Ingesting, Preserving and providing access to databases as records: experience at NDAD and elsewhere

Kevin Ashley
Head of Digital Archives Dept
ULCC
Overview

• What I won’t cover
• How NDAD works:
  ⇒ Pre-ingest activities
  ⇒ Ingest process
  ⇒ Preservation
  ⇒ User access
• Contrast with other database archives
• Challenges and future directions
What I won’t cover

- Costs (see DLM 1999 proceedings)
- Description
- Selection
- Details of access
- Our contract with The National Archives
- Details of alternative models:
  ⇒ See ERPAnet Bern workshop proceedings
  ⇒ or my article in Vine (2003)
What is NDAD?

- A service for UK government records which exist as ‘structured information’
- Contains data + contextual information
- Established in 1997 - service in March 1998
- First service by a national archive to provide online public access to preserved material
- Selection undertaken by National Archives and government departments
- Everything else at ULCC: under contract to TNA
Our role

- Acquire
- Preserve
- Describe
- Provide access + copies
- Provide support to users
- Promote
gov.uk records lifecycle

- Policies: Acquisition, Disposal, OSP
- Guidelines: Retention, Appraisal
- Schedules: Retention, Appraisal

Process:
- Creation
- 5-year review
- 25-year review

Places:
- External storage
- Disposed
- PRO
- Destroyed

Department
Resultant selection decisions

- Digital records continuing from preserved paper records
- Large, easily identifiable systems
- Systems on brink of retirement or replacement
- Many digital records still invisible to manual selection process
- Many actors involved are new to process of records management
NDAD’s relationships

- Users
- NDAD
- TNA
- Government Contractors
- Government Departments
Accession process

• Information collected from records manager and data owners: format, titles, documentation, etc
• Digital and paper material acquired and listed
• Check for:
  ⇒ completeness
  ⇒ accuracy
  ⇒ readability
• May lead to further acquisition or negotiation
Preservation

- Data transformed to canonical form - originals kept
- Paper documentation digitised
- Technical metadata produced or transformed
- Consistency checks applied:
  - For transformation process
  - Against original system
  - Against published information
  - Internal cross-checks
Description

- Professional archivists work with data specialists
- Goal is to document the system, its use and its context
- Information is historical and technical
- For some users, the catalogues are all that is needed
- For others, catalogues can assist in data reuse
Access

• Access is via website and linked to hierarchical catalogues
• Not all data open to access
• Simple viewing of rows from one or more tables:
  ⇒ Ability to choose fields
  ⇒ Ability to use query language or forms to select rows
  ⇒ Refine query or fields selected at any point
• Links back to catalogues + documentation
• Ability to order copies of data and/or documentation or download data
Access: restricted viewing

Not available until 2050

<table>
<thead>
<tr>
<th>ID</th>
<th>REG</th>
<th>Date</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>10246</td>
<td>X111ABC</td>
<td>1 Oct 98</td>
<td>Clapham</td>
<td>Deptford</td>
</tr>
<tr>
<td>56999</td>
<td>A217HGB</td>
<td>23 Dec 97</td>
<td>Poole</td>
<td>Sandy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REG</th>
<th>Make</th>
<th>Year</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>X111ABC</td>
<td>Yugo</td>
<td>1999</td>
<td>Grey</td>
</tr>
<tr>
<td>A217HGB</td>
<td>Trabant</td>
<td>1957</td>
<td>Brown</td>
</tr>
</tbody>
</table>
Access - goal

- Duplicate original system ✗
- Advanced analysis tools ✗
- Simple viewing via a generic tool ✓
- Multimedia datatypes ✓
- Extensible via object-like design ✓

- Traditional database systems not up to task without significant additional effort
- Hence much software home-grown
Inconsistencies and errors

- Schools census - 4 datasets per year for different school types
- But 1976 only has 3 - no nursery schools
- Further examination shows files have been merged
- Confirmation came from completed census forms held by schools - not by government department
NDAD: a final word

- NDAD was a quick solution to an urgent problem
- Keeps intellectual control with PRO: leaves details to service provider
- Provided a base for training archivists and information technologists
- Positive experience for all partners
A Database Taxonomy

Open:
- data still being added or amended

Closed:
- data is complete; no more changes

Dynamic:
- data altered in place

Static:
- data only added, never altered

From Shepherd, E. & Smith, C.
Cornell’s DP model

digital preservation

technology
organization
resources
Challenges

• Organisational: dealing with contractors
• Technical: reflecting dynamic systems well
• Resource: coping with many accessions in a cost-efficient way
Views of object collections

OLD

NEW