Identification and Collection

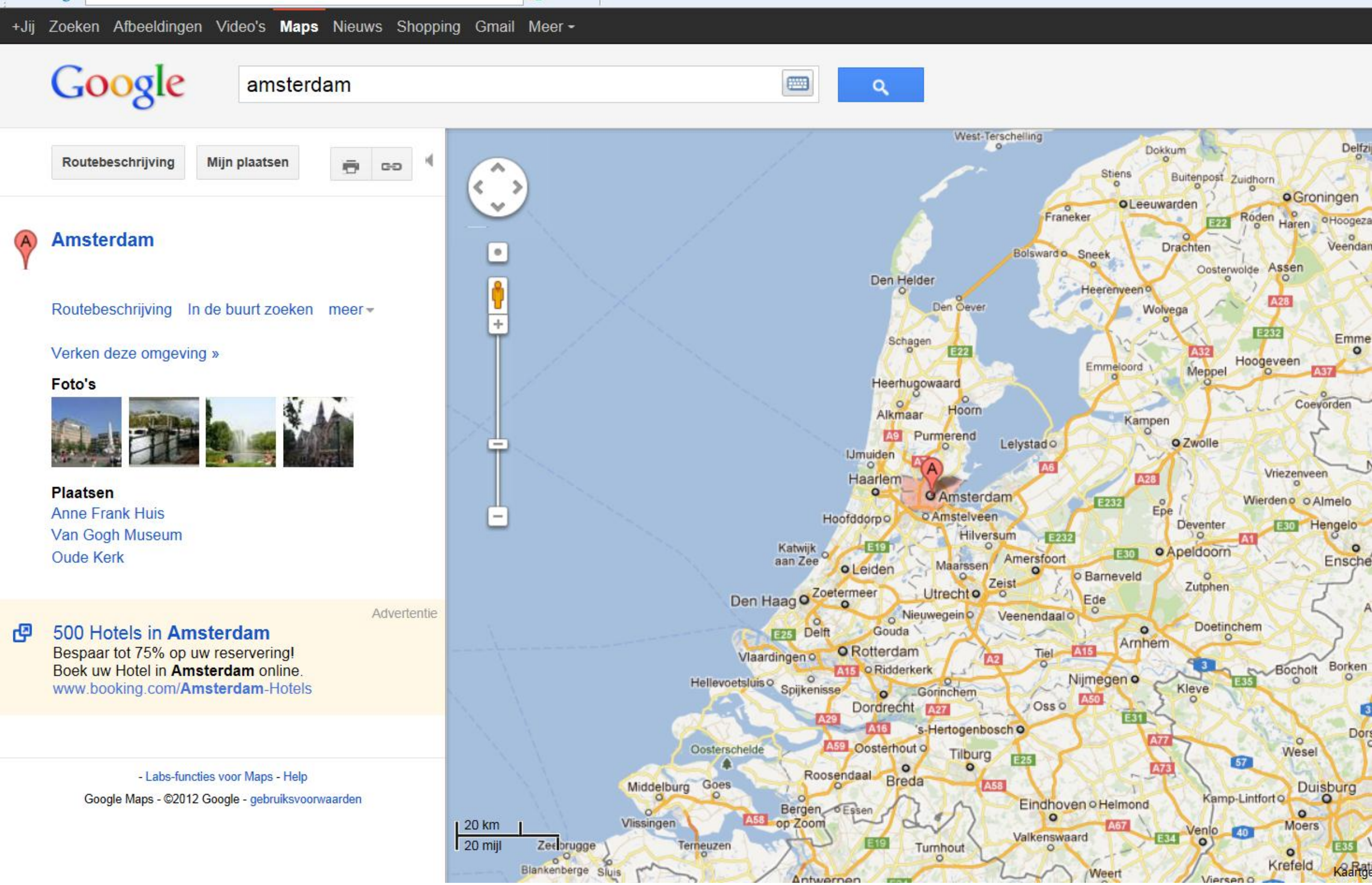
Seminar on E-Discovery, February 9th, 2012,
College of Information Studies, University of Maryland

Dr. Hans Henseler

Amsterdam University of Applied Sciences, The Netherlands

Information Technology & Computer Science
E-Discovery Lab

Hogeschool van Amsterdam
Media, Creatie en Informatie



Amsterdam

Routebeschrijving In de buurt zoeken meer

Verken deze omgeving »

Foto's

Plaatsen

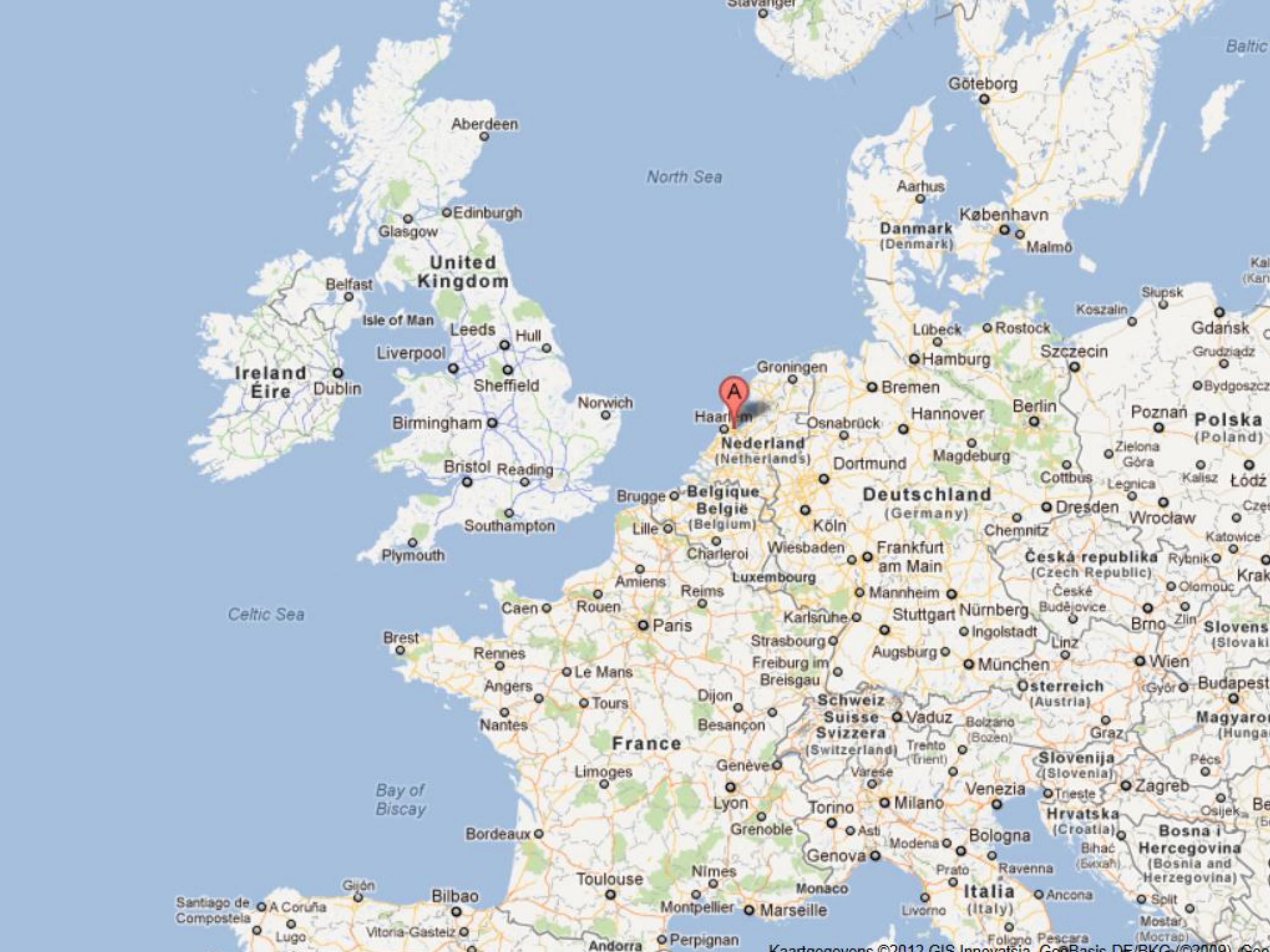
Anne Frank Huis

Van Gogh Museum

Oude Kerk

- Labs-functies voor Maps - Help

Google Maps - ©2012 Google - gebruiksvoorwaarden

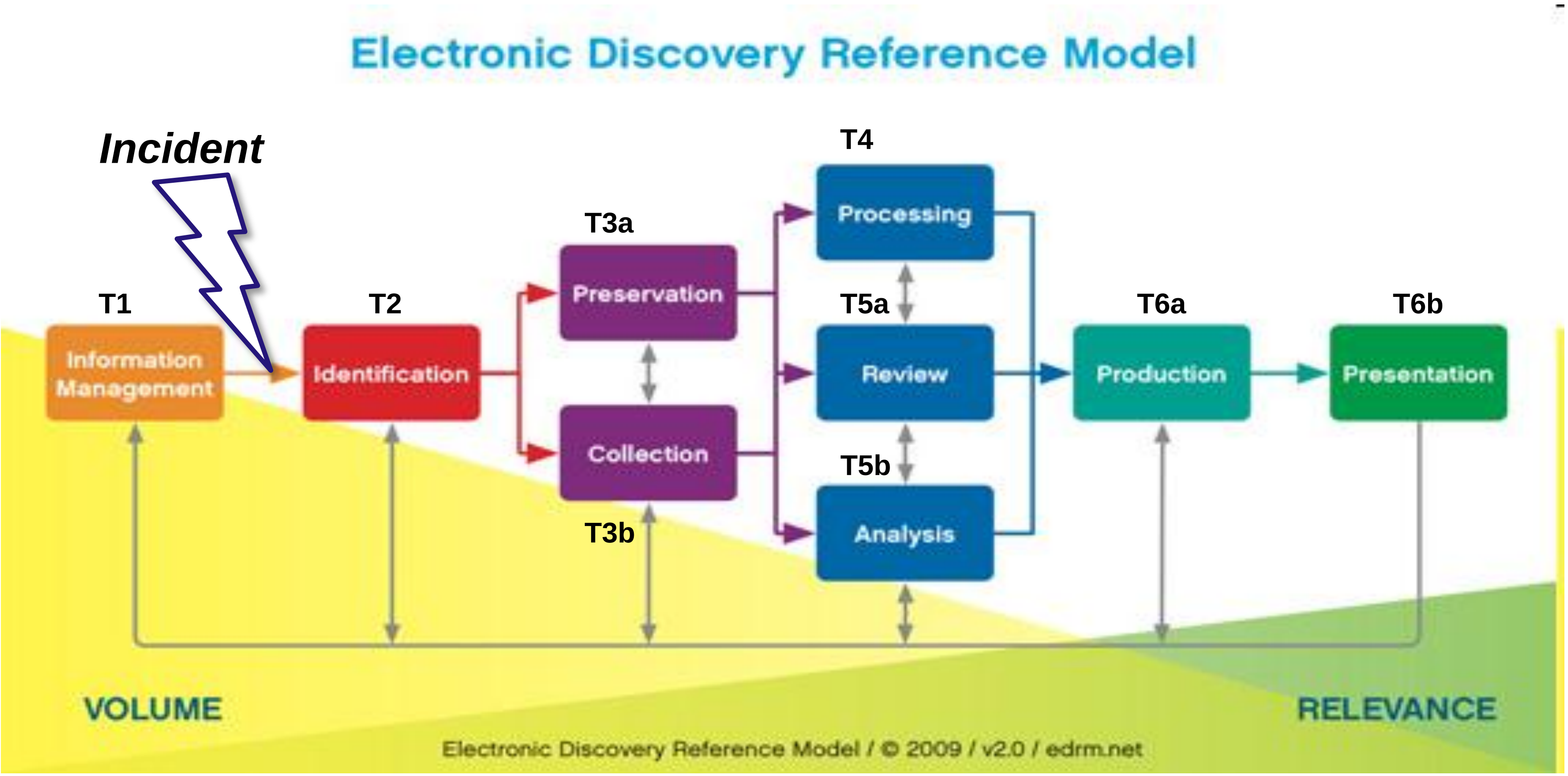


Kaartgegevens ©2012 GIS Innovatsia, GeoBasis-DE/BKG (©2009), Goo

Dr. Hans Henseler

- Ph.D. computer science (1993)
- Netherlands Forensic Institute (1992-1998)
- Netherland Institute of Applied Research (1998-2000)
- CTO at ZyLAB (2000-2006)
- Director at Pricewaterhouse Coopers (2006-2010)
- Adjunct Professor HvA (2009-)
- Partner at Fox-IT (2011-)

1. Recap: EDRM


Electronic Discovery Reference Model
Incident
T1
Information Management
T2
Identification
T3a
Preservation
Collection
T3b
T4
Processing
T5a
Review
T5b
Analysis
T6a
Production
T6b
Presentation
VOLUME
RELEVANCE
Electronic Discovery Reference Model / © 2009 / v2.0 / edrm.net

1. Recap: Track 1: Information Management

GOAL:

Develop defensible retention policies and e-discovery processes

HOW:

By managing all information sources:

- Complete information lifecycle: From creation, through using to archival and destruction.

Track 2: Identification

GOAL:

Determine what should be preserved and collected

HOW:

By identifying and localising potential sources of information:

- what kind of information is required?

- relevant time period?

Track 3a: Preservation

GOAL:

Preserve data to avoid spoliation claims/sanction

HOW:

By securing information that may potentially be relevant

- By ensuring that information can not be altered or destroyed.

Track 3b: Collection

GOAL:

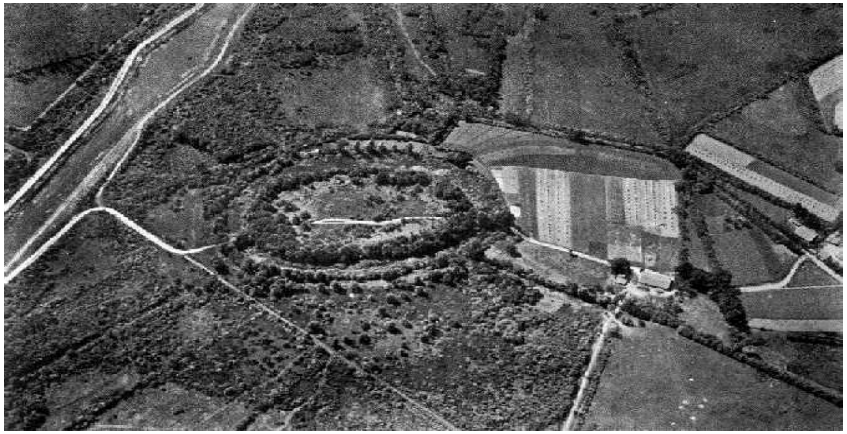
Retrieve forensically sound copies of critical data

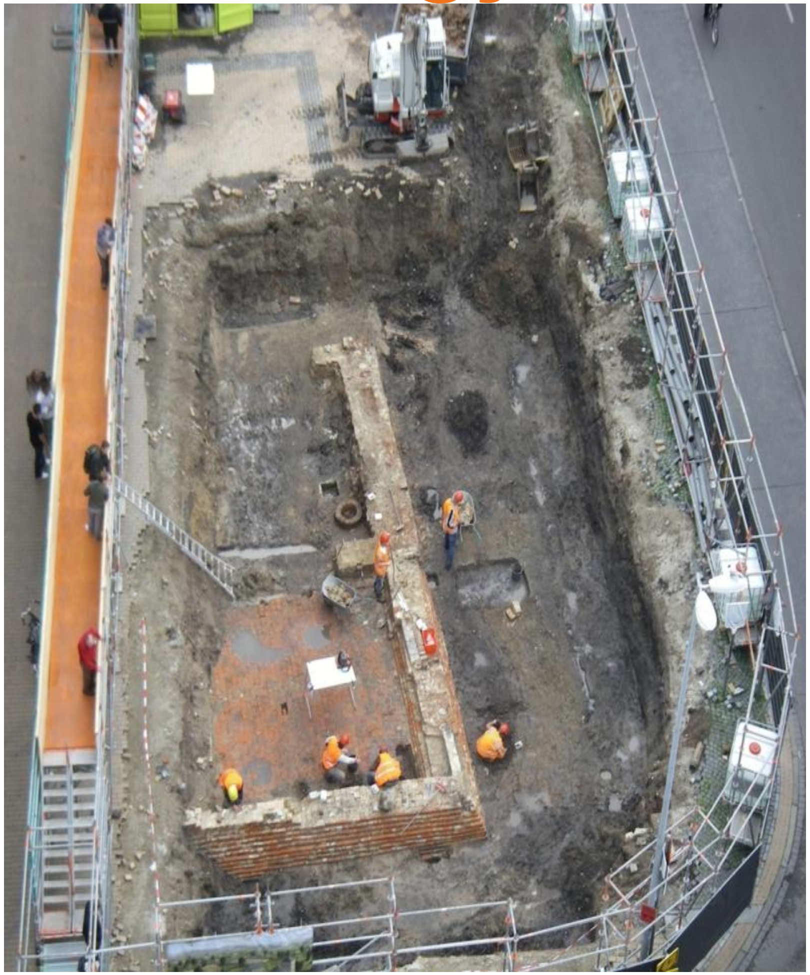
HOW:

By making digitale copies of electronic stored information and related meta data (information context)

- In such a way that the integrity and authenticity of the information can be verified

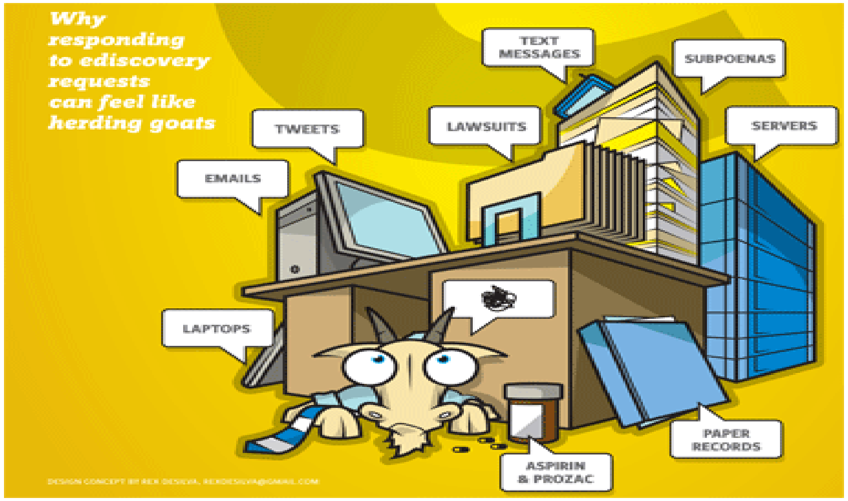
E-Discovery and Archeology



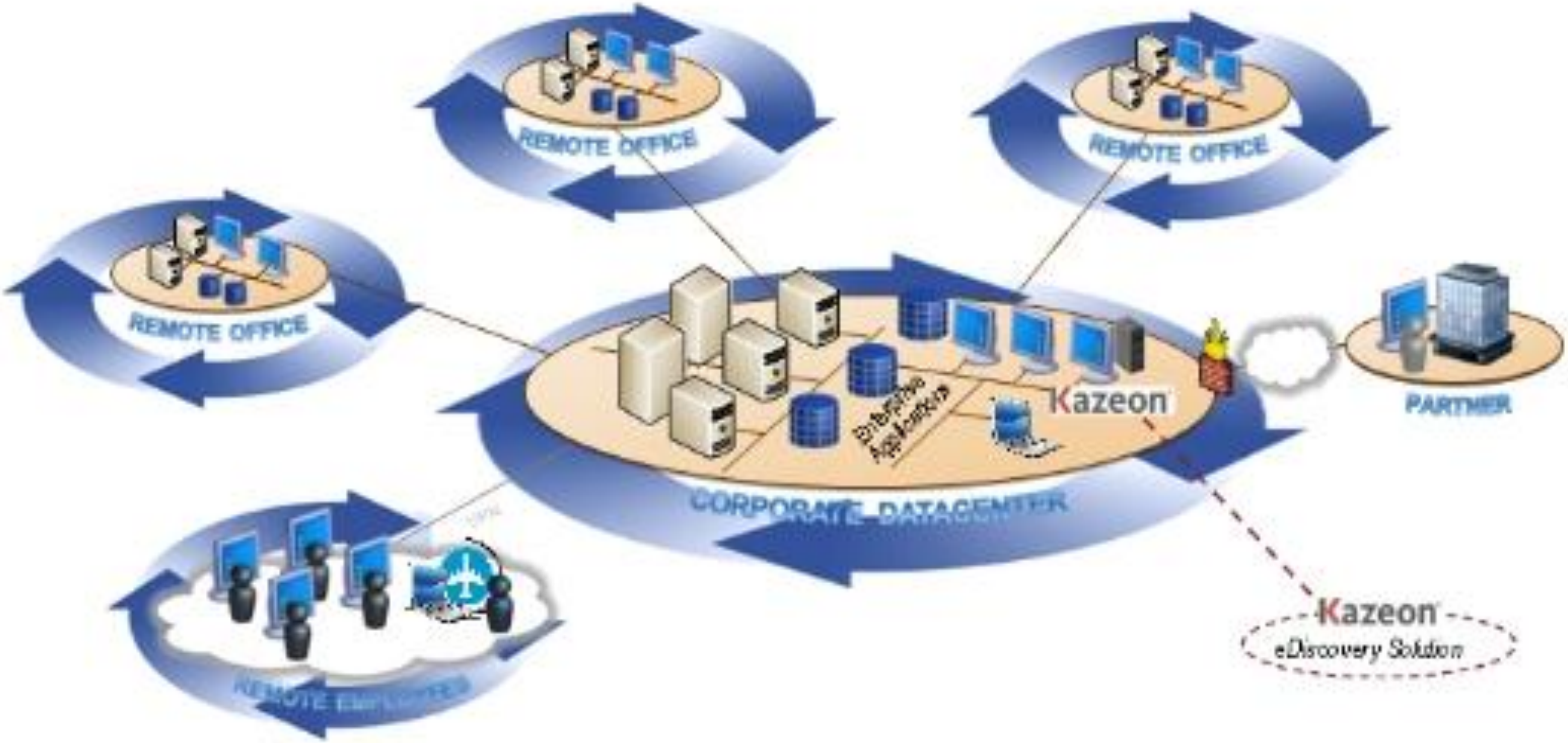


Identification

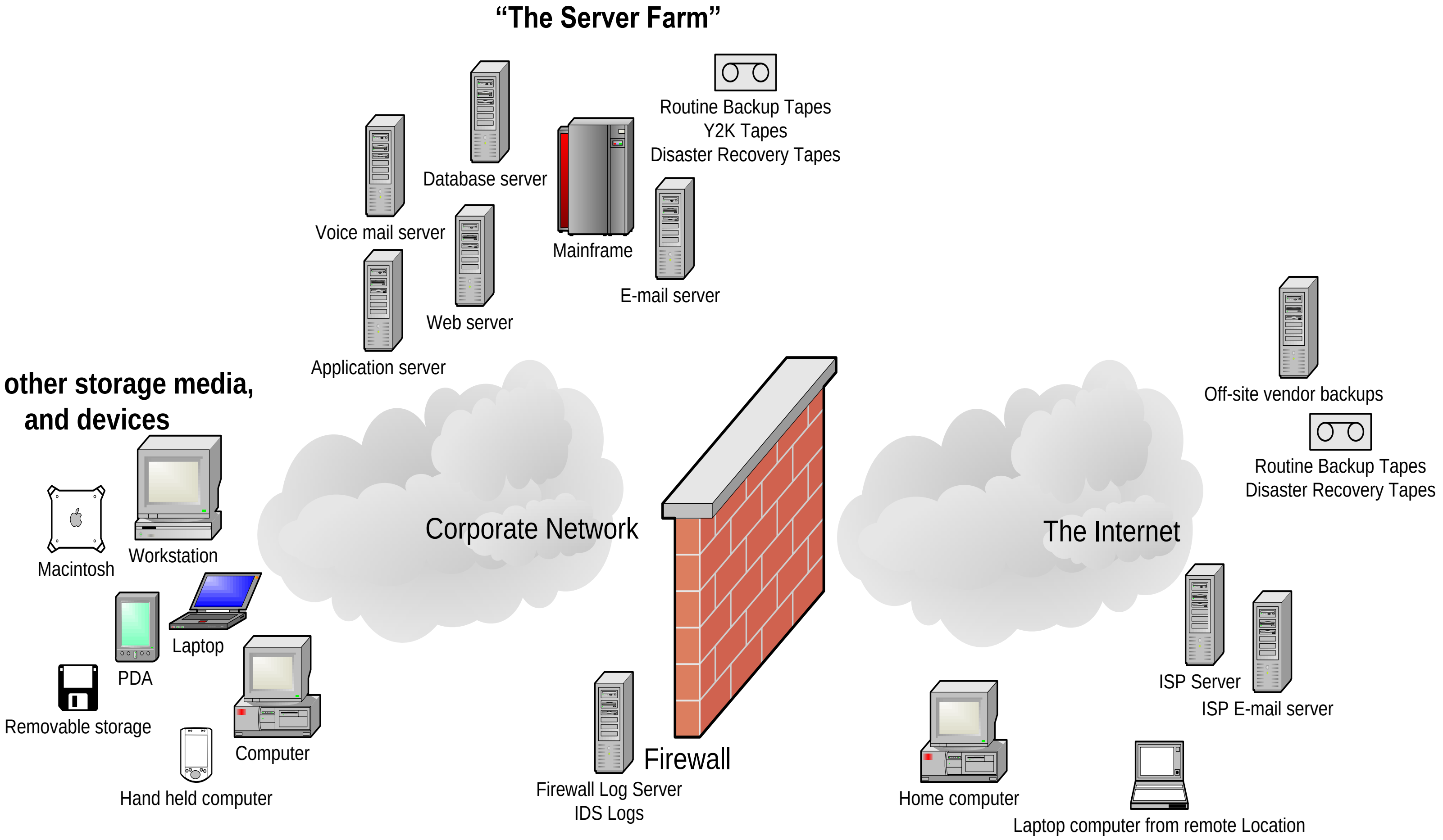
- **Identification is the first reactive step in response to an E-Discovery request.**
- **Identification involves:**
  - Localisation of potential sources of electronic information.
  - Determine the scope of the investigation
    - Which data (i.e. projects, employees, departments)
    - Which periods
- **Forensic Technology:**
  - Mapping the information landscape
  - Identifying relevant sources


Why responding to ediscovery requests can feel like herding goats
TEXT MESSAGES
SUBPOENAS
TWEETS
LAWSUITS
SERVERS
EMAILS
LAPTOPS
PAPER RECORDS
ASPIRIN & PROZAC
DESIGN CONCEPT BY REX DESILVA, REXDESILVA@GMAIL.COM

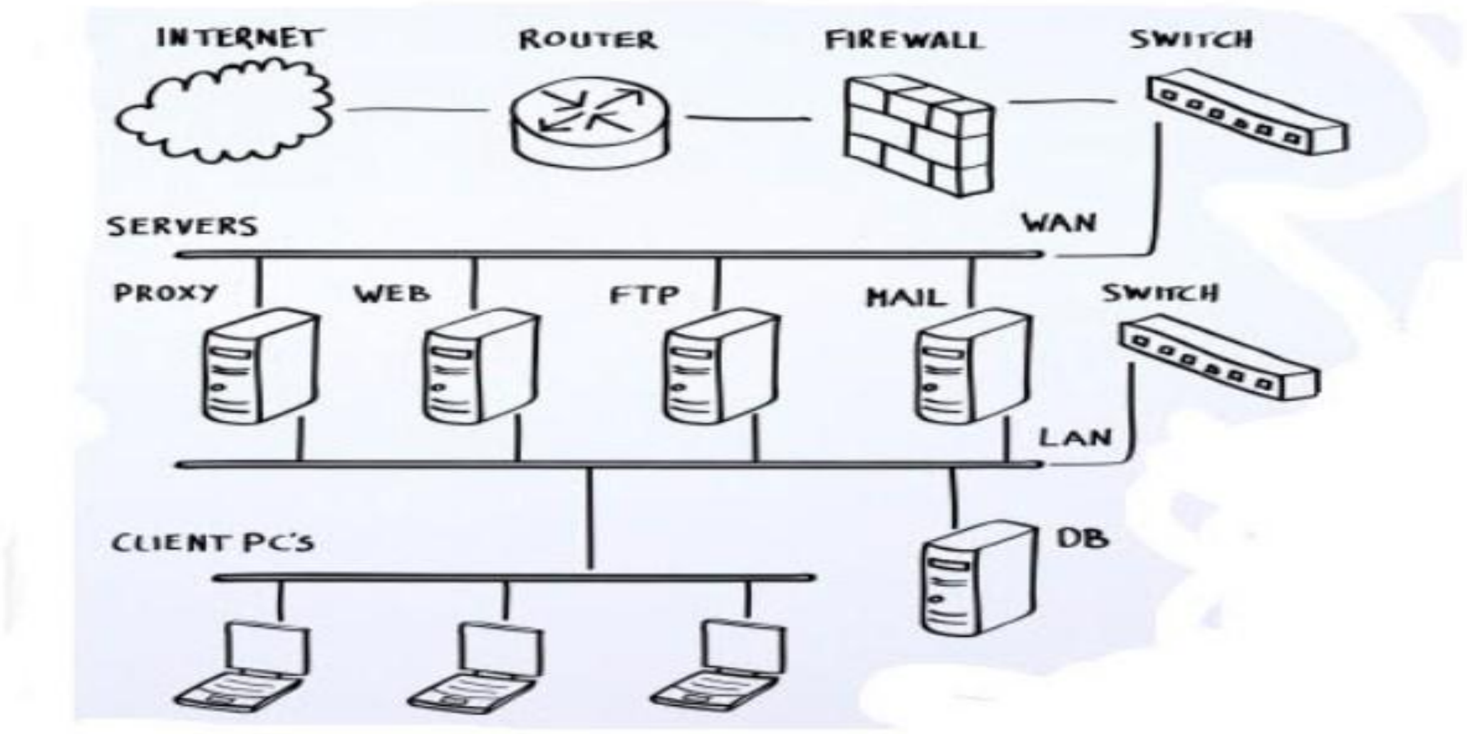
IT Infrastructure: Example 1



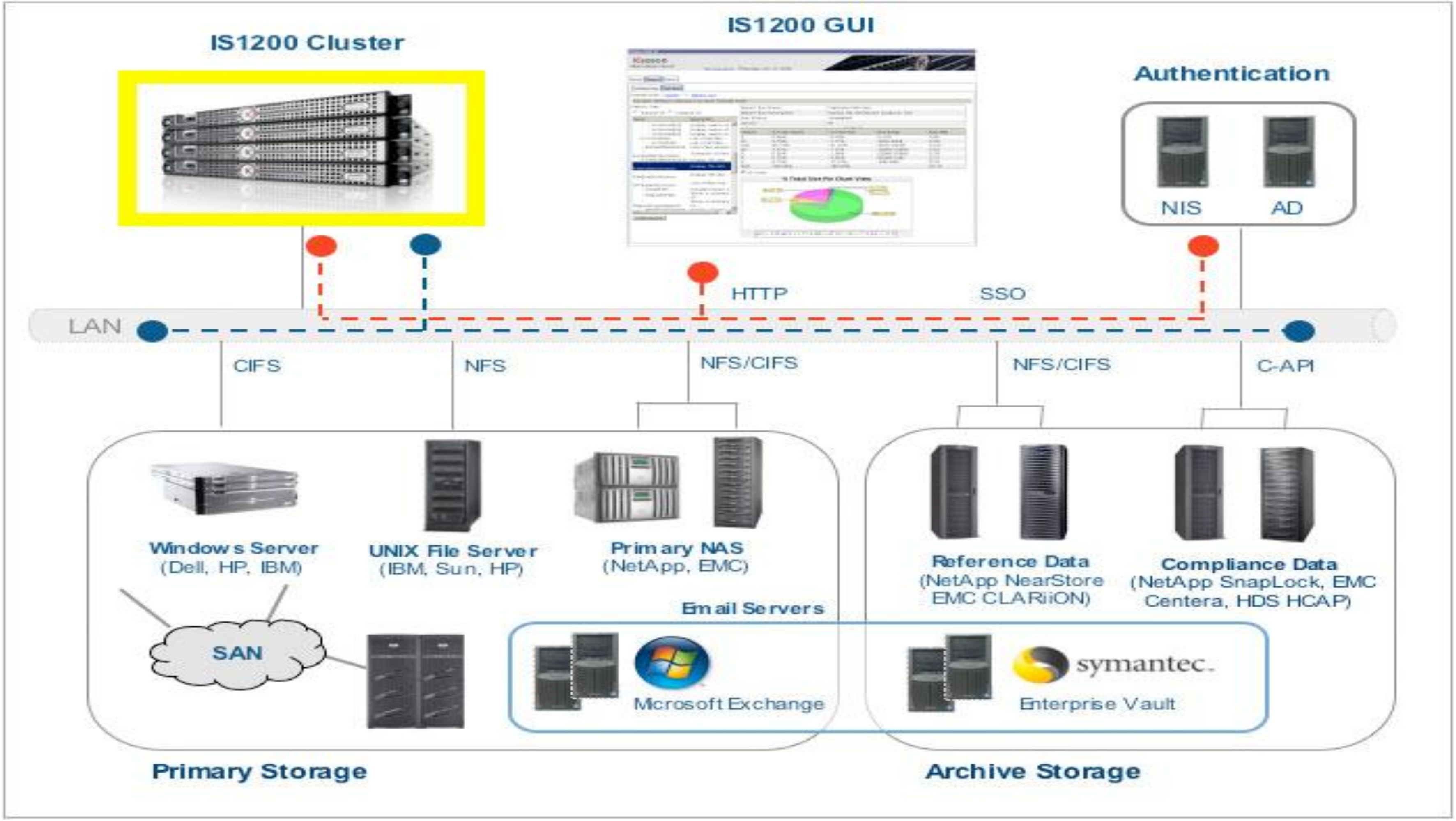
IT Infrastructure: Example 2



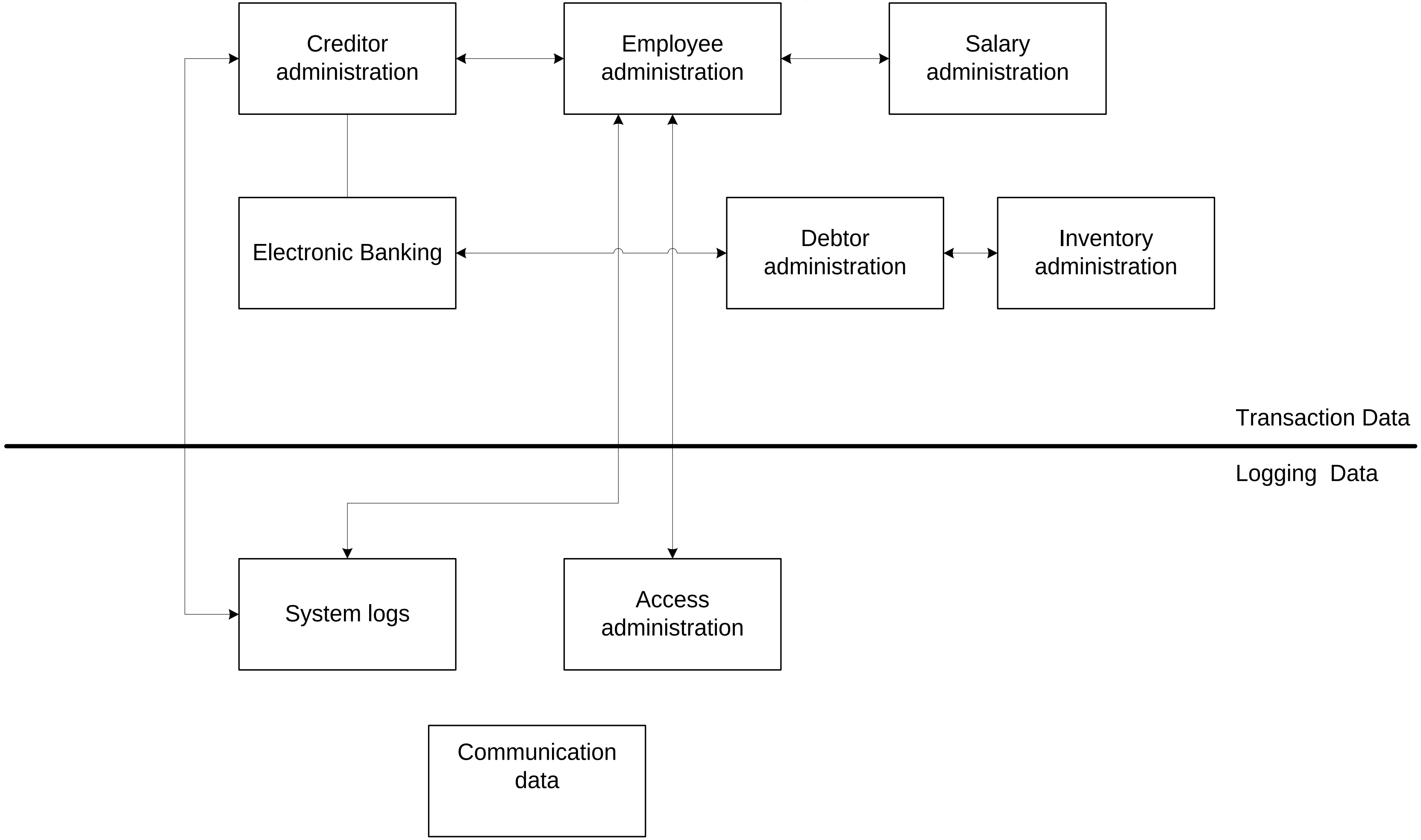
IT Infrastructure: Example 3



IT Infrastructure: Example 4



Systems: Accounting



Identifications of backups

Typical company (1800 employees) had the following backups available in July 2007:

-12x Backup July 2006 /June 2007

-1x Backup Friday 29/12/2006

-1x Backup Friday 30/12/2005

-1x Backup Friday 31/12/2004

Total 15 backups per custodian!

Data preservation

- **Goal:**
  - Preserve data to avoid spoliation claims/sanction
- **Measures:**
  - Issue a legal hold by sending out an internal company memo
  - Secure data to prevent it from being changed or destroyed (avoid data spoliation), for instance stop backup tapes from being recycled
  - Freeze records so they can not be destroyed

Collection

- **Relevant electronicalle stored information is copied in a forensically sound way.**
- **Forensic technology:**
  - Maintain original meta data of electronic information (i.e. filename, path, dates etc)
  - Forensic computer image versus logical file copy
  - Maintaining chain of custody
  - Calculate secure hash values of collected data

Collection: File Servers

- What to expect:
  - Files
  - Personal email archives (pst, nsf etc.)
  - Long and deep file paths
- Forensic tools:
  - Encase (Guidance Software)
  - Forensic Toolkit - FTK (AccessData)
  - Evidence Mover (Micro Forensics)
  - Robocopy (Microsoft)

Collection: Mobile Phones

- What to expect:
  - Mobile/Smart phones
  - Android Tablets, iPad
- Forensic Tools:
  - XRY (MicroSystemation) →

  - Device Seizure (Paraben)
  - UFED (Cellebrite)
  - FTK Mobile Phone Examiner (AccessData)
  - Encase Smartphone Examiner (Guidance Software)

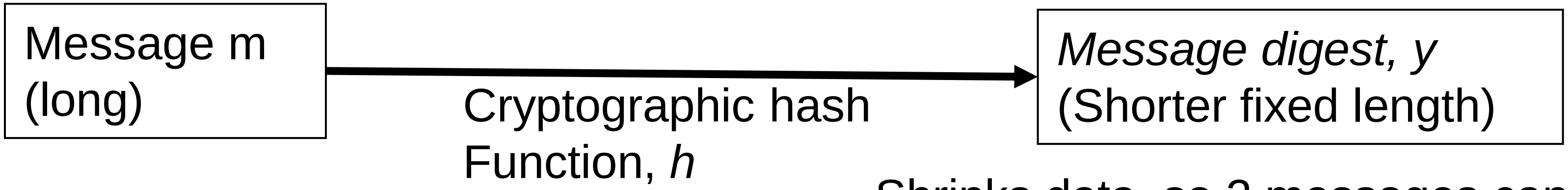
Collection: Databases

- What to expect:
  - Financial databases (SAP, Oracle Financials etc)
  - Firewall databases
  - SQL databases (MsSQL, Oracle, MySQL, Progress etc)
- Best practices
  - Use SQL queries
  - Exports vs. Dumps
  - SAP abap scripts vs. Oracle database dumps
    - (depends on size and available time)

Collection: Email Servers

- What to expect:
  - Lotus Notes (nsf)
  - Microsoft Exchange (edb)
  - Groupware
- Connect to life server (why?)
  - Exchange Server (2010 has interesting E-Discovery capabilities)
  - Encase Enterprise
- Process message store
  - Network Email Examiner (Paraben),
  - PowerControls (Kroll Ontrack)

Secure Hash: MD5 and SHA1



Message m
(long)

Cryptographic hash
Function, *h*

Message digest, y
(Shorter fixed length)

Shrinks data, so 2 messages can have the same digest: m_1 != m_2, but $h(m_1) = h(m_2)$

- **Goal: to provide a unique “fingerprint” of the message.**
- **How? Must demonstrate 3 properties:**

1. Fast to compute y from m.
2. One-way: given y = h(m), can’t find *any* m’ satisfying h(m’) = y easily.
3. Secure Hash: Strongly collision-free, i.e. can’t find any m_1 != m_2 such that $h(m_1)=h(m_2)$ easily

Procedures, Forms and Logs

1. Data freeze directive
2. Data request
3. Letter of consent
4. IT inventory template
5. Encase acquisition form
6. Chain of custody form
7. Evidence log for tracking collected electronic data
8. Physical document collection sheets and scanning log
9. Standard Operation Procedure for Data Collection

Information Technology & Computer Science
E-Discovery Lab

Hogeschool van Amsterdam
Media, Creatie en Informatie