Database Archiving in Sweden

Magnus Geber, National Archives of Sweden
Background

• E-archive transfers since 70ies

• Until last decade vast majority “databases”

• “Databases” two main types
  – “Registers”, data mainly without system, statistical data
  – Business systems based on a databases

• General preservation strategy
  – Normalization & Migration, simple standard formats, relation database structure
  – fully enforced since mid 80ies, in many ways still relevant

• Information / Record concept “Allmän handling”
Normalization / Transfer demands

- Flat files, fixed record length & field(column) length
- No variable length
- No different record types
- No repeating groups
- No packed numerical formats
- No control characters
- “Standard” character sets (Ebcdic, standard Ascii)
- Relation database structure (one file – one table)
- *Label/TAR/CD demands (format on transfer data tape etc, not anymore)*
- *Later delimited option*

Restructured and converted at export from original DB systems
Database E-archiving activities

- **Metadata**: paper documents, record description
  - technical, structure, semantic, context
  - a process of getting metadata more and more in digital form

- **Preservation**
  - Physical copying and format conversion when needed (mainly Ebcdic → ISO8859-1)
  - (Conversion of backlog from old transfers in obsolete formats, Project47)

- **Access**
  - Mainly copies of whole databases or selected (classified data) databases to researchers
  - Different tools
  - Sometimes created access system of transfer, “private boats”, RMS

- **Statistics**
  - Digital database holdings < 2 TB > 100 000 filer
  - Transfers/year ≈ 25
  - Access ≈ 15 (databases) ≈ 25 (info item)
Preservation object

• **Information in data files**
  – “Library information”
  – MRD

• **“Allmänna handlingar” (Public records)**
  – In Swedish Constitution
  – Recording which can be view or perceived
  – Screen presentations or out prints
  – Database view
Preservation object
ICA Reports RM & Business systems part 3

Figure 2: Identification of information components/data elements comprising an electronic record in a database

Note that it is possible for a single record to include multiple elements from a single database field or table, and that it is also possible for a single data element to form part of more than one record.
Preservation object
Documentation by Screen Layout & SQL

RECORD DESCRIPTION

ARENDE

<table>
<thead>
<tr>
<th>Namn</th>
<th>Typ</th>
<th>Start</th>
<th>Slut</th>
<th>Förklaring</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNRI</td>
<td>A</td>
<td>1</td>
<td>8</td>
<td>Diarienummer</td>
</tr>
<tr>
<td>REGDAT</td>
<td>A</td>
<td>9</td>
<td>14</td>
<td>Registreringsdatum</td>
</tr>
<tr>
<td>DOSS</td>
<td>A</td>
<td>15</td>
<td>18</td>
<td>Dossiebeteckning</td>
</tr>
<tr>
<td>FRAN</td>
<td>A</td>
<td>19</td>
<td>19</td>
<td>Avsändare/Mottagare</td>
</tr>
<tr>
<td>ATEXT</td>
<td>A</td>
<td>20</td>
<td>40</td>
<td>Anmärkningstext</td>
</tr>
</tbody>
</table>

KAL1

<table>
<thead>
<tr>
<th>Namn</th>
<th>Typ</th>
<th>Start</th>
<th>Slut</th>
<th>Förklaring</th>
</tr>
</thead>
<tbody>
<tr>
<td>KALROW1</td>
<td>N</td>
<td>1</td>
<td>2</td>
<td>Radnummer</td>
</tr>
<tr>
<td>AVSMOT</td>
<td>A</td>
<td>3</td>
<td>3</td>
<td>Huruvida Avsändare/Mottagare</td>
</tr>
<tr>
<td>ANMKAL</td>
<td>A</td>
<td>4</td>
<td>24</td>
<td>Händelse</td>
</tr>
</tbody>
</table>

LOGIC OF VIEW BY SQL

Vad gör utsökningen:
Listar information om ett visst ärende och dess händelse/åtgärd.
Ändra KAL1.DNRKAL1 till aktuellt diarienummer

SQL-fråga:

```sql
SELECT KAL2.DNRKAL2, KAL2.KALROW2, KAL2.BNR, KAL2.AVSMOT,
KAL1.DATUMUT, KAL1.DATUMIN, KAL2.KALSIGN, KAL1.ANMKAL,
KAL2.GODKDAT, KAL2.REVDAT, KAL2.REV, KAL2.UPDATKAL
FROM KAL1, KAL2
WHERE KAL1.DNRKAL1 = KAL2.DNRKAL2 AND
KAL1.KALROW1 = KAL2.KALROW2 AND
KAL1.DNRKAL1 = '2002-0045' AND
KAL2.KALROW2 = '000003'
```

LINKING SCREEN LAYOUT HEADINGS TO FILEDS/COLUMNS I RECORD DESCRIPTION

Huvud

<table>
<thead>
<tr>
<th>Namn</th>
<th>Fieldznamn</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAREINR</td>
<td>ARENDE.DNR1</td>
</tr>
<tr>
<td>DATUM</td>
<td>ARENDE.REGDAT</td>
</tr>
<tr>
<td>DOSSIE</td>
<td>ARENDE.DOSS</td>
</tr>
<tr>
<td>AVSÅNDRARE/MOTTAGARE</td>
<td>ARENDE.FRAN</td>
</tr>
<tr>
<td>ÅRENEMENING</td>
<td>ARENDE.ATEXT</td>
</tr>
</tbody>
</table>

Händelser

<table>
<thead>
<tr>
<th>Namn</th>
<th>Fieldznamn</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD</td>
<td>KAL1.KALROW1</td>
</tr>
<tr>
<td>AVSÅNDRARE/MOTTAGARE</td>
<td>KAL1.AVSMOT</td>
</tr>
<tr>
<td>HÄNDELSE</td>
<td>KAL1.ANMKAL</td>
</tr>
</tbody>
</table>

Magnus Geber

2012-02-07
XML-transfers

- Last decade
- Export option, normally hierarchical structure
- Demand on schema and style sheet
- Simpler controls
- More difficult with future statistical processing
- Can have less quality check
- Suitable if semantically and structurally standardized systems, ERMS etc
Architecture for information packages SIP/AIP
ADDML

- Archives Data Description Markup Language
- Describes Data Base Files
- Fixed or delimited
- Technical, structure & general describing metadata
- Record descriptions
- Also other data files on higher level
- Developed by National Archives of Norway
- Agreement with National Archives of Sweden
- Present version ADDML 8.2 (7.3)

www.arkivverket.no/arkivverket/Arkivbevaring/Elektronisk-arkivmateriale/Standarder/ADDML
ADDML, top levels

- addml
  - dataset
    - reference
    - flatFiles
    - dataObjects
ADDML
Additional elements & properties

• Generic elements & characteristics
• May be built as hierarchal structure
• An additional element may also have properties

Makes it possible to adjust ADDML to individual needs

May later be included in new version of the common standard
RADAR
Digital Archives system

Excel file

RALF
Tool for transfer

KRAM
Ingest

Digitala Kedjan
Ingest

ESSArch
Archival Storage

ARKIS
Data Management

KRAM
Access Dissemination
KRAM  Control framework
RALF

• Free distributed software
• Tool to be use before transfer
• Primarily for agencies submitting transfers
• Uses the excel file and data files
• Checks correctness of excel file and existence of data files
• Checks excel file against data file content
• Error report
• Creates simple SIP