# Digital Curation: Archives in Theory and Practice

Kevin Ashley

ULCC/NDAD
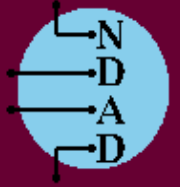
http://ndad.ulcc.ac.uk/

# Why am I here
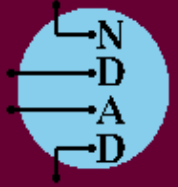
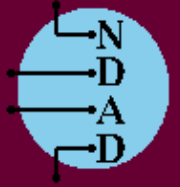- NDAD - National Digital Archive of Datasets
- Structured, digital records from UK government
- Operated under contract to PRO
- Acquires, documents, preserves, secures, catalogues, makes available for research…
- 300 Terabyte capacity (at present)
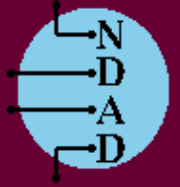- Also NDR™ - general preservation

# Why do archives exist ?

◆ Legal protection

◆ Creating historical collections

◆ Enabling reuse

◆ Informing the corporate memory

◆ Marketing

# ULCC: The back story

- 30 years of dealing with mass storage
- Scientific projects outlive computer systems
- Migration a way of life
- IEEE Mass Storage Systems Reference Model
- European Mass Storage Specifications Working Group
- Is OAIS to be our saviour?

# Archives: the back story

- ◆ Deal with traditional materials
- ◆ Collection driven by parent institution or subject policy
- ◆ Ingest organised and disorganised material
- ◆ High-level standards (ISAD(G) et al)
- ◆ Context and provenance key
- ◆ Sensitivity issues in access

# What forms do they take ?
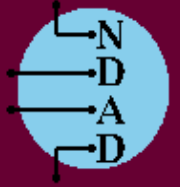
- Documents - 'digital paper' (including email)
- Spreadsheets
- Databases
- Digital audio/video/images
- Exotic forms: virtual worlds, games, etc.
- Programs
- Assemblies of the above: web sites, etc

# Characteristics
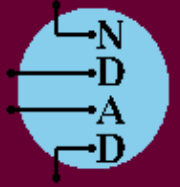
- Two forms: dynamic and static
- Static records are created once and not altered
- Dynamic records continually change
- Most common dynamic records are databases
- Static records must be captured to prevent change

# Metadata
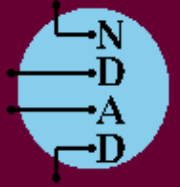
- Data about data

- It isn't specific to digital records

- Deals with:
  - resource discovery: 'Is there a document about X?'
  - resource description: 'What is this document?'

- May be embedded (TIFF) or external (catalogues)

- Most records contain some embedded data

# Metadata examples
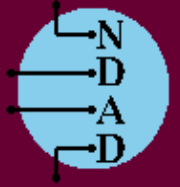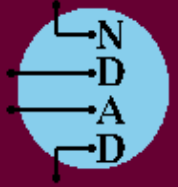
- Author
- Sensitivity/access conditions
- Retention period
- Subject
- Date of creation/use/retirement
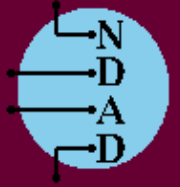- Keywords
- Abstract

# Non-digital metadata
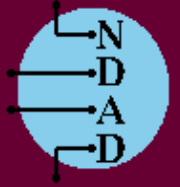
- Most computer systems need paper to be understood

- Manuals, specifications, reports

- Some essential information may only be in people's heads

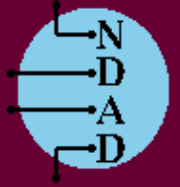- Most important when dealing with older records

# Confessions of a contractor

- Mix of professions essential
- Early focus on archival standards
- Post-contract standards difficult for both sides
- Clean exit strategies essential
- In-contract monitoring - what of ?
- Third parties (e.g. government) have different agenda

# OAIS - the complete answer ?

- Eases procurement of hardware and software
- Eases interworking with compliant systems
- Eases migration tasks
- but…
- Interworking with traditional repositories ?
- Dealing with mixed-mode records

# What is being preserved ?

◆ Existence may be more important than content

◆ Don't worry about emulating original views

◆ Do worry about describing original views and constraints

◆ Understanding is mediated by presentation

◆ Do allow for future exploitation

◆ 'Recordness' is critical