Noark
The Norwegian Records Management Standard

- The format of repository transfers from Noark systems
- Experiences with Noark transfers since the early 1990's

Prague 22 April 2009
Hans Fredrik Berg
The National Archives of Norway
The Noark standard

- A standard for ERM systems in Norwegian government and municipal agencies since 1984
- A mandatory standard since 1999 (also for registries embedded in specialized business applications)
  - Approximately all stand-alone ERM systems in public agencies are Noark systems certified by The National Archives
- Version 4 (1999) and after: An ERM standard including electronic archiving of documents

• **A standard for registry systems ("journal systems")**
  – Continuous registration of reference information (metadata) on documents received and sent by agencies
  – Including references to the filing of paper documents to enable management of physical archives
  – From version 3 (1994) Noark systems might also include references to electronic work copies of paper documents

• **Two levels of standardization**
  – Functional requirements
  – Transfer format (but samples were not tested and certified by The National Archives)
Three-schema architecture

- **External schema**: user view
- **Conceptual schema**: “business” logics
- **Internal schema**: data definition
Transfers to The National Archives had to conform to this model. Each table had to contain specific fields with specific names, data types, lengths, etc. The native format of Noark systems might be different (and more complex), but data had to be converted to this structure when transferred.
The transfer format of Noark version 1-3

- Exported data extracts from 3 or 4 tables
  - 40 columns (metadata) included
  - Table content exported as flat files in fixed (length) format

- **Purpose:** To enable regeneration of technical and logical structure
  - The Noark 1-3 model did not distinguish between these two structures
  - The four tables defined were not required to be implemented as the native (file or database) structure of Noark-systems

- One common format was used for repository transfer and data exchange between Noark systems

- Repository access: Dependent on Noark 1-3 viewers (or "survived" Noark 1-3 applications)
Exchange and transfer

Agency

Noark system 1
Exchange
Noark system 2

National Archives
Repository access
Transfer
Noark version 4 (1999)

• Specification of an ERM system
  – Technical structure specified in great detail (Noark-4, part 2)
  – 95 tables with 885 columns

• Two levels of standardization
  – Functional requirements
  – Transfer format (samples are tested by The National Archives during the certification process)
    ▪ But: it is not mandatory to implement the technical structure of Noark-4 as the application’s native (internal) structure
    ▪ What is mandatory, is the capacity to output transfers according to the structure defined in Noark-4, part 2.
Typical use of Noark 1-3

Archivist

Noark system

service

Executive officer
Typical use of Noark 4

Archivist

Noark system

Executive officer

Specialized business application

Archive interface

User interface

API or WS

SQL
The transfer format of Noark-4

• Data are exported from 40–60 tables
  – Data are exported in XML format
  – A DTD is defined for each table

• The transfer format is basically the same as the format for exchange between Noark-4 applications

• Like Noark v. 1-3, the purpose is to enable regeneration of technical structure

• Repository access
  – Database applications (Modified Noark-4 systems)
  – Database report generators
  – Database querying (SQL)
Problems with Noark-4 transfers

• Data in table extracts are frequently inconsistent
  – Referential integrity rules are violated
  – Duplicated primary key values occur in tables
  – The transfer model is dependent on referential integrity

• This type of problems was first observed in Noark-3 transfers. What may cause the problems?
  – Conversions from a native database structure
  – Earlier conversions of data from a “liberal” Noark-3 system to a Noark-4 system
  – Problems in fulfilling the requirements when selecting closed files for transfer (the principle of “periodization” in Noark)
  – Complexity in general

• Conclusion: A need for simplified transfer structures
Noark version 5 (2008)

• Specification of an ERM standard
  – Logical structure specified

• Two levels of standardization
  – Functional requirements
    ▪ Very few mandatory core requirements
    ▪ Optional requirements for integration with other tools to build complete stand-alone ERMS’ (like Noark-4)
    ▪ Optional requirements for embedding in business applications
  – Transfer format (samples will be tested by The National Archives during the certification process)

• Focus on the records rather than the applications
Noark 5 data model

Archive: Possible to include several archives (created by different agencies) in one system

Archive group: Division of the archive into separate periods of records, typically 5 years each

File: Arbitrary number of levels of sub-files

Component: used to handle different versions, redactions and formats of the same document

Transfer: Each document should be transferred as one file (bit stream). The National Archives accept the following document formats:

- Text (UTF-8 or ISO 8859)
- TIFF v. 6
- XML (ODF or OOXML?)
- PDF/A
The transfer format of Noark 5 (2008)

• Transfers are radically simplified - and more robust
  – Confined to 8 archival units (metadata objects)
  – 175 metadata in one hierarchical XML file as defined in an XML Schema

• Archival information is liberated from the technical structure
  – By focusing on archival records, preservation can be confined to the logical structure of the records themselves
  – Identification of the 8 archival units is not based on the original database primary key values
  – All codes employed by the application system are substituted by the texts represented by the codes
Consequences

• **Completely different language (XML vs. tables) for transfers**
  – Makes the distinction between transfer format and database structure obvious
  – Relations are represented by position (containment in the XML hierarchy) instead of explicit links (foreign keys)
  – Object identification is implicit in the XML path instead of explicit by unique keys

• **Distinction between exchange and transfer formats**

• **Repository access**
  – by style sheets and XML tools
  – import to an XML database and use XML queries (e.g. XPath)
  – convert and import to a database system and use database tools
Correct transfers!

Thank you for your patience
### Noark v. 1-3 transfer format: The File Table

<table>
<thead>
<tr>
<th>Position</th>
<th>Field name</th>
<th>Length</th>
<th>Pic</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Posttype</td>
<td>1</td>
<td>X</td>
<td>Fast verdi S</td>
</tr>
<tr>
<td>2-4</td>
<td>Grad</td>
<td>3</td>
<td>X(3)</td>
<td></td>
</tr>
<tr>
<td>5-14</td>
<td>Saksnr.</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-22</td>
<td>Årstall</td>
<td>4</td>
<td>9(4)</td>
<td>jf. kommentar #1</td>
</tr>
<tr>
<td></td>
<td>Skråstrek</td>
<td>1</td>
<td>X</td>
<td>Fast verdi /</td>
</tr>
<tr>
<td>15-22</td>
<td>Fortløpende nr.</td>
<td>5</td>
<td>9(5)</td>
<td>Høyrejustert, nuljer foran</td>
</tr>
<tr>
<td>15-22</td>
<td>Dato</td>
<td>8</td>
<td>9(8)</td>
<td>Form: ÅÅÅÅMMDD, jf. kommentar #1</td>
</tr>
<tr>
<td>23-24</td>
<td>Arkiv: 1. delfelt</td>
<td>2</td>
<td>X(2)</td>
<td>Angir delarkv</td>
</tr>
<tr>
<td>25-40</td>
<td>Arkiv: 2. delfelt</td>
<td>16</td>
<td>X(16)</td>
<td>Arivistkode/Arivibetegnelse</td>
</tr>
<tr>
<td>41-48</td>
<td>Blankt felt</td>
<td>8</td>
<td>X(8)</td>
<td>Fast verdi: Blank</td>
</tr>
<tr>
<td>49-52</td>
<td>Saksansv.: 1. delfelt</td>
<td>4</td>
<td>X(4)</td>
<td>Angir kontor</td>
</tr>
<tr>
<td>53-56</td>
<td>Saksansv.: 2. delfelt</td>
<td>4</td>
<td>X(4)</td>
<td>Saksansvarligs initialer</td>
</tr>
<tr>
<td>57-72</td>
<td>Presedens</td>
<td>16</td>
<td>X(16)</td>
<td></td>
</tr>
<tr>
<td>73-74</td>
<td>Kassasjon</td>
<td>2</td>
<td>9(2)</td>
<td>Kassasjonskode, høyrejustert</td>
</tr>
<tr>
<td>75-90</td>
<td>Kryssreferanse</td>
<td>16</td>
<td>X(16)</td>
<td></td>
</tr>
<tr>
<td>91-160</td>
<td>Sakstittel: 1. linje</td>
<td>70</td>
<td>X(70)</td>
<td></td>
</tr>
<tr>
<td>161-230</td>
<td>Sakstittel: 2. linje</td>
<td>70</td>
<td>X(70)</td>
<td></td>
</tr>
<tr>
<td>231-233</td>
<td>Antall dok</td>
<td>3</td>
<td>9(3)</td>
<td>Høyrejustert</td>
</tr>
<tr>
<td>234-241</td>
<td>Siste dok. (datofelt)</td>
<td>8</td>
<td>9(8)</td>
<td>Form: ÅÅÅÅMMDD, jf. kommentar #1</td>
</tr>
<tr>
<td>242</td>
<td>Sak avsluttet</td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>243-258</td>
<td>Sekundærkode</td>
<td>16</td>
<td>X(16)</td>
<td>Sekundærkodeangivelse</td>
</tr>
<tr>
<td>259-260</td>
<td>Blankt felt</td>
<td>8</td>
<td>X(8)</td>
<td>Fast verdi: Blank</td>
</tr>
</tbody>
</table>
Noark-4 data model

- Party to case
- Precedent
- Cross reference
- Sender/addressee
- Registry entry
- Filing plan
- Note
- Additional information

Electronic document
Noark-4: Electronic document management (1)

Diagram:

- Registry entry
- Document link
  - Document description
    - Keyword
    - Note
    - Additional information
  - Digital signature
  - Version
    - Electronic document
    - Document storage
Noark-4: Electronic document management (2)

Files

Registry entries

Document links

Record documents

Versions, variants (e.g. redactions) and formats
Archival periods in Noark systems (1)

File 1
File 2
File 3
File 4
File 5
File 6
File 7
File 8
File 9
File 10
File 11
File 12
File 13
etc.

Time

New period

Now
Archival periods in Noark systems (2)

These files turned out to be active after period shift

2 years later

New period
Archival periods in Noark systems (3)

These files remain in the active database.

New period

2 years later
These files are defined as closed - and selected for transfer

New period

Time

2 years later

File 1
File 2
File 5
File 9
File 10
File 13
To what extent can the Noark 5 transfer model be generalized?

• What do we want to preserve?
  – Data or archival records?
  – Systems or archival records?

• Is preservation dependent on the prolonged life of (regenerated) systems?
  – For reasons of integrity or authenticity?
  – Continued “system control” may be maintained by other means, for instance by the use of checksums

• Is the Noark 5 transfer model applicable outside the domain of recordkeeping systems?
  – Urgent need for simplification in response to the increasing complexity of systems
  – Necessary to focus on the distinction between logical information content and technical structure
Example: Business system of today on the logical vs. the technical implementation level

The (logical) presentation level selects data from 3 different databases.

What is the “system”? How do we define preservation?