



# *Enterprise eDiscovery*

## *Challenges, Methodologies, and Tools*

Christopher L. T. Brown, CISSP  
Technology Pathways, Founder & CTO  
[clbrown@techpathways.com](mailto:clbrown@techpathways.com)  
619-435-0906 / 888-894-5500

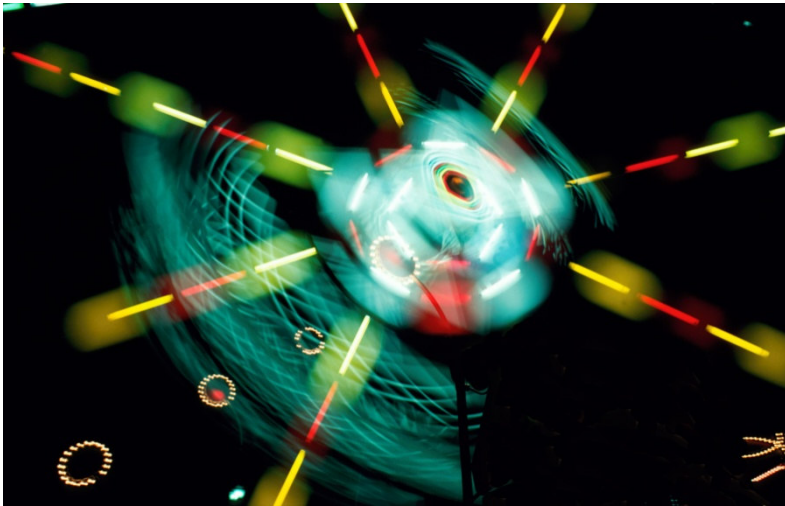
Copyright © 2007, Technology Pathways, LLC

# Presentation Objectives

- Overview of eDiscovery and Digital Forensics
- Discuss various methodologies and tools available to meet the challenges of Enterprise eDiscovery

# Agenda

- Setting the Stage (Forensics & eDiscovery)
- Challenges
- Methodologies
- Tools



# Setting the Stage *(Forensics & eDiscovery)*

# What we are/are not addressing

- This presentation focuses on the technical challenges of eDiscovery and regulation compliance.
- Complex legal issues arising surrounding eDiscovery & regulation compliance are not addressed here.
- Always seek legal counsel when dealing with such issues.

# Remember

- Digital Investigations
  - Inappropriate use
  - Fraud
  - IT Security
  - Other including criminal...

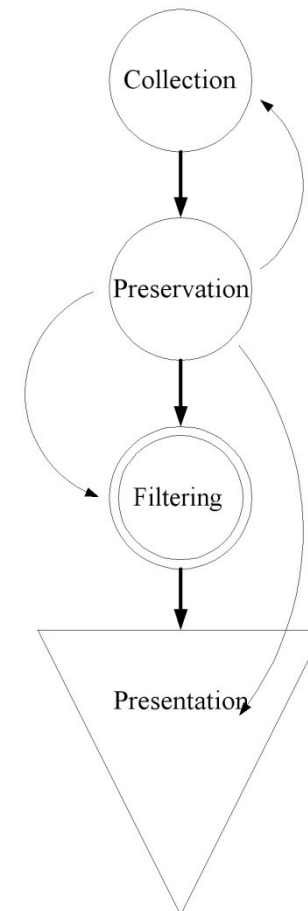
And...

- eDiscovery

*All are grounded in the computer forensics process*

# Digital Forensics Phases

- Collection (and identification)
- Preservation
- Filtering (analysis)
- Presentation



# Common Misconceptions

- “I can’t change anything!”
- “I have to create a bit-stream image of everything”
- “I must use a court approved tool”
- “This tool will do everything!”

# Confusion Enters

- While digital data is often dualistic (1 and 0's) the process cannot be.
- Each case is unique

## Always

1. Follow the four phases
2. Be as least intrusive/interactive as reasonably possible
3. EXTENT = Reasonableness + Case.

# EXTENT = Reasonableness + Case.

- What is reasonable in one case may not be in another.
- Always follow all four phases, but the extent to which you go, will be driven by the case and reasonableness.
- Example: it may be reasonable to bit-stream image 4 workstations, and completely search them for simple discovery, but not 150,000 workstations.

# Cost Factors Must be Managed

- Legal standards of “reasonableness”
- Weigh cost & capabilities
- Does not require perfection

Erin E. Kenneally & Christopher L. T. Brown, Risk Sensitive Digital Evidence Collection, Volume 2 Issue 2 Digital Investigations Journal, Available online at <http://www.compseconline.com/digitalinvestigation/tableofcontents.htm>, 2005

# Defining Reasonableness

- In the end the judge and jury will determine what is evidence and what is/was reasonable.
- Engineers should work with legal counsel early and often to establish and maintain a plan of action.
- Legal counsel should work with engineers early and often to establish and maintain a plan of action.



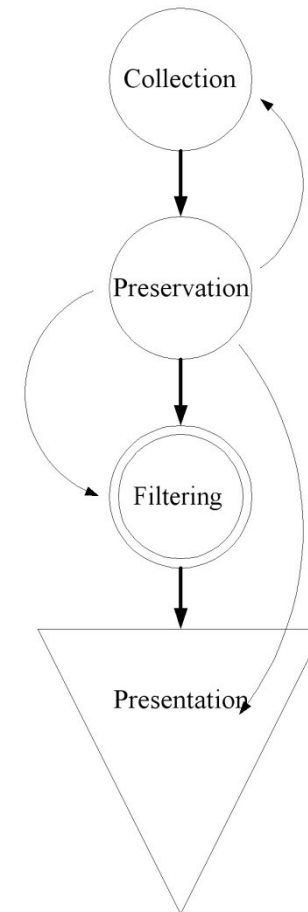
# Challenges

# Back on the eDiscovery Track

- Tremendous challenges arise even in medium sized networks of around 400 users.
- Technical and Legal teams should meet and focus on the four phases to help provide a smooth process.

# Review the Four Phases

- Identification & Collection
- Preservation
- Filtering & Analysis
  - (not discussed in detail here)
- Presentation
  - (not discussed in detail here)



# Defining Scope

- A great deal of information should go into defining scope of collection.
- Often too little time is spent on determining scope.
- Developing a key word list is not defining scope.
- The following checklist will help identify data locations.

# Identifying Data Locations (1)

1. Number, locations and types of employee used computers or data terminals. (include any authorized use of personal computers)
2. Number, locations and types of personal digital assistants or personal data storage devices issued to employees.
3. Number, locations and types of corporate servers including their purpose.
4. Operating systems, versions and patch levels in use.
5. Line of business applications in use. (include version and patch level).

# Identifying Data Locations (2)

6. Authorized general purpose applications in use. (include version and patch level).
7. Directory taxonomy or structure in use (servers based and host based).
8. File naming and storage standards.
9. Type of directory services in use (LDAP, Novel, etc).
10. Directory organization (users grouping).

# Identifying Data Locations (3)

11. Any server based logon scripts in use.
12. Network diagram specifically identifying data flow and devices which may provide log data and/or access control.
13. Firewall, Intrusion Detection System, and Identity Management configuration and logs.
14. Comprehensive Anti-virus, Anti-Spyware and Adware procedures addressing server and hosts protection.
15. Any documented vulnerability assessment and penetration tests conducted internally or by third parties.

# Identifying Data Locations (4)

16. Backup Tapes from on site, and off site storage.
17. Data backed up in any other forms such as network attached storage and storage area network snap shots or third party network storage providers.
18. Copies of any internal or third party data audit and control results.
19. Copies of published employee acceptable use policies.
20. Copies of published information technology guidelines, policies or procedures specifically addressing the handling, retention and storage of data.

# Checklists

- Previous checklist and others available in:  
“*Computer Evidence: Collection and Preservation*”,  
Christopher L. T. Brown, Charles River  
Media/Thompson Learning, Oct. 2006

# Create a Data Map

- Once we know the types and location of data... What then?
- The first hard steps are over 😊
- Now create a data map to start weighing how hard (dollars) it is going to be to get at the data you want.

# Data Map & Hard Decisions

- Types of data (email, database, user productivity documents, line of business applications)
- Can I search the data where it is or do I need to gatherer in a central location to identify relevant/responsive data?
- Each type of data will most likely have a greatly differing cost associated with the process.

# Most Sought After

- Email \*\*\*\*
- User productivity documents \*\*\*
- Database \*\*
- Line of business applications \*\*
- System Files & Unallocated (IR & Hacking Cases)\*
  - This information is helpful in many cases.

# Identification & Collection Approach



# Two Ends of the Spectrum

- Take it all (bit-stream image everything)
  - Too much in many eDiscovery situations.
- Take what you need (file copy)
  - Dangerous without great attention to detail and supplemental information.
  - Difficult to identify what is needed.

# First Extreme Benefits

- Cost are related/proportional to the size of the target disks.
- Protects original media from destructive tools & methods.
- Freeze the entire scene as it exists.
- Compartmentalizes the acquisition & analysis.

# Second end of Spectrum

- Targeting specific documents only.
  - Is not unwarranted in many situations
  - Often only email and/or user productivity documents are sought
  - Method of collection and handling needs to be:
    - least-intrusive
    - provide for integrity

# Compromise or Combo

- In eDiscovery reasonableness and cost factors often drive for some variance and/or combination of collection and identification approaches.
  - Think data reduction

# Choice

- Prior planning for eDiscovery often offers the producing party more choice than one extreme or another.
- Choice = Ability to Control Cost
- Even if outsourcing one person internally needs to possess expert level knowledge.



# Tools

# No Magic Bullet

- No one tool will do the entire job.
- Specific tools address specific areas of need.
- Technology changes rapidly and often a new tool or methodology must be implemented.

# Remember Most Sought After

- Email \*\*\*\*
- User productivity documents \*\*\*
- Database \*\*
- Line of business applications \*\*
- System Files & Unallocated (IR & Hacking Cases)\*
  - This information is helpful in many cases.

# Tool Area Breakdown

- Tier One: (all of)
  - Collection
  - Search
  - Analysis
  - Review
  - Production
- Tier Two: (some of)
  - Collection
  - Search
  - Analysis
  - Review
  - Production

# Digital Forensics Suites (1)

- Guidance Software, EnCase<sup>®</sup>
- Technology Pathways, ProDiscover<sup>®</sup>
- Access Data, Forensics Tool Kit (FTK)<sup>®</sup>
  
- All offer tier one and some tier two capabilities. (collection, search, analysis, some production)
- Currently very little work flow

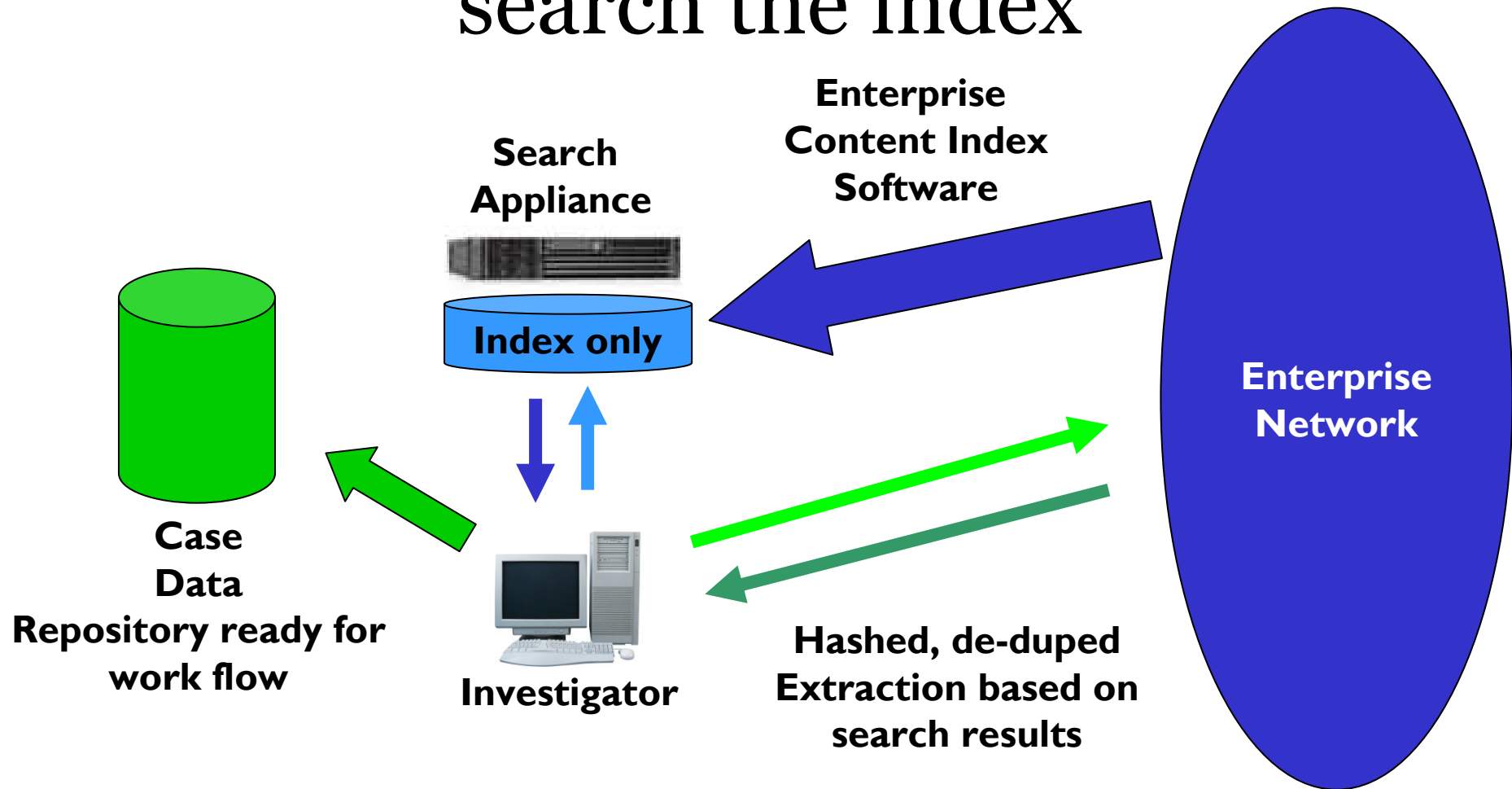
# Digital Forensics Suites (2)

- EnCase and ProDiscover have enterprise focused versions.
- Designed to meet the highest standards of reliability and completeness by working in a “clean” file system.
- The larger the enterprise or collection the more the basic design works against them., but both can be scripted and crawl the enterprise.

# Focus & Scope

- Remember many times only email and/or user documents are sought.
- Two approaches to supplement enterprise forensics suites for large scale collection:
  - Aggregate, then search repository for relevant/responsive data
  - Pre-index the enterprise, then search the index for relevant/responsive and extract live.
    - Offers the ability to reduce the corpus and collect at a lower level

# Pre-index the enterprise, then search the index



# Index Software

- Both methods are easily accomplished with:
  - Google<sup>®</sup> Enterprise
  - Other enterprise search indexers
  - dtSearch<sup>®</sup> Network (SMB)
  - Custom collection scripting
- Both work well with existing enterprise forensics suites
- Technology Pathways introducing packaged solution 4<sup>nd</sup> quarter 2007

# Discovery Workflow Management

- Concordance
- Summation
- Ringtail
- Etc...

# What about Enterprise Email

- Clearwell
- Kroll/Ontrack, Power Control Tools
- Paraben Network Email Examiner for Exchange EDB database

# Other Tier Two Tools

- Line of business applications and specialized email often require specialized analysis tools.
- Many Enterprise storage and software vendors are becoming aware of the need for discoverable solutions.

# COMPUTER EVIDENCE

## *Collection & Preservation*

- Teaches investigators how to ensure case integrity when dealing with computer evidence
- Provides a practical resource for collection and preservation that will help ensure legal acceptability
- Covers key areas such as rules of evidence, evidence dynamics, network topologies, collecting volatile data, imaging methodologies, and forensics labs and workstations
- Includes a CD-ROM with shareware and commercial demo software tools as well as document templates, worksheets, and references



Networking & Security Series

CHRISTOPHER L.T. BROWN

# Thank You Questions?

**Technology**  
**Pathways**

**703 First Street**  
**Coronado, Ca. 92118**

**Phone: 888-894-5500**  
**FAX: 619-435-0465**

**[www.TechPathways.com](http://www.TechPathways.com)**

Technology Pathways provides comprehensive, affordable computer forensic tools for Law Enforcement, Corporate and Government.

ProDiscover solutions include: investigations, incident response, computer forensics, and electronic discovery.

ProDiscover can forensically examine live systems over networks and has been accepted in criminal and civil proceedings.