV 2120, 1996 U.S. Dist. LEXIS 563. “The law is clear that data in computerized form is discoverable even if paper ‘hard copies’ of the information have been produced, and that the producing party can be required to design a computer program to extract the data from its computerized business records, subject to the Court's discretion as to the allocation of the costs of designing such a computer program.”

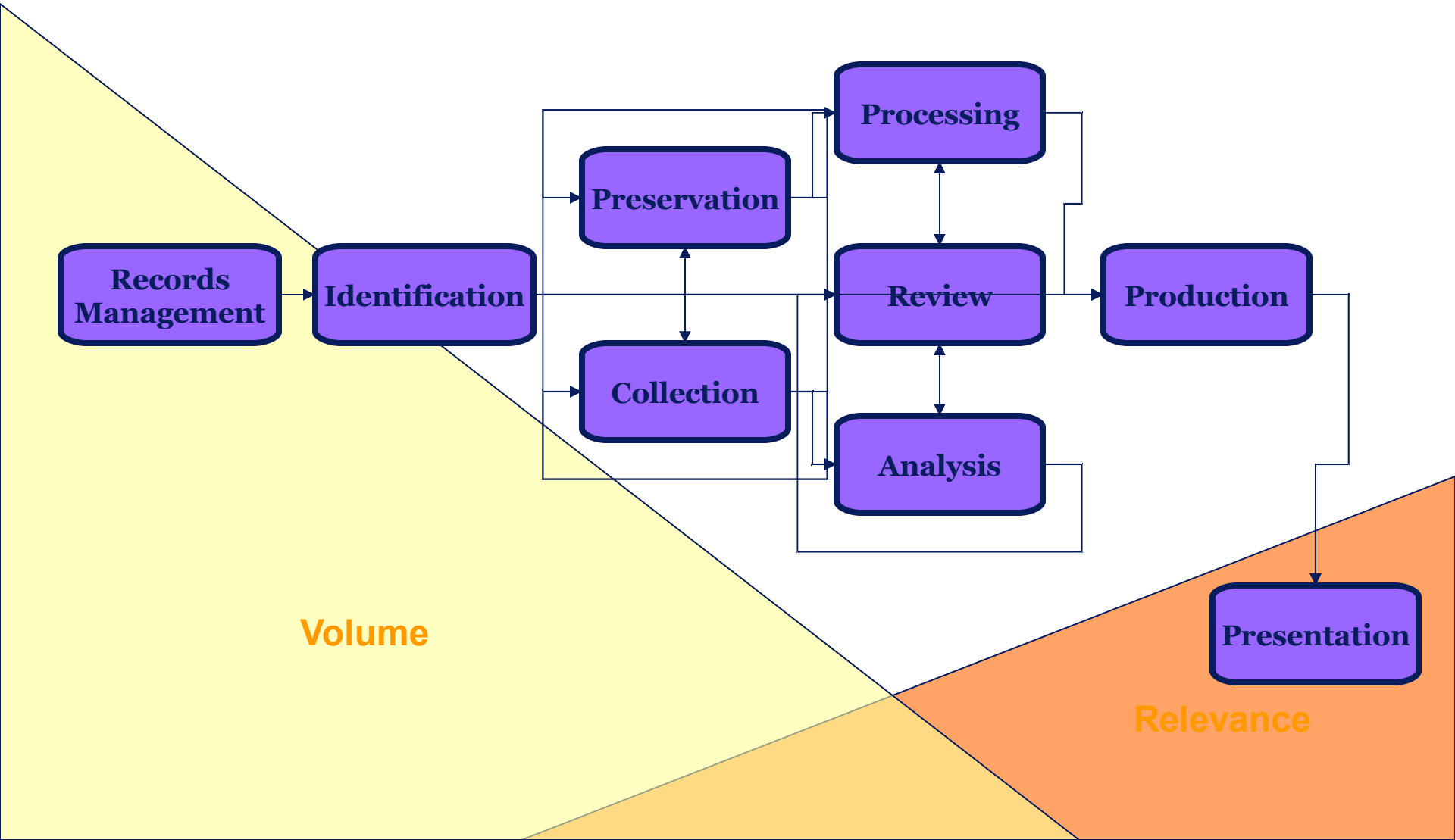
Horror Stories That Could Be You…

- Recently, Morgan Stanley fined $15,000,000 by SEC for improper records keeping
  - Fined $1.85 million in 2002 for similar reasons
- Coleman Parent Holdings, Inc versus Morgan Stanley (2005)
  - Improper destruction of emails led to partial summary judgment and $1.45 billion total award to claimant
- Zubulake versus UBS Warburg, LLC (2005)
  - Jury awarded $29 million to Laura Zubulake after receiving adverse inference instructions over destruction of hypothetical records, including emails, on back-up tapes
- Prudential Insurance Fined $1,000,000 for destroying records during a class action suit.

Electronic Discovery Process - Overview

- Identification
- Preservation
- Collection and Culling
- E-File Processing
- Search, Cull and Review
- Production

The Electronic Discovery Reference Model



Identification

- Conduct inventory of all data storage locations
- Document process by which all data is:
- created, secured, stored, retained, deleted, transferred, modified and archived
- Develop strategy for identifying, locating, and retrieving discoverable data

Locations of Data

- File servers
- E-mail servers **[!]**
- Instant messaging
- PC hard drives
- Backup tapes
- PDAs
- CDs
- Floppy disks
- Flash drives
- iPods
- Telephones
- Home computers
- Countless other possibilities

Preservation

- Preservation Letter - Formal notice that a claim is anticipated or has been filed with the court
- Must take immediate reasonable steps to preserve evidence
- Defines **electronic** data for the purposes of the request
- Must refrain from taking steps that would lead to spoliation

Collection and Culling

- Once data sources are identified, need to collect.
- Data collection can be done in many ways, depending upon the case matter, discovery order, etc.
- Two main choices:
  - Forensic data acquisition
  - Logical data acquisition
- After collection, perform initial culling
  - Remove known system files (MD5 Hash)
  - Remove other “non-printables”
  - Possibly de-duplicate (within custodian or across)
  - May date range filter or perform other pre-processing searches

Computer Forensics

- The process of obtaining digital evidence from storage media for use in legal proceedings.
  - Preserving the original media or a “duplicate” (bit-for-bit) copy for use in litigation.
  - Searching and documenting relevant evidence, including exculpatory evidence.
  - Safeguarding of the privacy interests of the victim and suspect.
- All data is handled as evidence – Maintaining the Chain-of-Custody is critical

How Much Data?

- E-Mail
  - 1 megabyte = approximately 71 pages
  - 70% of documents produced
- Total average data per person
  - 1 – 2 gigabytes (post culling)
  - Industry standard is between 55,000 and 70,000 pages per gigabyte
  - 1 gigabyte when printed translates to approx. 33 boxes

Forensic Acquisitions

- Document that a particular file is on a system or piece of media.
- Renamed files can be recovered by their “fingerprint”
- Document that a file has been deleted and depending on the operating system, possibly tell you who deleted it and when.
- Recover files from a reformatted HDD.
- Recover deleted files.
- Find other potentially relevant data in such places as temp files, slack and swap space.
- Gain access to password protected and encrypted files.

What May Be Located

- E-mail
- Evidence no longer available in paper form
- Embedded information
- File Signatures
  - Secrets.txt renamed to personal.bmp
- Financial databases and systems
- Temp files (.tmp)
- Backup files (.bk)
- Registry Information
- Deleted files
- And more….

More hidden information

- Recover “meta-data” from certain documents
  - Microsoft documents (Word, Excel, PowerPoint, Access) contain:
    - Original author
    - User who last saved the document
    - Revision number
    - Date last printed
- Recover previously deleted text within a file (previously saved versions) with revisions

E-File Processing

- Traditional e-discovery process includes:
  - Explode all archives (.pst, .zip, .tar, etc.)
  - Maintain all parent-child relationships
  - Extract text
  - Extract metadata
  - Create image (Tiff of .pdf)
- Goal is to create a fully searchable data collection and support attorney review process
  - Term and Boolean searches
- Newer approaches allow first pass processing in native file format
  - Extract text, metadata and create links to native files for review
  - Less costly approach – Better for some file formats (.xls)
  - Post review, image only the responsive files for production to the courts

Search, Cull and Review

- Once data is e-file processed, can search for responsive documents and documents that might be privileged
- Current state-of-the-practice is still term and Boolean searches
- Some newer concept-based tools being applied to larger data collections for initial searches
- Most review tools provide same basic functionality:
  - Search
  - Organize (create folders, tag, create notes, redact, etc)
- Goal is organize data into manageable units for review and analysis
- Prepare for production to courts and/or other side

Production

- In most cases we are talking about delivering data in a “useable” format to the courts and/or opposing side
  - Delivery may be in the form of a “load file” that contains extracted text, metadata, and images
  - In many cases, images are Bates numbered and endorsed and/or branded (possibly with a confidentiality stamp)
  - Rarely are productions done in native format today due to issues with redactions, Bates numbering, and metadata
  - Paper productions becoming very rare, but can still occur
    - Printed images

Why is Process Important?

- May find data that makes or breaks a case
- Many times we find data that was never intended for external or public consumption
- Once data is processed and searchable, can analyze for unique patterns and timelines that may or may not support a case
- Many cases have several terabytes of data to wade through – impossible to do manually – becoming the norm
- Industry has seen its first petabyte case…

Case Studies…Smoking Guns!

"*Screw Sun . . . . Let's move on and steal the Java language*."

9/17/97 e-mail from Microsoft manager Prashant Sridharan, quoted in Justice Department's Memorandum of Law in Support of Preliminary Injunction in *U.S. v. Microsoft Corp.,* 84 F. Supp. 2d 9 (D.D.C. 1999)

Case Studies…Smoking Guns!

"*Our recent handling of the employee Handbook receipt and agreement process left a lot to be desired. Virtually all aspects of the communication and distribution process that could have gone wrong did – moreover, the process was too complicated and legalistic …*"

8/17/98 e-mail to all CIGNA employees, quoted in *Leodori v. CIGNA Corp.,* 175 N.J. 293, 298-299 (2003)

Case Studies…Smoking Guns!

"[Excite@Home] is such a piece of crap!"

6/3/00 internal Merrill Lynch analyst e-mail coinciding with the Firm's ratings of "accumulate" and "buy," quoted in affidavit submitted to court by N.Y. Attorney General in *Spitzer v. Merrill Lynch & Co.,* Index No. 02/401522 (N.Y. Co. 2002)

Case Studies…Smoking Guns!

"[After our break-up, my life became an] absolute complete disaster, and it is a struggle every day just to get through it."

"[U]nder these circumstance[s] and how you have handled the end of our relationship, I don't see how we can work together. You have done all of this."

3/23/00 e-mails from a supervisor to his subordinate, who had recently ended their affair, as quoted in *Kaminski v. Freight-A-Ranger*, 2002 WL 31174461 (N.D. Ill. 2002)

New Technologies and Advances

- Concept-Based Searching
- Email Analytics
- Near De-Duplication
- Automated “First” Review
- Intuitive Graphical User Interfaces

Concept-Based Searching

- Based on several different technologies, such as latent semantic indexing
- Goal is to relate similar documents to one another and “cluster” for the reviewer based on the concept discussed in the documents
- Great for wading through large data collections, can save significant time and resources
- Excellent tool for complex investigations, such as fraud cases, terrorism cases, etc.

Email Analytics

- Since a significant portion of most data collections is email, need a better way to look at the big picture
- Social network analysis is extremely useful
  - Who is talking to who, both directly and indirectly
- Timeline analyses also very interesting
  - When was there a lot of message traffic, just before or right after a suspected event?
- Overlaying timeline analysis and social network analysis and the picture begins to get much clearer!

Near De-Duplication

- With data collections starting off in the multi-terabyte and beyond ranges, we are still looking at better ways to cull down the data collection and locate only those relevant documents
- Exact matches, based on MD5 hash values help, but in many cases we have very similar documents that can also be removed
- Various technologies and approaches from statistical analysis of like words, to hashing more fields, etc.

Automated Review Tools

- In addition to “Smart” GUIs and Concept-based search engines we are seeing a growing interest in addressing the document review problem
- Application of Rule-Based technology and Artificial Intelligence technology to automate review process
- Still requires human review – un-proven and not yet accepted by the courts
- Examples:
  - H5 Technologies

Intuitive GUIs

- Significant research is going on in the user interface arena
- We are seeing more analytical types of interfaces that leverage graphics, hotlinks, etc.
- Examples:
  - Attenex Patterns
  - Stratify
  - Others

Current Trends

- Both costs and risks associated with preservation, review and production are high and increasing
- Requesting parties routinely seek sanctions for alleged failures to preserve and produce electronic information
- Criminalization and large fines for lapses in preservation and production obligations
- Many courts and regulators are losing patience with both inside and outside counsel who claim to be uniformed about the policies and procedures of their clients

Conclusions

- Preserve large, produce small
- Preserve and secure electronic data using methods that have withstood judicial scrutiny
- Obtain all data potentially relevant to a matter
- Minimize cost and business disruption
- Screen for relevance
- Prepare for production

Questions & Answers

Contact Information

- Dr. Bruce V. Hartley, CISSP
  Firm Director
  Deloitte Financial Advisory Services LLP
  brhartley@deloitte.com
  (202) 378–5175

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu, a Swiss Verein, its member firms and their respective subsidiaries and affiliates. Deloitte Touche Tohmatsu is an organization of member firms around the world devoted to excellence in providing professional services and advice, focused on client service through a global strategy executed locally in nearly 150 countries. With access to the deep intellectual capital of 120,000 people worldwide, Deloitte delivers services in four professional areas, audit, tax, consulting and financial advisory services, and serves more than one-half of the world's largest companies, as well as large national enterprises, public institutions, locally important clients, and successful, fast-growing global growth companies. Services are not provided by the Deloitte Touche Tohmatsu Verein and, for regulatory and other reasons, certain member firms do not provide services in all four professional areas.

As a Swiss Verein (association), neither Deloitte Touche Tohmatsu nor any of its member firms has any liability for each other's acts or omissions. Each of the member firms is a separate and independent legal entity operating under the names "Deloitte", "Deloitte & Touche", "Deloitte Touche Tohmatsu" or other related names.

In the US, Deloitte & Touche USA LLP is the US member firm of Deloitte Touche Tohmatsu and services are provided by the subsidiaries of Deloitte & Touche USA LLP (Deloitte & Touche LLP, Deloitte Consulting LLP, Deloitte Financial Advisory Services LLP, Deloitte Tax LLP and their subsidiaries), and not by Deloitte & Touche USA LLP. The subsidiaries of the US member firm are among the nation's leading professional services firms, providing audit, tax, consulting and financial advisory services through nearly 30,000 people in more than 80 cities. Known as employers of choice for innovative human resources programs, they are dedicated to helping their clients and their people excel. For more information, please visit the US member firm's web site at www.deloitte.com/us.


Deloitte.