Digital Preservation Issues

Antony Gordon

British Library National Sound Archive

IAML Berkeley, 5th August 2002

Summary

- Not about digitization for preservation
- Preservation issues in existing media
- Carrier formats – change is continual
- What is digital? Why is context important?
- Obsolescence problems
- Digital media degradation
- Digital preservation and digital stores

Digital is forever?

This is a dangerous illusion!

Analogue audio carriers

- cylinders (mostly wax) – *fragile*
- coarsegroove discs (mostly shellac)
- Metal/glass platters coated with a thin layer of lacquer / cellulose nitrate – *very fragile*
- microgroove discs (vinyl)
- steel wire
- flexible ferromagnetically-coated plastic strips (tape)

Digital audio carriers

- Digital Audio Tape (R-DAT)
- compact disc (CD)
- “enhanced” CD
- betamax/VHS tape cassettes (PCM-encoded)
- minidisc
- digital compact cassette (DCC)
- DVD (digital versatile disc)
- Where next? … 100Gb and upwards …

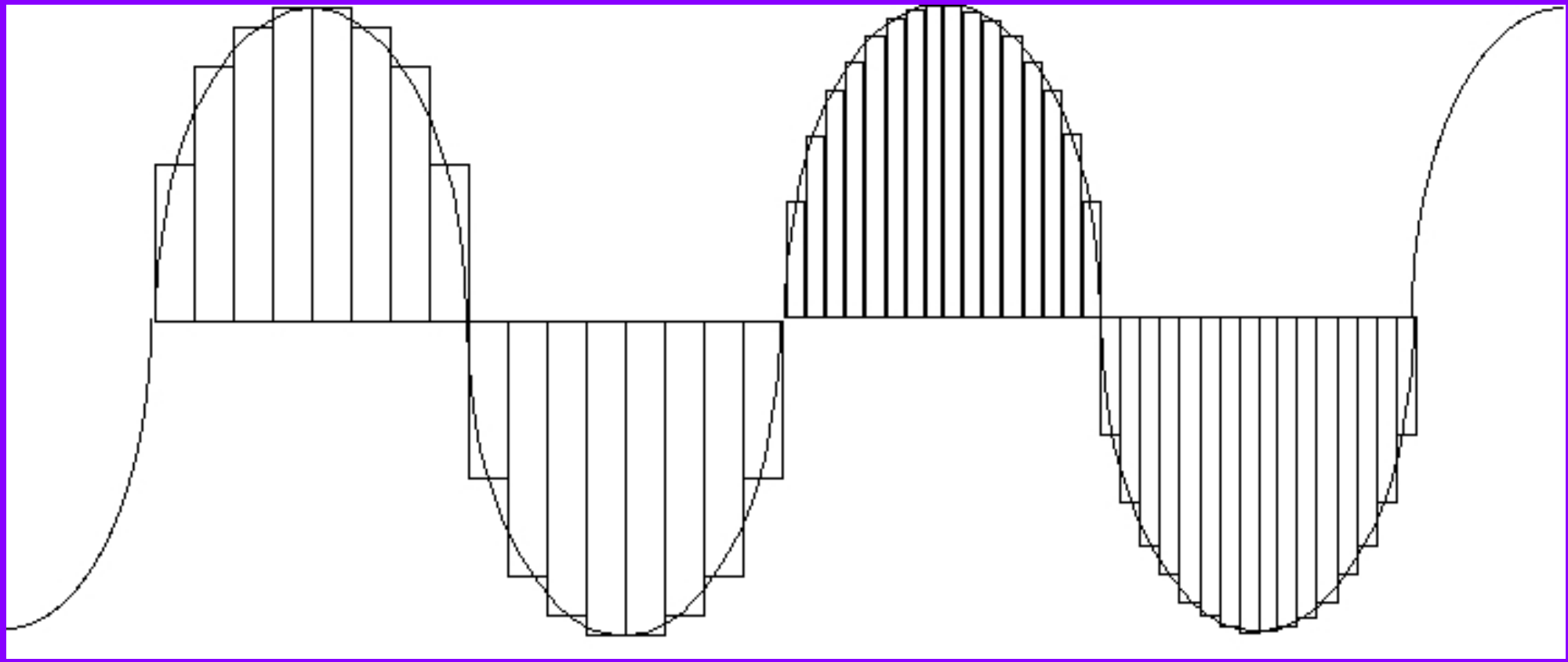
Signal compression

- Lossy compression used for minidisc, DCC, DVD, mp3 – original can never be re-created
- Only lossless compression – or no compression at all – should ever be used for archiving of digital originals

What do we actually mean by digital anyway?

- *Definition*
  - The representation of complex (analogue) data (sound, graphical, etc.) by encoding in any two-state (binary) system.

Sampling a waveform



Case example

“Red book” compact disc (16 bit resolution at 44.1kHz)

- this registers 65,536 different levels, generating data at 705,600 bits per second per channel = a little under 606 Mb per hour for stereo
- DVD Audio samples at 24 bit resolution and a rate of 192kHz (around 4Gb per hour for stereo)

Data packets, redundancy and error checking

- Grouping of bits into: bytes (8 bits) or words (16, 20, 24, 32, 64 or 128 bits)

0001**0**1001001010000010100110101001100**1**1101

IAML

0001**1**1001001010000010100110101001101**0**1101

§†¶¶

- Error checking and data redundancy

The issue of context

11010011001010100001101010111100011010011001010100001101010111100011010011001010100001101010111100011010011001010100001101010111100011010011001010100001101010111100011010011001010100001101010111100011

- **text file (e.g. ASCII, RTF, Word…)**
- **graphic image (e.g. BMP, GIF, TIFF, PNG)**
- **sound file (e.g. WAV, RDF, AIFF)**
- **moving image file (e.g. QTM, MPG)**

Hardware obsolescence

- BBC microcomputer
- Acorn Archimedes
- Amiga
- NeXT box
- IBM-compatible PC (from 8086 to Pentium IV)
- Apple machines (from 68000 to G4)
- *Case study:* BBC Domesday Project

Digital media formats

- 8" (200mm) diskette, or floppy disk
- 5.25" (130mm) floppy disc
- 3.5" (90mm) floppy disk
- Zip disks, Jaz disks, portable hard drives (SCSI /Firewire/USB 2)
- Tape formats: DAT or DLT cartridges

Software obsolescence

- IBM-compatible PCs
  - DOS
  - Windows 3.1
  - Windows 95 and 98, NT, 2000, Me and now XP
- Apple Macs
  - System 6
  - OS 7, 8 and 9
  - OS X
- Connector changes can have an impact !

Degradation of storage media

- Magnetic media
  - Drop outs
  - Substrate problems
  - Binder problems
  - Tape edits
- Optical media
  - Scratches to the playing surface
  - Oxidation of the reflectant layer

Born-digital materials

- With digitized analogue materials we still have the original
- With born-digital materials we are conserving the original, or a digital clone

Digital preservation

… the series of managed activities necessary to ensure continued access to digital materials for as long as necessary … beyond the limits of media failure and technological change.

Digital Preservation Coalition handbook (Jones and Beagrie)

Preservation strategies

- Technology preservation
- Data refreshing
- Data migration
- Emulation
- Virtual envelopes
- Data archaeology
- Use of open standards

Preservation strategies II

- Print it out !
- Save in a simple common format
- Not a viable option for audio
- HD-ROM ??

Digital stores

- Metadata
  - traditional catalogue data
  - file format
  - hardware and software required for replay
  - origination/copying dates
- Data pertaining to media and file formats

Trusted Digital Repositories

Defined as:

"… one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future."

(RLG-OCLC report entitled 'Trusted digital repositories: attributes and responsibilities', May 2002 .)

Conclusion

"It is an exercise in futility to try to pick the technological winner, for 'the latest' will be hopelessly outdated in ten years' time. What is necessary is a permanent strategy for handling perpetual change."

The archive as an ecosystem / Julia Martin and David Coleman (Journal of Electronic Publishing, April 2002)

Links

- **Open Archival Information System**

  www.rlg.org/longterm/oais.html
- **Preserving Access to Digital Information**

  www.nla.gov.au/padi/
- **Digital Preservation Coalition handbook**

  www.jisc.ac.uk/dner/preservation/workbook/
- **BBC Domesday Project**

  www. atsf.co.uk/dottext/domesday.html