Digital Records Forensics
Master Class

Luciana Duranti & Adam Jansen
CITRA
Oslo, Norway
16 September 2010

Digital Records Forensics Project

Archival Science and the Law

Records are "much better than navy yards, much more efficacious than munitions factories, as it is finer to win by reason rather than by violence, by right than by wrong“ Baldassarre Bonifacio 1632

Archival concepts are grounded in Roman Law

- Archives as a place—trusted custody
- Authenticity based on a chain of trusted custody—wax tablets
- Reliability based on antiquity and on form (Justinian Code)

Archival methods are born out of legislative acts

- Swedish Law of 1766—freedom of information act
- Decree 25 July 1793—public records belong to the people
- Decree of 1841—principle of respect des fonds

Archival science is at its heart as a legal science

Diplomatics and the Law

The rule of law was easily circumvented, authenticity and reliability needed to be tested using scientific methods, beyond Valla's textual criticism

Diplomatics (1681), a new science studying the nature, genesis, formal characteristics, structure, transmission and legal consequences of records, gave origin to **Palaeography, Sigillography, Heraldry, Philology, Exegesis, Semiotics, etc.**

The **Bella Diplomatica** gave origin to the **Law of Evidence:** by mid 18th century all faculties of law in Europe taught these "forensic" disciplines

Teaching of Forensic Disciplines

In addition to Jurists, other professionals were educated in forensic disciplines:

1811 Naples—Scuola per Archivisti

1821 Paris—Ecole des Chartes

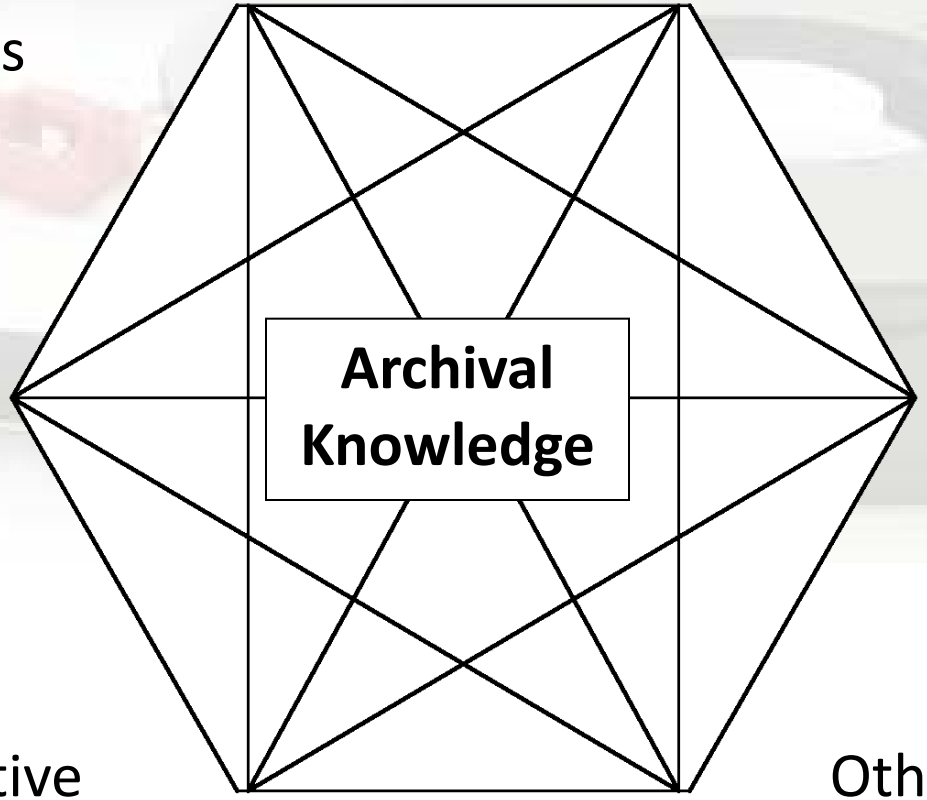
1821 Marburg—Archivschule

1854 Vienna—School on Auxiliary Disciplines of History

1925 Roma—Scuola Speciale per Archivisti e Bibliotecari

They all taught the forensic disciplines, legislation, and the archival methods rooted in legislation

Archivists



Diplomatics

Archival Methods

Palaeography

History of the Law

Administrative Law and Law of evidence

Other auxiliary disciplines

Digital Forensics

Digital Forensics is the use of scientifically derived and proven methods toward the collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events, or helping to anticipate unauthorized or inappropriate actions

Its methods are based on conceptual assumptions about records, trustworthiness, and recordkeeping

Records Managers, Archivists and Digital Forensics Experts

Records Managers and Archivists are called to act as forensics experts, e.g. ensuring the identity and integrity of digital records through time and attesting to it, and acquiring such records, often from obsolete systems or portable media, without altering them in the process

Digital forensic experts are called to act as archivists, e.g. identifying what digital materials fall under the definition of business records, and keeping them intact for as long as needed. They are also called to attest to and sometimes provide quality assurance for digital system that produce and/or contain records, to assess whether spoliation has occurred, to fulfill e-discovery requirements.

We Need Each Other's Knowledge

Digital forensic experts need our knowledge on

- Concepts of Record and Recordkeeping
- Records Trustworthiness

We need digital forensic experts' knowledge on

- Types of integrity
- Processes of access, reproduction, identification and extraction

Records Managers and Forensic Knowledge

The issue of **what is a record in the digital environment** keeps coming up at trials and in political discussions.

- British Columbia Rail case: the judge pointed out that legislation speaks of preserving "records," and the Liberal MLA Ralph Sultan asked "What is the definition of a record?" referring "to the controversy over to what extent e-mails qualify"
- The Supreme Court of Canada is deciding whether hyperlinks in a text are akin to footnotes or make of the material to which they connect the reader a component of the document being read

Record: the Legal and Archival Views

- A document made or received in the usual and ordinary course of business and kept for the purposes of such business at a time close to the fact at issue by a person responsible for doing so (laws of evidence)
- A document made or received in the course of activity as a by-product of or instrument for it and kept for action or reference (diplomatics/archival science)

Record:
the Computer Science View

- Any document that is exclusively machine readable.
- Comprised of Binary Data – stored in Base2
- One value is a bit – short for BInary digiT

0 = 0
1 = 1
2= 10
3 = 11
4 = 100
5 = 101
6 = ????
110

Bits and Bytes

- Bits bundled into 8-bit collections
- 256 values ranging from 0 to 255

  0= 00000000
  1= 00000001
  2= 00000010
  255= 11111110
  256= 11111111

- ASCII characters given value 1-127
- Example:

  F o u r a n d s e v e n
  70 111 117 114 32 97 110 100 32 115 101 118 101 110 32 = 00100000

Big Bytes

Name	Abbr	Size
Kilo	K	2^10 = 1,024
Mega	M	2^20 = 1,048,576
Giga	G	2^30 = 1,073,741,824
Tera	T	2^40 = 1,099,511,627,776
Peta	P	2^50 = 1,125,899,906,842,624
Exa	E	2^60 = 1,152,921,504,606,846,976
Zetta	Z	2^70 = 1,180,591,620,717,411,303,424
Yotta	Y	2^80 = 1,208,925,819,614,629,174,706,176

Digital Record: Our View

- **Act**: an action in which the record participates or which the record supports (naturalness and impartiality)
- **Persons Concurring to Its Creation**: author, writer, originator, addressee, and creator
- **Archival Bond**: explicit linkages to other records inside or outside the system (uniqueness)
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, technological (interrelatedness)
- **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**

Fixed Form

- An entity has fixed form if its binary content is stored so that the message it conveys can be rendered with the **same documentary presentation** it had on the screen when first saved (different digital presentation: Word to .pdf)
- An entity has fixed form also if the same content can be presented on the screen in several different ways in a **limited series of possibilities:** we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table)

Stable Content

- An entity has stable content if the data and the message it conveys are **unchanged and unchangeable**, meaning that data cannot be overwritten, altered, deleted or added to

- **Bounded Variability**: when changes to the documentary presentation of a determined stable content are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications

The Parts of a Digital Record

- **Formal Elements:** constituent parts of the record documentary form as shown on its face, e.g. address, salutation, preamble, complimentary close
- **Metadata:** the attributes of the records that demonstrate its identity and integrity
- **Digital Components:** stored digital entities that either contain one or more records or are contained in the record and require a specific preservation measure

Stored and Manifested Record

- **Stored record:** it is constituted of the digital component(s) used in re-producing it, which comprise the data to be processed in order to manifest the record (content data and form data) and the rules for processing the data, including those enabling variations (composition data)
- **Manifested record:** the visualization of the record in a form suitable for presentation to a person or a system. Sometimes, it does not have a corresponding stored record, but it is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database)

Static and Interactive Records

- **Static Records:** They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages
- **Interactive Records**: They present variable content, form, or both, but the rules governing the content and form of presentation are fixed. Ex. Interactive web pages, online catalogs, records enabling performances

Digital Record: Digital Forensics View

Problematic in relation to the hearsay rule: in common law, documents are hearsay because they contain human statements made outside the court—if they are records they fall under the business records exception to the rule

- **Computer Stored Documents:** They contain human statements and are considered hearsay (they can be tested for truthfulness and accuracy under the business records exception to the hearsay rule): e.g. e-mail messages, word processing documents, and Internet chat room messages.

Digital Record: Digital Forensics View (cont.)

- **Computer Generated Documents**: They do not contain human statements, but they are the output of a computer program designed to process input following a defined algorithm (they can be tested for authenticity on the basis of the functioning of the computer program): e.g. server log-in records from Internet service providers, ATM records.
- **Computer Stored & Generated:** e.g. a spreadsheet record that has received human input followed by computer processing (the mathematical operations of the spreadsheet program).

Substantive Evidence vs. Demonstrative Evidence

Records Trustworthiness: Our View

- **Reliability:** The trustworthiness of a record as a statement of fact, *based on* the competence of its author, its completeness, and the controls on its creation
- **Accuracy:** The correctness and precision of a record's content, *based on* the above, and on the controls on content recording and transmission
- **Authenticity:** The trustworthiness of a record that is what it purports to be, untampered with and uncorrupted, *based on its* identity, integrity and on the reliability of the system in which it resides

Authenticity: Our View

- **Identity:** The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records (e.g. date, author, addressee, subject, identifier).
- **Integrity:** A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered (e.g. chain of custody, security, technical changes).

Records Trustworthiness. The Digital Forensics View: Reliability

- **Reliability:** the trustworthiness of a record as to its *source*, defined in digital forensics in a way that points to either a reliable person or a reliable software.

  This would be an open source software, because the processes of records creation and maintenance can be authenticated either by describing a process or system used to produce a result or by showing that the process or system produces an accurate result

Records Trustworthiness. The Digital Forensics View: Accuracy

- **A component of authenticity and, specifically, integrity.** Digital entities are guaranteed accurate if they are repeatable.
- **Repeatability**, which is one of the fundamental precepts of digital forensics practice, is supported by the documentation of each and every action carried out on the evidence.
- **Open source software** is the best choice for assessing accuracy, especially when conversion or migration occurs, because it allows for a practical demonstration that nothing could be altered, lost, planted, or destroyed in the process

Records Trustworthiness. The Digital Forensics View: Authenticity

The data or content of the record are what they purport to be and were produced by or came from the source they are claimed to have been produced by or come from. Again, the term "source" is used to refer to either a person (physical or juridical), a system, software, or a piece of hardware.

Like in diplomatics, authenticity implies integrity, but the opposite is not true, that is, integrity does not imply authenticity.

Integrity: Our View

The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity

The same for data, documents, records, copies, systems

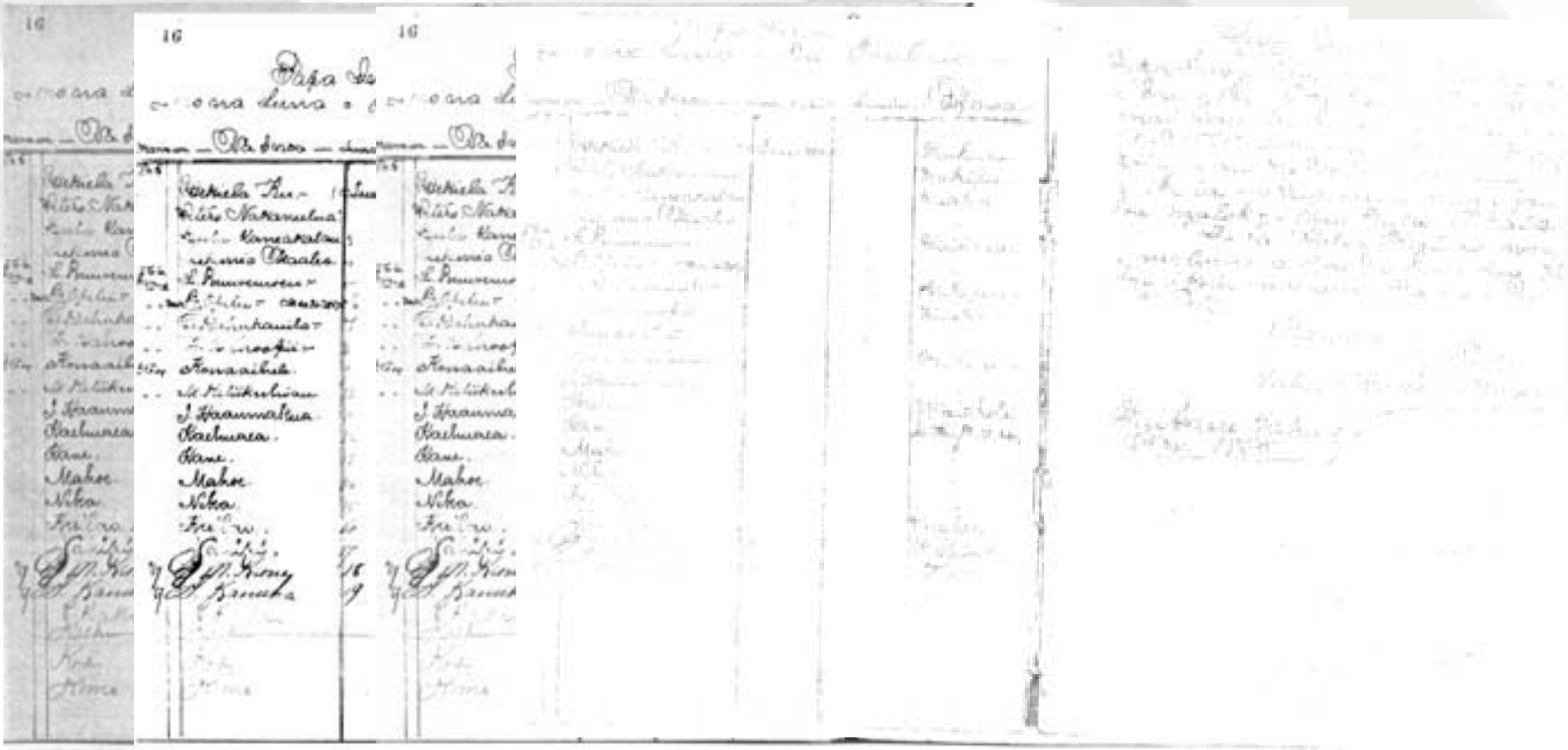
Integrity
Digital Forensics View

Data integrity: the fact that data are not modified either intentionally or accidentally "without proper authorization."

What is Bitwise Integrity?

- Maintaining the original bits in a complete and unaltered state from the time of capture
- Exact and same order and value of the bits
- Small change in a bit means a very different value presented on the screen or action taken in a program or database.

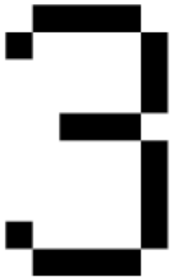
Loss of Fidelity: Analog vs. Digital



Digital Records Forensics Project

Loss of Fidelity (cont.)

- If Original Bits 101
- Change state to 110
- Continues to a 011



- Same bits, but
  Different value

Data Alteration

- Intentional alteration preventable through permission and access controls
- Accidental alteration avoidance requires additional hardware and/or software be in place

Data Alteration (cont.)

- Requires method of determining if the record has been altered, maliciously or otherwise
- Cannot rely on file size, dates or other file properties
- Requires audit logs and strong methods

Checksum

- Form of data authentication
  - If checksum doesn't match, data corrupted or incomplete
- Add up value of bits in a packet
  - If less than 255 actual value used
  - More than 255 value after divided by 256
- Example:

Byte1	Byte2	Byte3	Byte4	Byte5	Byte6	Byte7	Byte8	Total	Checksum
212	232	54	135	244	15	179	80	1151	127

1,151 / 256 = 4.496 (round to 4)
4 x 256 = 1,024

1,151 - 1,024 = 127

Parity Check

- Counts the number of ‘set’ bits in a seven bit stream
- If number is even, parity bit is zero
- If number is odd, parity bit is one
- Therefore every 8bit sequence will be even
- Both sender and receiver must agree on parity

HASH Algorithms

- Computed from the base number using an algorithm
- Nearly impossible to derive without original data
- Typically use 128bit or greater algorithms, that's 2^{128}
  - Like trying to find a grain of sand in the Sahara
- Example:

Input	Hash	Value
10,667	Input x143	1,525,381

HASH Values

- Compresses bits of a message into a fixed-size value
- Extremely difficult to come up with original record based on hash value
- Common Hash functions
  - SHA-1 160 bit
  - RIPEMD-160 160bit
  - MD5 – 128 bit

Integrity
Digital Forensics View

Duplication integrity: the fact that, given a data set, the process of creating a duplicate of the data does not modify the data (either intentionally or accidentally) and the duplicate is an exact bit copy of the original data set. Digital forensics experts also link duplication integrity to time and have considered the use of time stamps for that purpose.

Data Duplication

- RAID
  - Redundant Array of Independent Disks
- Hardware or software based solution
- Data automatically written to multiple drives
- Resource intensive – performance benefits
- Most common are 0,1, 5

Integrity
Digital Forensics View (cont.)

Computer integrity: the computer process produces accurate results when used and operated properly and it was so employed when the evidence was generated.

System Integrity: a system would perform its intended function in an unimpaired manner, free from unauthorized manipulation whether intentional or accidental

Both imply **hardware and software integrity**

What is Computer or System Integrity?

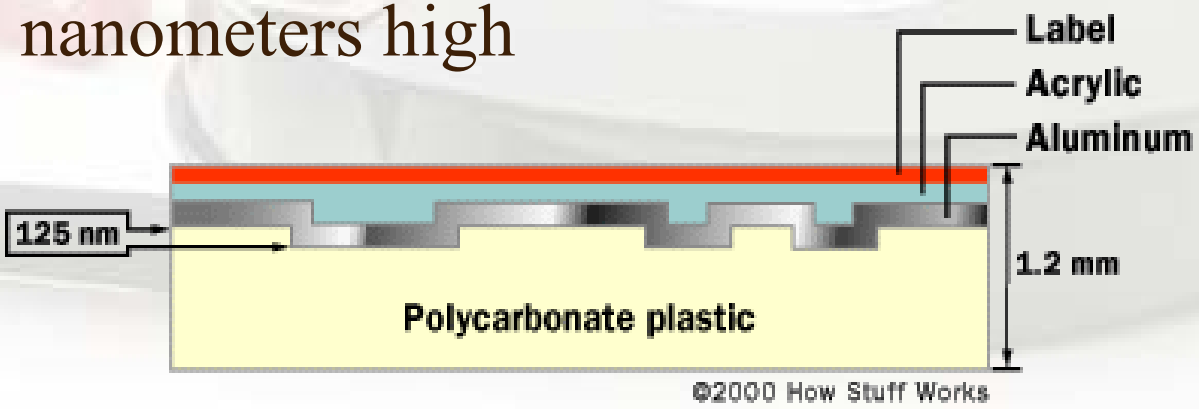
- Sufficient security measures to prevent unauthorized or untracked access to the computers, networks, devices, or storage.
- Stable physical devices that will maintain their ‘statefulness’ – the value they were given in maintained until authorized to change.
  - Data
  - Users/permissions
  - Passwords
  - Logs
  - Firewalls

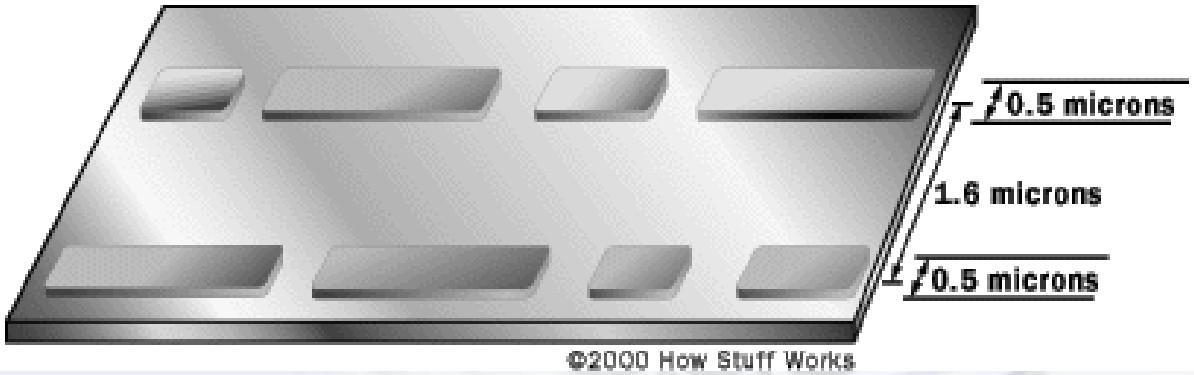
Software Integrity: Cathedral or the Bazaar?

- Black Box vs. White Box testing
- Closed Source vs. Open Source software

Device Integrity: CDs/DVDs

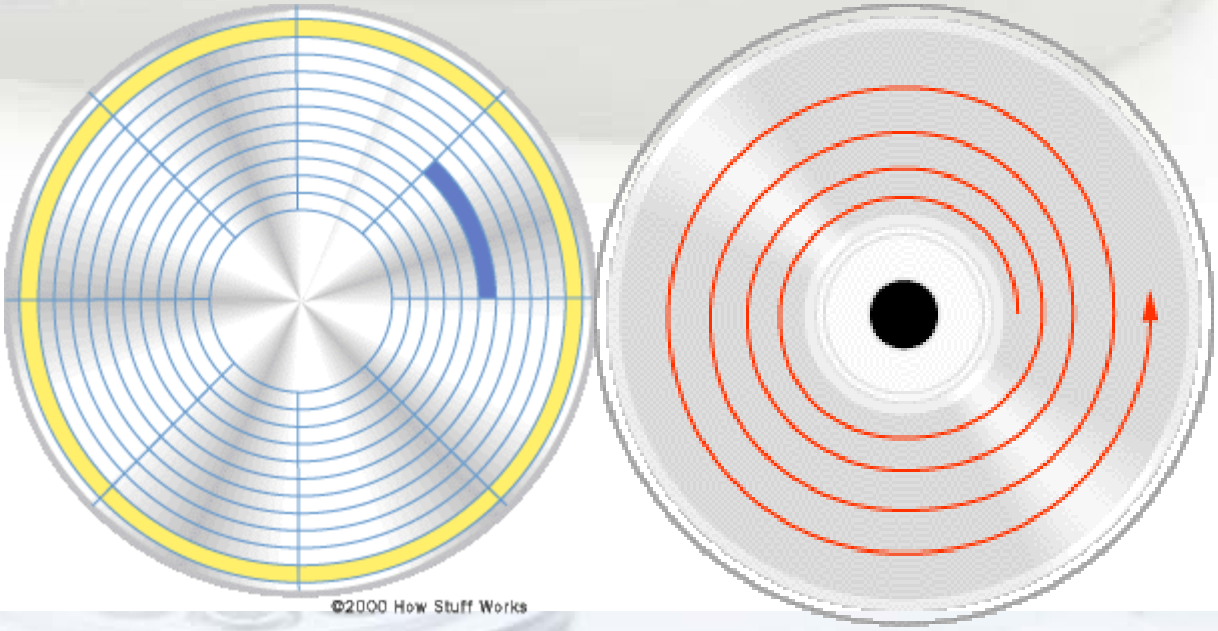
Data track .5 micros in width, .83 microns long and 125 nanometers high



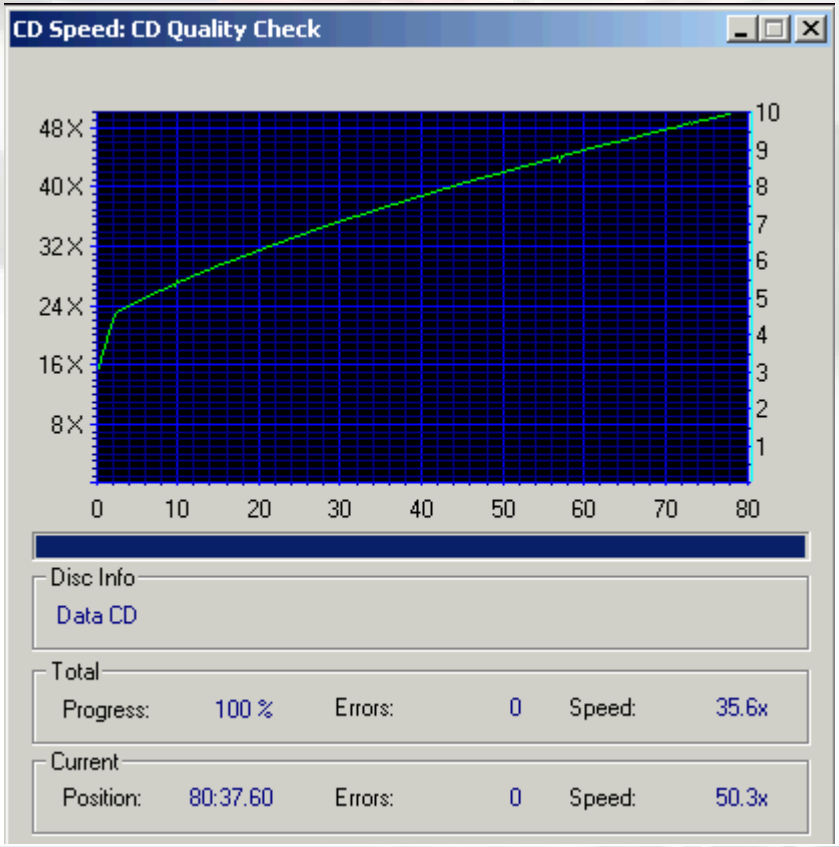


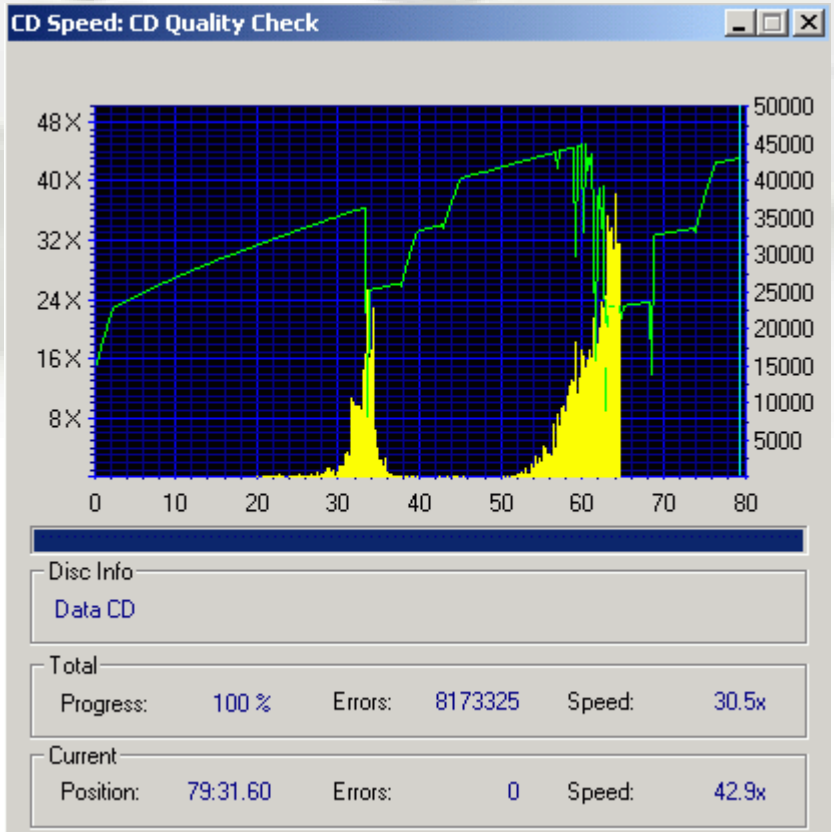
Sectors and Tracks

- Tracks are the concentric circles (yellow)
- Sectors are pie shaped wedges of tracks (blue)



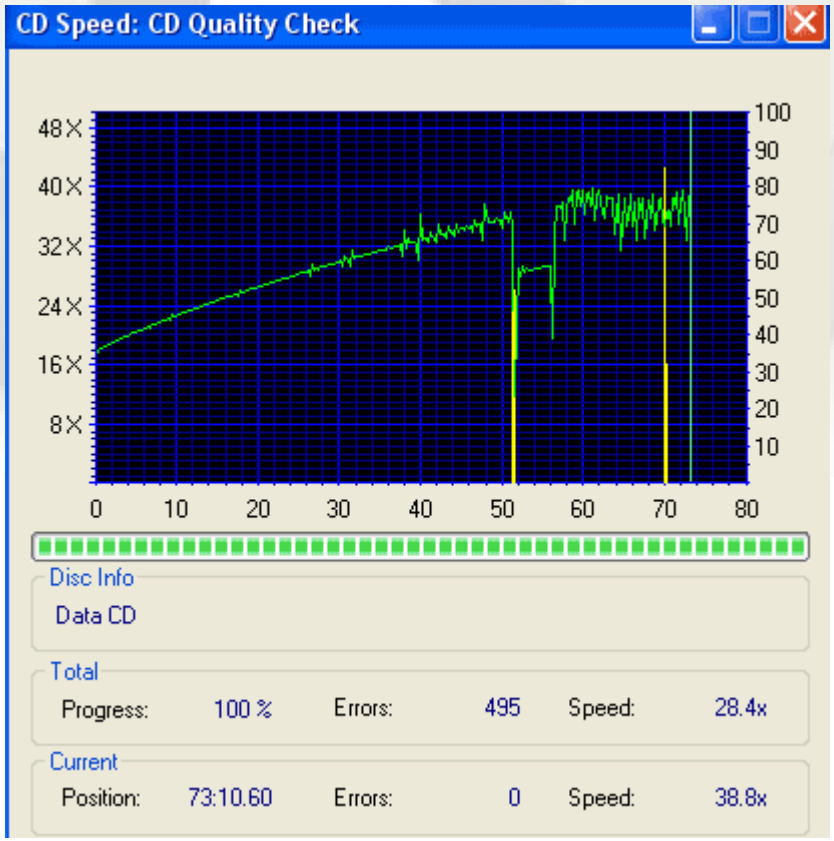
Burn Quality





Long-term Preservation???

After two years….



Virtual Environments

- Transformation of hardware into software; allow software to replicate functions of hardware
- Can be system implementations, resources of devices
- Limited view, constrained to stay ‘inside the box’

VM Requirements

Popek and Goldberg Virtualization Requirements:

Fidelity – Exhibit identical behavior in VM as ‘real world’

Safety – VM in complete control of resources

Performance – machine instructions must run without help

Example of a Virtual Machine

Types of Virtual Machines

- Emulation – (legacy software)
- Native – (sandbox)
- Operating Systems – (optimization)
- Application – (standalone installations)
- Cross Platform – (Win on a Mac)
- Resource Virtualization – (Drives and CDs)

Why VMs?

- Ability to completely isolate from external resources
- Good, known, *unaltered* state at every start-up – free of corruption or alteration
- Ensure that compromise in one area does not affect others

System Logs and Auditing

Set of files *automatically* created to track the actions taken, services run, or files accessed or modified, at what time, by whom and from where

- Web logs
- Access logs
- Transaction logs

Typical Web Log

- Client IP Address
- Request Date/Time
- Page Requested
- HTTP Code
- Bytes Sent
- Browser Type
- OS Type
- Referrer

Typical Access Log

- User account ID
- User IP address
- File Descriptor
- Bind record results
- Actions taken upon record
- Unbind record
- Closed connection

Typical Transaction Log

- History of actions taken on a DBMS to ensure ACID over crashes
- Sequence number
- Link to previous log
- Transaction ID
- Type
- Updates, commits, aborts, completes

Example of a Log

Auditing Logs

- Increasing required by law (SOX, HIPPA) to demonstrate integrity of the system
- Properly configured, restricted provide checks and balances
- Ability to determine effectively of security policies
- Ability to trap errors that occur
- Provide instantaneous notification of events
- Monitor many systems and devices through ‘dashboards’

Auditing Logs (cont.)

- Ability to determine accountability of people, resources for measure events
- Provide the necessary snapshot for post-event reconstruction (‘black-box’)
- Answers Who-What-Where-When
- Only if retained for sufficient time (space vs. money vs. risk vs. knowledge)

Digital Forensics Integrity

Formalized legal requirements for the collection, recovery, interpretation and presentation of digital evidence.

Example: UK ACPO:

- No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.
- In exceptional circumstances, where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
- An audit trail or other record of all processes applied to computer based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

Assessment of Integrity Digital Forensics View

The assessment is based on **repeatability, verifiability, objectivity** and **transparency**

Inference of system integrity derives from the fact that:

- the theory, procedure or process on which the system design is based has been tested or cannot be tampered with
- it has been subjected to peer review or publication (standard)
- its known or potential error rate is acceptable
- it is generally accepted within the relevant scientific community

Integrity
Digital Forensics View (cont.)

Non-interference: the method used to gather and analyse [or acquire and preserve] digital data or records does not change the digital entities

Identifiable interference: if the method used does alter the entities, the changes are identifiable

These principles, which embody the ethical and professional stance of digital forensics experts, are consistent with the traditional impartial stance of the archivist, as well as with his/her new responsibility of neutral third party, of trusted custodian

Authentication: Our View

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the **digital signature**. Functionally equivalent to medieval seals (not signatures): verifies origin (identity); certifies intactness (integrity); makes record indisputable and incontestable (non-repudiation)

But, medieval seals were associated with a person; digital signatures are associated with a person and a record

Authentication: The Digital Forensics View

Proof of authenticity provided by a **witness** who can testify about the existence and/or substance of the record on the basis of his/her familiarity with it, or, in the absence of such person, by **a computer programmer** showing that *the computer process or system produces accurate results when used and operated properly and that it was so employed when the evidence was generated.*

The strength of circumstantial digital evidence could be increased by metadata which records (1) the exact dates and times of any messages sent or received, (2) which computer(s) actually created them, and (3) which computer(s) received them.

Other Means of Authentication

A chain of legitimate custody is ground for inferring authenticity and authenticate a record.

Digital chain of custody: the information preserved about the record and its changes that shows specific data was in a particular state at a given date and time.

A declaration made by an expert who bases it on the **trustworthiness of the recordkeeping system** and of the procedures controlling it (quality assurance).

Other Means of Authentication (cont.)

Biometric identification systems and cryptography **are not** considered the prevalent means of authentication.

Inference of system integrity: Circumstantial evidence that a system would perform its intended function in an unimpaired manner, free from unauthorized manipulation of the system, whether intentional or accidental.

Other Relevant Concepts

- **Chain of Custody vs. Chain of Documentation** (the conditions under which the evidence is gathered, the identity of all evidence handlers, duration of evidence custody, security conditions while handling or storing the evidence, and the manner in which evidence is transferred to subsequent custodians each time such a transfer occurs)
- **Prevention vs. Preparation** (for detection and response)
- **Identification and Acquisition vs. Search and Seizure**
- **Cryptography vs. Steganography** (a covert form of information hiding)
- **Copy vs. Image** (a bit by bit reproduction of the storage medium)

Forensic View: Disk Image

A full disk copy of the data on a storage device – regardless of operating system or storage technology – made prior to performing any forensic analysis of the disk.

Creating a disk image is important in forensics for several reasons:

- Ensure that disk information is not inadvertently changed.
- Reproduce forensic test results on the original evidence.
- Capture information normally invisible to the operating system when in use (including memory, page files, boot sector, BIOS)

Why not Copy?

- Copy is selective duplicate of files
  - Can only copy what you can see
- Rarely includes confirmation of completeness
- Moved as Individual files
- Provides incomplete picture of the digital device

Example of Disk Image

What for?

In a records management context:

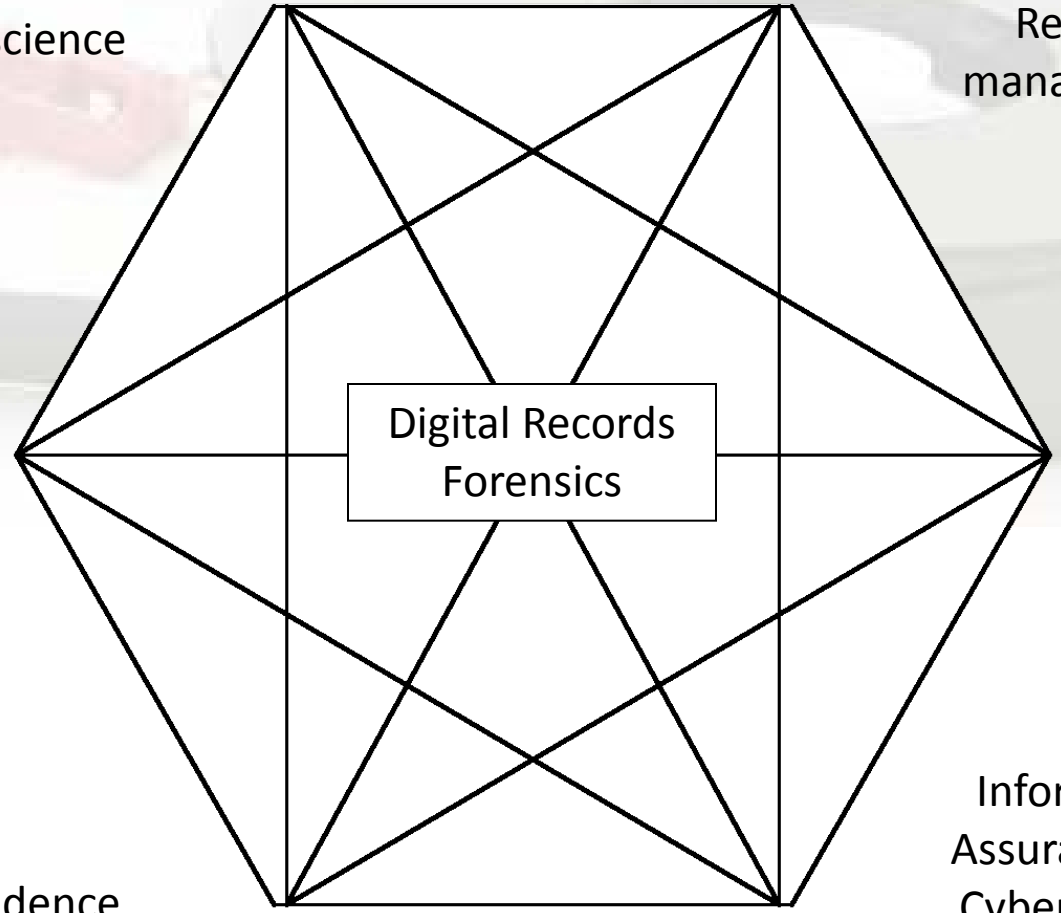
- **Discovery**: the compulsory disclosure of pertinent facts or documents to the opposing party in a civil action, usually before a trial begins. The **discovery process** is the process of identifying, preserving, collecting, reviewing, analyzing and producing information during legal actions.
- **E-discovery:** the extension of the discovery process to information stored electronically (ESI), including email, instant messages, word processing files, spreadsheets, social networking content, and anything else stored on desktops, laptops, file servers, mainframes, smartphones, employees' home computers or on a variety of other platforms.
- **Information Assurance and Cybersecurity**

What for? (cont.)

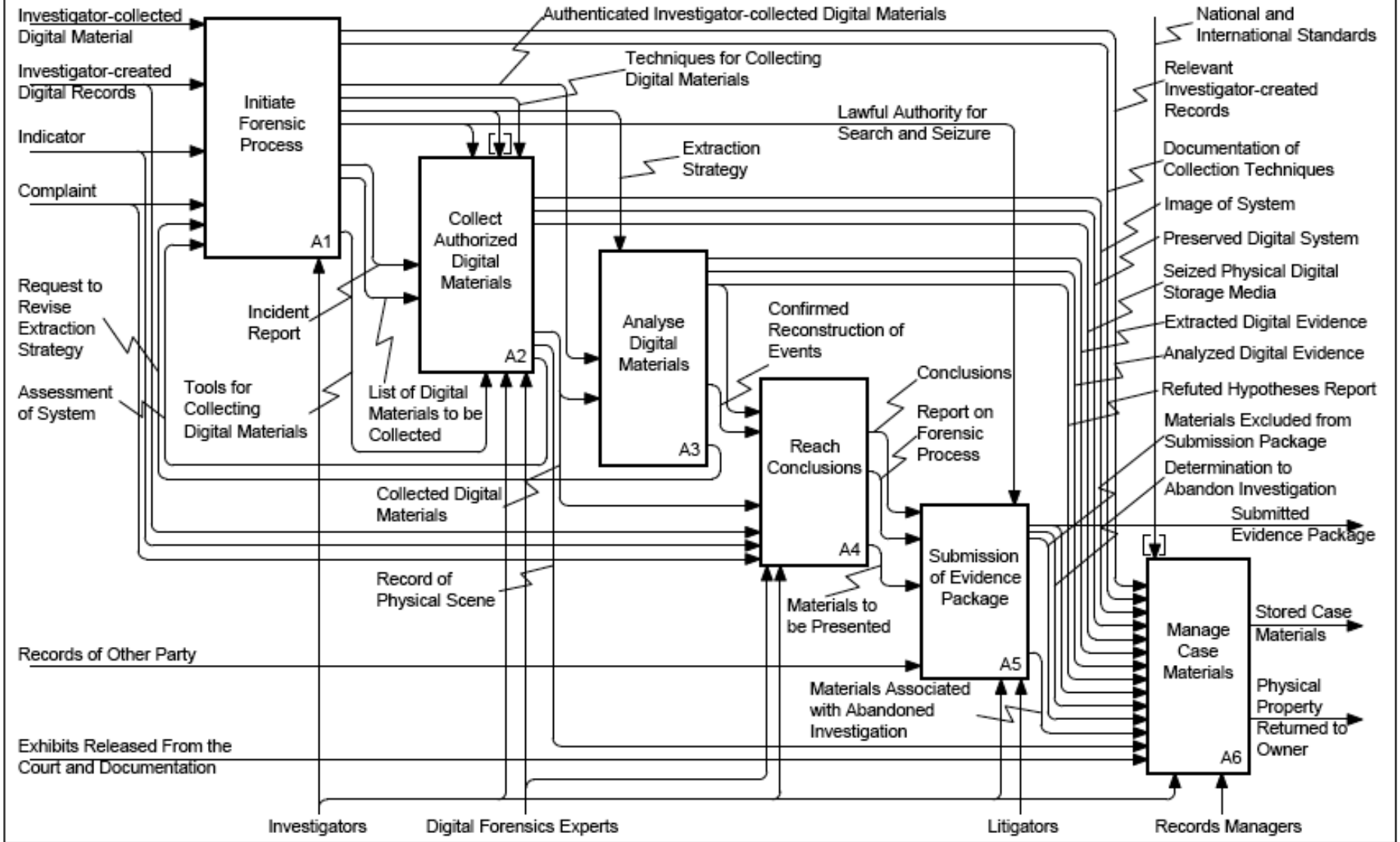
In an historical archives context:

- Extraction of digital materials from obsolete hardware and software
- Authentication of digital material of uncertain provenance
- Documentation of the technological context of records
- Protection of digital material over the long term
- E-discovery

Digital Records Forensics



USED AT:	AUTHOR:	DATE: 14-Nov-2008		WORKING	READER	DATE	CONTEXT:
	PROJECT: DRF Model	REV: 05-Jun-2010		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION			A-0



NODE:	TITLE:	NUMBER:
A0	Conduct Digital Forensics	

CASE STUDY EXAMPLE: VPD

Example of a Forensic Case

Questions?