Secrets of Electronic Discovery

David F. Axelrod, Director
Ernie Liu, Manager
Forensic & Dispute Services
Deloitte Financial Advisory Services LLP

CONTENTS

• Litigation Technology Infrastructure
• Phases of Electronic Discovery
  – Define scope
  – Preserve data
  – Collect and cull
  – Process
  – Review and analyze
  – Produce
• Time and Cost Factors in Electronic Discovery
• Computer Forensics in Electronic Discovery
• Appendix – Glossary of Technical Terms
Litigation Technology Infrastructure
Litigation Technology Infrastructure

• Infrastructure (in•fra•struc•ture):
  – The *underlying foundation or basic framework*
  – Also: the *resources required for an activity* (as in personnel, buildings, or equipment)
  – Source: Merriam-Webster Online Dictionary

“Infrastructure is to information as a bottle is to wine: the technology is the packaging that allows the information to be distributed to the end consumers.”

Litigation Technology Infrastructure: Why it’s important

• **Evidence admissibility**

  *Forensics: “Relating to the use of science or technology in the investigation and establishment of facts or evidence in a court of law”* (1)

• **Data volume**

  “While 20 years ago PCs were a novelty and e-mail did not exist, today by some estimates more than 90 percent of all information is created in an electronic format.” (2)

• **Opportunity to use technology to leverage the electronic characteristics of the data**

(1) The Free Dictionary by Farlex

"I’d rather not challenge the expert based on his/her analysis. I’d rather go after the viability of the data used and throw the expert out on that basis."

--Anonymous (and various) attorneys
Litigation Technology Infrastructure: Data

Unstructured Data

ABC Bro...
Solutions

Report and Analysis

Organize Case-Related Information Using:

Case Management Applications

“Unstructured Data”

Structured Data

Review Using:

Computer Forensics
E-Discovery Applications

Review Using:

Traditional Account Analysis
Investigative Analysis
Anomaly Detection
Phases of Electronic Discovery
An Electronic Discovery Reference Model

- Records Management (Phase 0)
- Define Scope (Phase 1)
- Collect & Cull (Phase 3)
- Preservation (Phase 2)
- Analysis (Phase 5)
- Review (Phase 5)
- Processing (Phase 4)
- Production (Phase 6)

Volume

Relevance
Phases of Electronic Discovery

• Phase 1: Define scope of discovery

• Phase 2: Preserve data

• Phase 3: Collect and cull

• Phase 4: Process

• Phase 5: Review and analyze

• Phase 6: Produce
Phase I – Define Scope of Discovery

• **Goal**
  – Strategy for identifying, locating, and retrieving relevant data

• **Benefits**
  – Limit document collection (don’t collect what you don’t have to!)
  – Project the costs of downstream processes
  – Facilitate discussion of electronic discovery issues at Rule 26(f) conference

• **Challenges**
  – So much data, and so many sources; difficult to compile an inventory
  – Documenting the process by which all data is created, secured, stored, retained, deleted, transferred, modified and archived
Phase I - Define Scope of Discovery

Mobile and Other Devices
- Cell Phones
- Palm Pilots
- Blackberries
- USB Drives (a/k/a “thumb” or “flash” drives)
- Onboard Systems
- Digital Cameras
- Voice mail

External Sources
- Former Employees
- Clients
- Experts
- Consultants
- Licensees
- Suppliers
- Industry Data

Internet

Internet Traffic

Local Area Network

Main Frame

Mail Server

File Server

Offsite Storage

Tape Backup

Hard Disk

CD-ROMs, DVDs

Disks (Floppy, Zip, etc.)

Finc

Actg

Mktg Laptop

Operations Desktop

Hard Disk

Laptop

Internet Traffic

Traffic

Finc

Actg

Mktg Laptop

Operations Desktop

Laptop

CD-ROMs, DVDs

Disks (Floppy, Zip, etc.)

Main Frame

Mail Server

File Server

Offsite Storage

Tape Backup

Hard Disk

Mobile and Other Devices

External Sources

Finc

Actg

Mktg Laptop

Operations Desktop

Laptop

CD-ROMs, DVDs

Disks (Floppy, Zip, etc.)

Main Frame

Mail Server

File Server

Offsite Storage

Tape Backup

Hard Disk

Internet Traffic
Phase I – Define Scope of Discovery

• **Recommendations**
  
  – Get the right people in the room
    
    • Counsel
    • Client information technology
    • Business unit personnel (custodians)
  
  – Impose reasonable limits
    
    • See discussion of time and cost factors, below
  
  – Build an inventory of identified data sources
    
    • A “control” document
    • Basis to estimate costs per unit
  
  – Early and open dialogue with the other side
  
  – If possible, proactively assist your client with this process (prior to litigation)
Phase II – Preserve Data - Litigation Holds

- **Distinguish preservation of records from collection, review and production**
  - Preservation often inexpensive compared to collection, review and production

- **Litigation Hold definition**
  - The process of notifying employees of obligations to preserve appropriate records and suspend normal records management and destruction, while continuing routine destruction of non-relevant active and archived data

- **Legal standard**
  - Triggered when litigation is reasonably anticipated
  - Requires preservation of all potentially relevant evidence
  - Relevance in a “discovery” rather than “admissibility” sense
Phase II – Preserve Data - Litigation Holds (cont’d)

• The Overall Challenge
  – Implementation as required by law, while minimizing disruption of normal business activities
  – This involves both human and technical challenges

• Human Challenges
  – Reliance on untrained, lay personnel (custodians and others)
  – Need for a shared language
    • Lawyers (in-house and outside)
    • Consultants
    • IT personnel
    • Custodians
  – Pinpointing the specific time when litigation becomes reasonably anticipated
  – Determining what records should be preserved
  – Determining how to preserve them
Phase II – Preserve Data - Litigation Holds (cont’d)

• **Technical Challenges**
  – Technical challenges arise from five characteristics of electronic records:
    
    • **Volume** - The rate at which electronic records are created has accelerated rapidly in recent years
    
    • **Duplicability** - Electronic records can be replicated rapidly, with or without human intervention
    
    • **Dynamic content** - Electronic records often have dynamic content that can change over time
    
    • **Metadata** - Metadata is descriptive information about documents, files and e-mail that assists users and facilitates the storage and retrieval of electronic records
    
    • **Dispersion** - Though freezing paper records may be difficult, they may be collected in discrete filing cabinets, while electronic records can reside in many places simultaneously
Phase II – Preserve Data - First Steps

• **Secure physical media**
  - Computers and servers
  - Peripheral media
  - Backup tapes
  - Many other possible data repositories (see next slide)

• **Discontinue automated data deletions and modifications of relevant records – *e.g.*,**
  - Tape rotations
  - Automated email archives and deletes
Phase II – Preserve Data

Examples of Data Repositories

Mobile and Other Devices
- Cell Phones
- Palm Pilots
- Blackberries
- USB Drives (a/k/a “thumb” or “flash” drives)
- Onboard Systems
- Digital Cameras
- Voice mail

External Sources
- Former Employees
- Clients
- Experts
- Consultants
- Licensees
- Suppliers
- Industry Data
Phase II – Preserve Data Recommendations

• Identify an overall “owner” of the process, and subordinate owners of specific aspects

• Issue the Litigation Hold notice early

• Require confirmation of receipt

• Require confirmation of actions taken

• Issue very specific guidance to personnel responsible for preserving paper and electronic data, wherever it resides

• Seek legal and technical guidance to validate preservation processes

• Plan for subsequent preservation notices
  – Periodic reminders may be required
  – Modifications in scope

• Consider “evergreen” preservation requirements

• Track status using inventory of data sources
Phase III – Collect and Cull Data

• Goal
  – Gather data potentially responsive to discovery requests

• Benefits
  – Places the responsive data in one (or a few) places
  – Helps begin elimination of obviously non-responsive and redundant documents ("initial cull")
    • System files and other non-printables
    • De-duplication
    • Other filters (date, etc.)
  – Chain of custody documentation validating procedures
  – Under certain circumstances, can economically combine the preservation and collection steps
Phase III – Collect and Cull Data

• **Challenges**
  - Over-collection
  - Preserving chain of custody and data integrity
    • *E.g.*, metadata
  - Data management
    • Data may be voluminous
  - Non-standard data
    • Structured vs. unstructured data
    • Wide variety of file formats (see next slide)
  - International standards for data transfer across borders
Phase III – File Types: The Kitchen Sink*

*Non-Exhaustive List
Phase III – A Note on Interviews

- **Problem**: Where do individuals keep their data?

- **Interview Benefits**
  - Understand where custodians keeps their data
  - Allows for follow up questions to ensure completeness of searches

- **Interview Costs**
  - Time and expense
  - Disruption of business

- **Alternative approaches**
  - Selective use of interviews - tiers of custodians or other categories
  - Questionnaires to custodians (suggest places to look for data)
  - Group meetings
  - Consider the experience of the organization in dealing with document requests
Phase III – Collect and Cull Data

• **Recommendations**
  
  – Consider the range of collection options
    • Forensic image (full or logical) is the “gold standard” for evidence handling
    • Companies sometimes deploy other collection mechanisms because of cost or concern about over-collection
    • Balance chain of custody requirements against cost to the organization
  
  – Seek legal and expert guidance to validate collection processes
  – Negotiate collection requirements with the other side
  – Consider alternative means to help custodians locate their potentially responsive data
  – Apply filters early, if possible, to pare down the data
Phase IV - Processing

• **Goal**
  - Create a data set that fully supports the attorney review process

• **Benefits**
  - Create a consistent file format
  - Create a fully searchable data set
  - Apply searches to the data
  - **File format**
    • Traditionally, create a static image (TIFF or PDF) of each file
    • Industry gradually moving to “native” file review and production (see Phase VI: Production and Hosting)
    • Generally still see a hybrid approach (TIFFs and natives, especially for production purposes)
Phase IV - Processing

• **Challenges**
  – Maintaining all parent-child relationships (*e.g.*, emails and attachments)
  – Exploding all archives (.*pst*, *.*zip*, etc.)
  – Cost
  – Often review unstructured data only
    • *E.g.*, emails, Word documents
  – If structured data is found, usually treated as an “exception”
    • Either it is ignored or at best a picture of the data is rendered (*e.g.*, Access databases, transactional systems)
  • **Result**: Incomplete analysis, missing pieces of the puzzle, and narrative may be taken out of context
    • See Deloitte Research Foundation White Paper: “Transaction Data in Discovery”
Phase IV – A Note on Searching

• **Apply filters to identify potentially responsive data**
  – Use keywords to limit the population and save money
  – Tremendous search technology available

• **Key issue: What to cull and when**
  – Applying terms early in the process saves downstream cost
  – Risk of “going back to the well” if litigation circumstances change

• **When filters are typically applied**
  – Phase 3: Collect and cull (dates, file types)
  – Phase 4: Processing (key terms and phrases, concept based)
  – Phase 5: Review (ad hoc searching by attorneys on text and metadata)
Phase IV – Processing

• **Recommendations**
  – Begin with the end in mind
    • Primarily a back office process, but needs to support the attorney review
    • Consider what exceptions are likely and how to treat those exceptions
  – Negotiate key terms with the other side
  – Track progress and cost from a time and unit cost perspective
  – Consider best means to make structured data available
Phase V - Review and Analyze

• **Goal**
  – Analyze and classify the data, including legal classifications (e.g., privilege), to understand facts and circumstances

• **Benefits**
  – Review efficiency
  – Leverage tools that exploit electronic data
    • Simple, complex, and *ad hoc* searching
    • Simple and complex analytical and operational reporting
    • Ability to review and analyze thoroughly
    • Document/page level management and control
  – Ability to administer “batch” operations
    • “Mass” privilege or other tagging
    • Organize and assign documents for attorney review
Phase V - Review and Analyze (continued)

• Challenges
  – Current state-of-the-practice is still keyword, Boolean, and fuzzy logic searches
  – Tremendous variety of review tools
  – Most review tools provide same basic functionality: search and organization (create folders, tag, notes, redact, etc.)
  – Practical challenges integrating unstructured data review with structured data review

• Recommendations
  – Leverage your litigation support group when determining the appropriate review platform and processes
  – Recognize that different review platforms do different things well
  – Don’t forget about the value of data contained in transactional systems
Phase VI – Production

• **Goal**
  - 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

• **Benefits**
  - Traditionally, a memorialized/static production set
    - Usually images, but may still be paper
    - Bates numbers, endorsed, and/or branded (e.g., confidentiality)
  - In principal, data exchange between parties is fluid
  - Production reporting and control
    - Multiple productions to single or multiple parties
    - “Document history” features in review platforms
Phase VI – Production

• **Challenges**
  
  – Native format productions (see next slide)
  – Focus still on unstructured data (email, user files)
  – Structured data is still treated as an exception
  – Review platform alternatives
    • Reviewers know and are comfortable with industry standards (e.g., Concordance, Summation)
    • Complex technology selection process; all perform different functions well
    • Not all applications manage complex productions well
Phase VI – A Note on Native File Production

• Industry movement to native review and production
  – Can provide for a more “complete” review
    • Better context — “see what the custodian sees/saw”
    • Reveals “hidden” or “embedded” data in files that is only accessible (or discernable) through the native application

• Current challenges
  – Document control — redactions, bates numbering, metadata, etc.
  – Data volatility post production
  – Result: most productions are still provided in TIFF or PDF format
Phase VI – Production

• **Recommendations**
  
  – Negotiate and clearly delineate production requirements with the other side
  
  – Conduct early discussions with vendor/s regarding treatment for native or non-standard production requirements
  
  – Leverage advanced tools to track productions in complex litigation environments
  
  – Consider production of transactional data (parallel process)
An Electronic Discovery Reference Model (revisited)
Time & Cost Factors in Electronic Discovery
A 40GB Hard Drive Can Hold...

- The text in a stack of documents about 2000 feet high.
- 10,000 4-minute MP3s encoded at 'CD quality' (128 bps), which equals over 27 uninterrupted days of listening.
- The equivalent of about 10 DVD movies stored in MPEG-2 format.
- 3 hours of digital video (for camcorder editing).
This 60GB Hard Drive has 109,362 files
Define Scope of Discovery – Consider Potential Data Repositories

- Internet
- Internet Traffic
- Local Area Network
- Main Frame
- Mail Server
- File Server
- Hard Disk
- Offsite Storage
- Tape Backup
- Mobile and Other Devices
  - Cell Phones
  - Palm Pilots
  - Blackberries
  - USB Drives (a/k/a “thumb” or “flash” drives)
  - Onboard Systems
  - Digital Cameras
  - Voice mail
- External Sources
  - Former Employees
  - Clients
  - Experts
  - Consultants
  - Licensees
  - Suppliers
  - Industry Data

- File Server
- Operations Desktop
- Finc
- Actg
- Mktg Laptop
- CD-ROMs, DVDs
- Onboard Systems
- Digital Cameras
- Voice mail
- Former Employees
- Clients
- Experts
- Consultants
- Licensees
- Suppliers
- Industry Data
Typical Scope “Limiters”

- **Custodians**
  - Key
  - Secondary

- **Media purpose**
  - Work use
  - Personal use
  - Disaster recovery

- **Media Type**
  - Hard drives
  - Company servers
  - Backup tapes

- **Data Type**
  - Emails and user files
  - System files
  - Financial / ops databases
  - Other (web, proprietary)

- **Data Ownership**
  - Individual
  - “Shared” servers

- **Time Frames**
  - Period in time
  - Ongoing (“evergreen”)
Tapes: Some Questions to consider asking at the beginning

- How many tapes?
- Are catalogs available for the tapes?
- How often are the backups ran?
- Are the backups Full, Differential or Incremental?
- Is disaster recovery performed on a regular basis?
- What type of Backup Software used?
- What Platform?
- How many servers?
- How many custodians?
- Where are the custodians located?
- Can the network admin be contacted directly?
- How many tapes in backup sets?
Things You Will Need
Equipment

Tape drives can run in the high $1000’s
Equipment
Tapes, Tapes and More Tapes
Potential Issues in dealing with tapes

- Numerous tape sizes – 40GB – 800GB and growing
- Numerous tape formats – such as AIT, DAT, DLT, LTO2 and LTO3
- Hardware issues – need to have the tape drive to match the tape
- Software issues – need to use the same software as used to create
- IT Personnel – not always good about testing the tape to verify it works
- IT Personnel – not always good about keeping logs of tapes
- Need to duplicate the tape – 200GB takes approximately 5 hours
- Need to catalog the tape to determine what is on the tape, take approximately 2 hours
- Time involved to restore tapes – 200GB takes approximately 5 hours

Times are approximates and can vary depending on hardware
Identifying Tape Software

Veritas Software

Date & Time
12-26-03 12:25

Veritas Backup Exec
Version 8.60
Tapes - Sampling

• If you have 1000 backup tapes a set number of tapes can be sampled in order to identify or locate the data being requested. Instead of restoring 1000 tapes, which is a very time consuming process

• This can be accomplished by cataloging the tapes, which will give a list of the items on the tape
Computer Forensics in Electronic Discovery
The File System

• Commonly referred to as FAT or NTFS
  
  Is an index that points to the location on the hard drive where the file resides

• Deleting a file
  
  The location in the index is marked as being available thus the operating system will show you that the file has been deleted. However, the area on the hard drive in which the data resides will remain until it is overwritten

If you take a book and cross through Chapter 3 in the index, is the Chapter 3 data still there?
File Deletion

I am going to delete this file and it will be gone forever

Right?
File Deletion
File Deletion
File Deletion
File Deletion

I am going to delete this file and it will be gone forever.
Hard Drive Imaging – Forensic vs. Ghost

• **Forensic Imaging hard drive**

  A 'forensically-sound' duplicate of a drive is created by a method which does not alter any data on the drive being duplicated. A forensically-sound duplicate must contain a copy of every bit, byte and sector of the source drive, including unallocated 'empty' space and slack space.

• **Forensic Logical file imaging:**

  Duplicate of folder/folders created by a method which does not alter the data in the folder/folders being duplicated. It will contain a copy of all files logically imaged and will allow for the recovery of deleted files.

• **Ghost:**

  Is a logical copy of the active files, which can change the time stamps of the files and doesn’t allow recovery of deleted files.
Ghost Image
Forensic Logical Image
Forensic Logical Image

I am going to delete this file and it will be gone forever.

Logical File creation

Logical evidence file

I am going to delete this file and it will be gone forever.
# Data Gathering

<table>
<thead>
<tr>
<th>File Extension</th>
<th>Brief Description</th>
<th>Email</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>.xls</td>
<td>Microsoft Excel</td>
<td>.pst</td>
<td>Outlook Mail</td>
</tr>
<tr>
<td>.xla</td>
<td>Microsoft Excel</td>
<td>.ost</td>
<td>Outlook Mail</td>
</tr>
<tr>
<td>.mdb</td>
<td>Microsoft Access</td>
<td>.mbx</td>
<td>Outlook Express Mail or Eudora</td>
</tr>
<tr>
<td>.ppt</td>
<td>Microsoft PowerPoint</td>
<td>.dbx</td>
<td>Outlook Express Mail</td>
</tr>
<tr>
<td>.pps</td>
<td>Microsoft PowerPoint</td>
<td>.eml</td>
<td>Email File</td>
</tr>
<tr>
<td>.zip</td>
<td>WinZip File</td>
<td>.msg</td>
<td>Message File</td>
</tr>
<tr>
<td>.pdf</td>
<td>Adobe Acrobat</td>
<td>.nsf</td>
<td>Lotus Notes Mail</td>
</tr>
<tr>
<td>.tif/ .tiff</td>
<td>Tiff Image</td>
<td>.ntf</td>
<td>Lotus Notes File</td>
</tr>
<tr>
<td>.txt</td>
<td>Text File</td>
<td>.toc</td>
<td>Eudora Mail</td>
</tr>
<tr>
<td>.htm</td>
<td>Web File</td>
<td>.idx</td>
<td>Eudora Mail</td>
</tr>
<tr>
<td>.html</td>
<td>Web File</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.rtf</td>
<td>Rich Text Format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.mpp</td>
<td>Microsoft Project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.nsf</td>
<td>Lotus Notes File</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.csv</td>
<td>Comma Separated Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.123</td>
<td>Lotus 123 Spreadsheet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.wk4</td>
<td>Lotus 123 Spreadsheet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.wpd</td>
<td>Word Perfect File</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.dbf</td>
<td>Database File</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.log</td>
<td>Log Files</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I am totally changing this document.
Metadata

Document Name: I am totally changing this document.doc
Path: C:\Documents and Settings\galowe\Desktop\Screen shots\Tracked Changes
Document Format: Word Document

Built-in document properties:
- Built-in Properties Containing Metadata: 2
- Title: I am totally changing this document
- Company: Deloitte & Touche

Document Statistics:
- Document Statistics Containing Metadata: 6
- Creation Date: 4/17/2007 8:13:00 PM
- Last Save Time: 4/17/2007 8:13:00 PM
- Time Last Printed: [Blank]
- Last Saved By: [Blank]
- Revision Number: 1
- Total Edit Time (Minutes): 2 Minutes

Custom document properties:
- No Custom Document Properties

Last 10 authors:
- No Last 10 Author Data

Versions:
- No Versions

Track Changes:
- Tracked Changes: 2, Tracked Changes are On
- 1 Type: Delete Author: [1 Space]
  This is a document that I am going to make changes too in order to show metadata of the tracked changes
- 2 Type: Insert Author: [1 Space]
  I am totally changing this document
This is the wording in the document
This is the wording in the file slack, as you can see it is not a completed set of sentences.
I am a new resident in Metro Atlanta area, still looking for a suitable career opening. Your Company seems to have just

...
Unallocated Space

![Image of a computer screenshot showing a search process with status completed, search hits, and added search hits.

Copyright © 2007 Deloitte Development LLC. All rights reserved. 70
Appendix

Glossary of Technical Terms
Glossary of Technical Terms

Hardware

i. Hard drive
   • The primary storage unit on PCs
   • Magnetic media on which digital information can be written and erased magnetically

1 This glossary is based on definitions contained in “The Sedona Conference Glossary: E-Discovery & Digital Information Management” (May 2005 version) (available at www.thesedonaconference.org)
Glossary of Technical Terms

i. **Client/Server**

- **Client** – any computer system that requests a service of another computer system.
- **Server** – any central computer on a network that contains data or applications shared by multiple network users on individual (client) PCs.
 iii. Backup tapes

• Backup/archival – to create a copy of data as a precaution against loss or damage to original data. Most computer networks utilize automatic backup software

• Disaster recovery tapes – portable media used to store data for backup purposes.
Glossary of Technical Terms

iv. **Cache** – a high speed storage location, which can be used for temporary storage of frequently used data.

v. **RAM (Random Access Memory)** – hardware inside a computer that retains short term memory and stores information while a computer is in use, i.e., the “working memory” of the computer.

vi. **Clipboard** – a “holding area” that temporarily stores information copied or cut from a document.
vii. Slack space – Unused space on a cluster that exists when the logical file space is less than the physical file space. Can contain information deleted from the record, information from prior records stored in the same physical location, metadata fragments and other information useful for forensic analysis of computer systems.

1. Logical file space – the actual amount of space occupied by a file on a hard drive. Differs from the physical file space because when a file is created on a computer, a sufficient number of clusters (physical file space) are assigned to contain the file. If the logical file space isn’t large enough to completely fill the assigned clusters (the physical file space), some unused space will exist within the physical file space.
Glossary of Technical Terms

• Slack Space

The slack space is just above. You can see where the actual word document ends, before the slack.
vii. **Unallocated space** – the area of computer media, such as a hard drive, that doesn’t contain easily accessible data. Usually results from deletion of a file.

Unallocated space is seen as one large file. This is one screen of data that resides in such a file.

In this case the unallocated space is 40 GB
Glossary of Technical Terms

Software and Data

i. **Application** – Commonly used in place of “program” or “software. An application is a collection of one or more related software programs that enable an end user to enter, store, view, modify or extract information from files or databases. May include, among other things:

- Word processing programs
- Internet browsing tools
- spreadsheets
- Email clients
- PDA data (contacts, calendar, etc.)
- Janitor program – an application which runs at scheduled intervals to manage business information by deleting, transferring or archiving online data such as e-mail, which is at or past the end of its scheduled active life
Glossary of Technical Terms

ii. **Active data** – information residing on the direct access storage media (disc drives or servers) of computer systems, which is readily visible to the operating system and relevant application, and immediately accessible to users without restoration.

iii. **Archival data** – Archival data – information maintained for long-term recordkeeping purposes that is not immediately accessible to the user of a computer system.

   1. May be written to removable media or maintained on system hard drives.
   2. Some systems allow users to retrieve archival data directly, while other systems require the intervention of an IT professional.

iv. **Legacy data** – information that was created by software and/or hardware that has become obsolete or replaced (“legacy systems”). May be costly to restore or reconstruct.
Glossary of Technical Terms

v. **Backup data** – an exact copy of system data which serves as a source of recovery in the event of a system problem or disaster. Generally stored on portable media, separate from active data.

vi. **Distributed data** – information belonging to an organization that resides on portable media and non-local devices such as remote offices, home computers, laptops, PDAs, wireless communication devices (Blackberrys, Treos, etc.) and internet repositories (such as email hosted by internet service providers).

vii. **Fragmented data** – data broken up and randomly placed throughout available file storage on a computer’s hard drive or other storage medium, such as removable disks. This may occur because data that would normally be saved in contiguous clusters exceeds the contiguous free space.
Glossary of Technical Terms

viii. **Residual data** – data that is inactive on a computer system, such as data found on media free and slack space, or within deleted files

ix. **Metadata** – information about a particular data set or document, which describes how, when and by whom it was collected, created, accessed, modified and how it is formatted.

1. Can be altered intentionally or inadvertently.
2. Can be extracted when native files are converted to image.
3. Some metadata, such as file dates and sizes, are readily visible to users; other metadata can be hidden, embedded or unavailable to users who are not technically adept.
Metadata

Analyzing C:\Documents and Settings\galowe\Desktop\Screen shots\Tracked Changes\I am totally changing this document.doc

Document Name: I am totally changing this document.doc
Path: C:\Documents and Settings\galowe\Desktop\Screen shots\Tracked Changes
Document Format: Word Document

Built-in document properties:
Built-in Properties Containing Metadata: 2
Title: I am totally changing this document
Company: Deloitte & Touche

Document Statistics:
Document Statistics Containing Metadata: 6
Creation Date: 4/17/2007 8:11:00 PM
Last Save Time: 4/17/2007 8:13:00 PM
Time Last Printed: [Blank]
Last Saved By: [Blank]
Revision Number: 1
Total Edit Time (Minutes): 2 Minutes

Custom document properties:
No Custom Document Properties

Last 10 authors:
No Last 10 Author Data

Versions:
No Versions

Track Changes:
Tracked Changes: 2. Tracked Changes are On.
1 Type: Delete Author: [1 Space]
This is a document that I am going to make changes too in order to show metadata of the tracked changes
2 Type: Insert Author: [1 Space]
I am totally changing this document
**Temp file** – files stored on a computer for temporary use only. Often created by internet browsers, to store information about websites that a user has visited, and allow for a more rapid display of the web page when the user revisits the site. Also created by common office applications such as word processing and spreadsheet applications.
Glossary of Technical Terms

**xi. Cookie** – a message given to a web browser by a web server.

1. The browser stores the message in a text file. The message is sent back to the server each time the browser requests a page from the server.

2. The main purpose of cookies is to identify users and possibly prepare customized web pages for them.
Cookies

File and Folder Tasks
- Make a new folder
- Publish this folder to the Web
- Share this folder

Other Places
- galowe
- My Documents
- My Computer
- My Network Places

Details
Cookies
File Folder
Date Modified: Today, April 30, 2007, 3:28 PM

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Type</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>galowe@2o7[1].txt</td>
<td>3 KB</td>
<td>Text Document</td>
<td>2/8/2006 3:04 PM</td>
</tr>
<tr>
<td>galowe@2o7[2].txt</td>
<td>2 KB</td>
<td>Text Document</td>
<td>5/5/2005 1:01 PM</td>
</tr>
<tr>
<td>galowe@2o7[4].txt</td>
<td>2 KB</td>
<td>Text Document</td>
<td>4/26/2007 6:32 PM</td>
</tr>
<tr>
<td>galowe@3com[2].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>12/21/2005 11:32 AM</td>
</tr>
<tr>
<td>galowe@3m[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>9/30/2005 8:36 AM</td>
</tr>
<tr>
<td>galowe@4imprint[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>12/9/2005 4:40 PM</td>
</tr>
<tr>
<td>galowe@11alive[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>3/20/2007 8:01 PM</td>
</tr>
<tr>
<td>galowe@112.2o7[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/4/2007 8:22 AM</td>
</tr>
<tr>
<td>galowe@12021449[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/30/2007 3:13 PM</td>
</tr>
<tr>
<td>galowe@13041680[2].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>9/30/2005 9:28 AM</td>
</tr>
<tr>
<td>galowe@25041515[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/26/2005 7:33 AM</td>
</tr>
<tr>
<td>galowe@4183375[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>10/19/2005 3:47 PM</td>
</tr>
<tr>
<td>galowe@43836137[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>10/27/2005 8:04 AM</td>
</tr>
<tr>
<td>galowe@46369213[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>8/22/2005 12:00 PM</td>
</tr>
<tr>
<td>galowe@53320982[2].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/26/2005 10:44 AM</td>
</tr>
<tr>
<td>galowe@69937441[2].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>2/6/2006 10:52 PM</td>
</tr>
<tr>
<td>galowe@69974735[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/20/2005 10:57 AM</td>
</tr>
<tr>
<td>galowe@1067451516[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>1/26/2006 3:31 PM</td>
</tr>
<tr>
<td>galowe@1070216689[1].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>8/22/2005 11:56 AM</td>
</tr>
<tr>
<td>galowe@1070216689[2].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>10/19/2005 12:50 AM</td>
</tr>
<tr>
<td>galowe@1070216689[3].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/12/2007 7:24 PM</td>
</tr>
<tr>
<td>galowe@1070216689[4].txt</td>
<td>1 KB</td>
<td>Text Document</td>
<td>4/25/2007 12:23 PM</td>
</tr>
</tbody>
</table>
xii. **HTML (HyperText Markup Language)** – the document standard for the Internet.

xiii. **Native format** – the file structure defined by the original creating application. Because viewing or searching documents in native format may require the original application, documents are often converted to a vendor-neutral format as part of the record acquisition or archive process.

xiv. **Static formats (e.g., TIFF or PDF)** - designed to retain an image of the document as it would look when viewed in the original creating application but do not allow metadata to be viewed or document information to be manipulated.
<html>
<head>
<title>Delta Air Lines - Airline Tickets and Airfare to Worldwide Destinations</title>
<meta name="description" content="Airline tickets, flight schedules, and air travel deals to destinations worldwide at Delta.com."/>
<meta name="keywords" content="delta air lines, airline tickets, flight, cheap airfare, discount airfare, delta.com" />
<meta http-equiv="pragma" content="no-cache">
<meta name="verify-v1" content="JNg732/mwIkUyXHCgKHq/iwplz7T+WWgdsFttnc40=" />
<link rel="shortcut icon" href="https://a248.e.akamai.net/7/248/19260/v1/images.delta.com/delta/misc/favicon.ico"/>
<link rel="stylesheet" type="text/css" href="/components/css/home.css"/>
<link rel="stylesheet" type="text/css" href="/components/css/forms.css"/>
</head>
</html>
Glossary of Technical Terms

xv. **FAT (File Allocation Table)** – an internal data table on hard drives that keeps track of where the files are stored.

xvi. **Export** – data extracted or taken out of one environment or application, usually for import into another environment or application.

  – **Import** – data brought into an environment or application which has been exported from another
Glossary of Technical Terms

xvii.ASCII (American Standard Code for Information Interchange) – a non proprietary text format that consists of only text with no formatting, and can be read by most computer systems.

xviii.Byte, Kilobyte, Megabyte, Gigabyte, Terabyte, Petabyte and Exabyte
– a byte is the basic measurement of most computer data and consists of 8 bits. Storage is measured in larger increments of bytes

1. Bit (binary digit) – the smallest unit of computer data, consisting of either zero or one (0 or 1). There are 8 bits in a byte

2. E.g., a kilobyte is a unit of 1024 bytes.
Glossary of Technical Terms

Processes

i. Deletion – removal of data from active files and other storage structures on computers, rendering it inaccessible except through the use of special data and recovery tools.

1. When a file is deleted, it is not actually erased, but is simply no longer accessible through normal means. The space that it occupied becomes unallocated space on the drive that can be reused to store new information.

2. Until portions of the unallocated space are used for new data storage, the old data usually remains and can be retrieved using forensic techniques.

ii. Boot – to start up or reset a computer.
Glossary of Technical Terms

iii. De-duplication – the process of comparing electronic records based on their characteristics and removing or marking duplicate records within the data set.

1. The definition of duplicate records should be agreed upon (i.e., whether an exact copy from a different location is considered to be a duplicate)

2. Duplication can be selective, based on agreed criteria.
   - Metadata Comparison – a method of de-duplication that compares file metadata and ignores content
   - Case de-duplication – eliminates duplicates to retain only one copy of each document. For instance, if an identical e-mail message resides with three custodians, only the first custodian’s copy will be saved
iv. **Ghost image** – Is a logical copy of the active files, which can change the time stamps of the files and doesn’t allow recovery of deleted files.

v. **Forensic image** – an exact copy of an entire physical storage media (hard drive, CD-ROM, DVD-ROM, tape, etc.) including all active and residual data and unallocated space on the media. Compresses and encrypts to ensure authentication and protect chain of custody.

vi. **Hash** – a mathematical algorithm that represents a unique value for a given set of data, similar to a digital fingerprint

vii. **Hash coding** – to create a digital fingerprint that represents the binary content of a file. Assists in ensuring that data has not been modified
Glossary of Technical Terms

vi. **Restore** – to transfer data from a backup medium (such as tapes) to an on-line system, often for the purpose of recovery from a problem, failure or disaster.

1. Restoration of archival media is the transfer of data from an archival store to an on-line system for purposes of processing (such as query, analysis or extraction of data).

2. Archival restoration of systems may require not only data restoration but also replication of the original hardware and software operating environment.

vii. **Backup tape recycling** – the process whereby an organization’s backup tapes are overwritten with new data, usually on a fixed schedule determined jointly by records management, legal and IT sources.
Glossary of Technical Terms

viii. **Harvesting and Mining** – the process of retrieving or collecting electronic data from storage media or devices; an e-discovery vendor harvests electronic data from computer hard drives, file servers, backup tapes, etc.

ix. **Sampling** – usually refers to the process of testing a database for the existence or frequency of relevant information. Can be a useful technique in addressing items relating to litigation, including decisions about what repositories of data are appropriate to search in particular litigation, and determinations of the validity and effectiveness of searches or other data extraction procedures.

x. **OCR (Optimal Character Recognition)** – a process that converts printed matter on an image into a format that a computer can manipulate, and therefore, renders text searchable.
Contact Information

David F. Axelrod
Director
Forensic & Dispute Services

Deloitte Financial Advisory Services LLP
155 E. Broad St.
18th Floor
Columbus, Oh 43215
Tel +1 614-228-4525
Fax +1 614-233-6338
Mobile +1 614-307-5576
Email daxelrod@deloitte.com
www.deloitte.com
Member of Deloitte Touche Tohmatsu

John C. Moorman
Principal
Analytic & Forensic Technology
Forensic & Dispute Services

Deloitte Financial Advisory Services LLP
111 South Wacker Drive
Chicago, IL 60606-4301
Tel +1 312-486-4559
Fax +1 312-247-4559
Mobile +1
Email jmoorman@deloitte.com
www.deloitte.com
Member of Deloitte Touche Tohmatsu