Preservation Issues

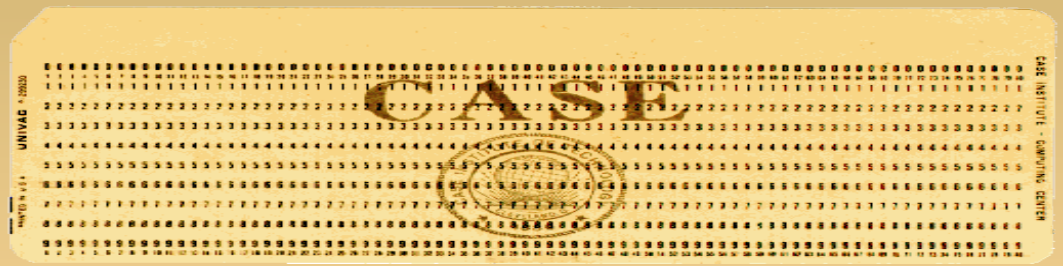
- Physical vs. Digital
  - Example
  - Viewing problem
- Problems
- Reliability of long-term digital storage
  - Visible faults
  - Latent faults

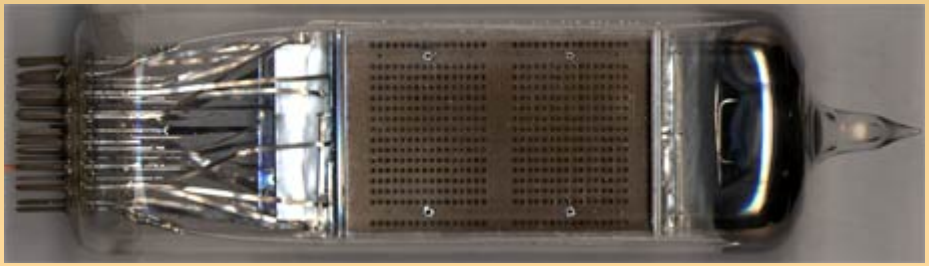
Physical vs. digital

Physical vs. digital

- Traditional Media
  - Preserving physical artifacts
  - Future researchers work with originals
- Digital Media
  - Information disembodied from physical artifact
  - Analog to digital copy
  - Digital to digital copy

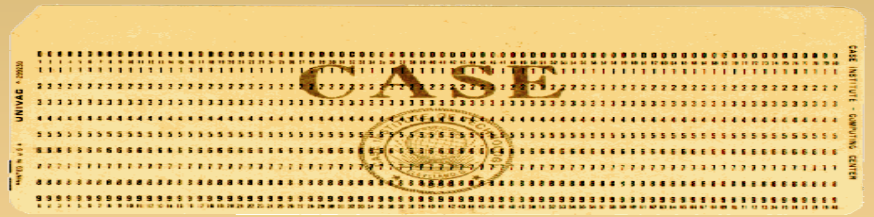
Digital storage

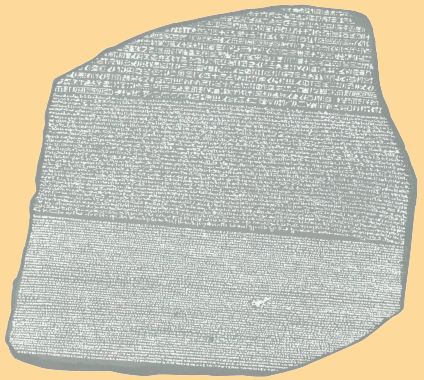




Perspective

- Punch Card
  - 100 years ago
- Floppy Disk
  - 40 years ago
- Rosette Stone
  - 196 BC.
  - ca. 2200 years ago





Change

- Viewing
  - Interpretation
  - Everything encoded as 0,1
- Infrastructure
  - Reading devices
  - Display devices
  - Scanning devices

Preservation Issues

- Physical vs. Digital
  - Example
  - Viewing problem
- **Problems**
- Reliability of long-term digital storage
  - Visible faults
  - Latent faults

Problems

- Viewing
- **Scrambling**
- Inter-Relation
- Translation
- Custodial

Scrambling

- Compression
  - Lossless
  - Lossy
- Encryption
  - Digital Right Management (DRM)
  - Algorithm
  - Keys
- → Added complexity
- Data appears to be random

Inter-Relation

- Information is increasingly inter-related
- Rosetta Stone
  - Everything in one place
- Web
  - Everything is referenced
  - Best practice: many files
  - How many references should be saved?

Translation

- Analog → Digital
  - Information is lost
  - Sampling
  - Information we do not know about
- Digital → Digital
  - Lossless only in theory
  - Metadata needs to be converted
  - File ← → Application
  - No historic data

Translation - Vocabulary

- Refreshing
  - Moving files from on media to another
  - Punch Tape → Floppy → HDD
- Migration
  - Digital to digital translation
- Emulation
  - Focus on the application environment
  - High complexity

Custodial

- Responsibility
- Not anything can be archived
- Authenticity needs to be guaranteed
  - Information is changed in the translation process
- Guidelines
  - Importance might not be obvious

Preservation Issues

- Physical vs. Digital
  - Example
  - Viewing problem
- Problems
- **Reliability of long-term digital storage**
  - Visible faults
  - Latent faults

Reliability of long term storage

- Concrete Model
- Threads to preservation
  - Large-scale disaster
  - Human error
  - Component faults
  - Hardware obsolescence
  - Software obsolescence
  - Economic faults
  - Media faults

Visible faults

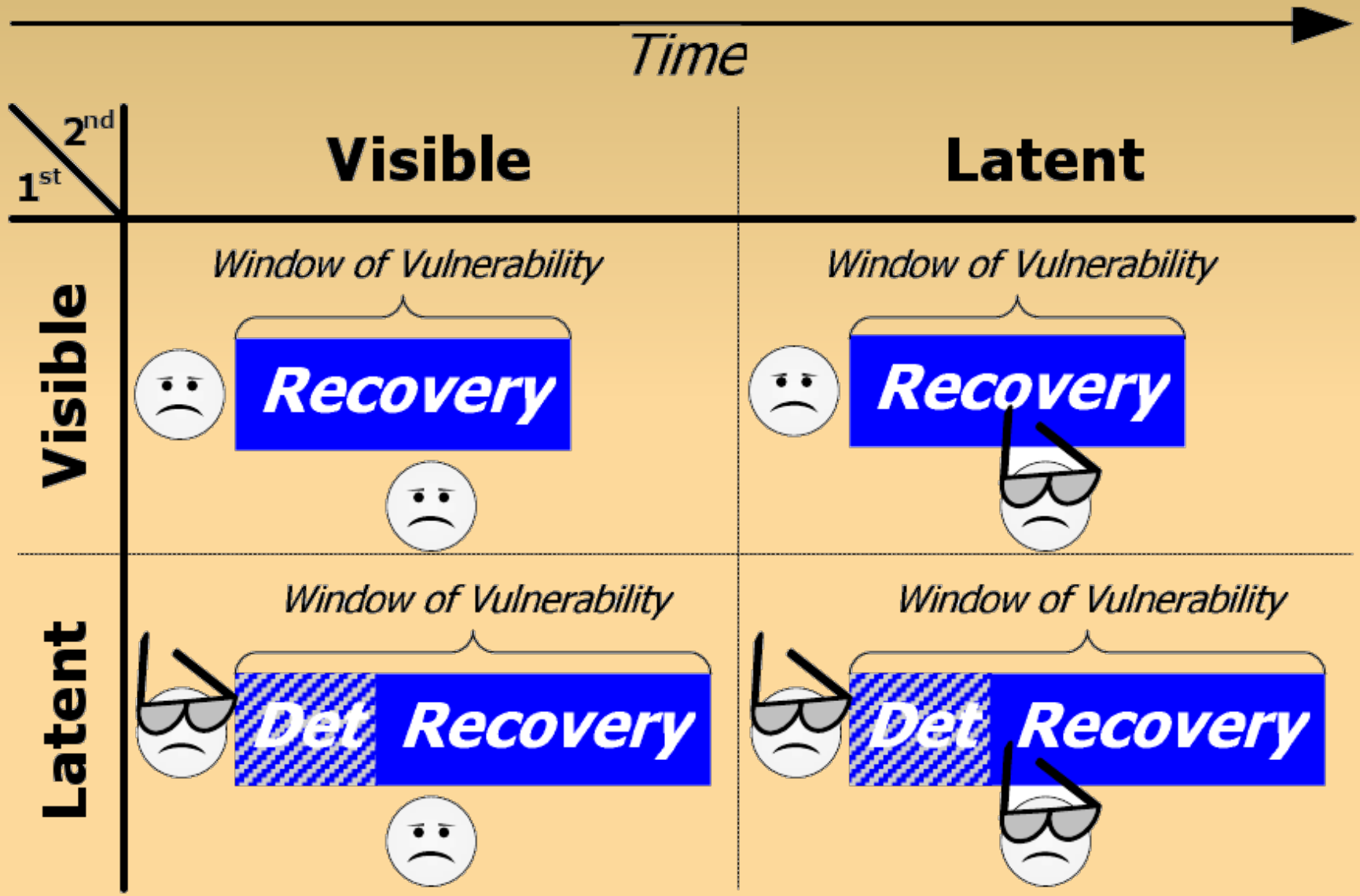
- Example:
  - HDD fault
  - Controller failure
  - Theft

- The obvious solutions are replicas
  - Different physical location
  - Distributed administrative controll

Latent Faults

- Example
  - Censorship
  - Bit rot
- Checksums
- Periodic checks
  - Checks might degrade quality
  - Restore from replica

Overwiev


Time
1st
2nd
Visible
Latent
Visible
Latent
Window of Vulnerability
Recovery
Window of Vulnerability
Recovery
Window of Vulnerability
Det
Recovery
Window of Vulnerability
Det
Recovery

Summary

- Differences between digital and analog
  - Encoding
  - Infrastructure
- New procedures
  - Selecting important information
- Information is lost easier
  - Data is lost if nothing is done to prevent the loss
  - Analog data behaves the other way around