Digital archiving in Denmark – strategy and implementation

Jan Dalsten Sørensen, Rigsarkivet
Reykjavík
26-05-2016
Agenda

- Why archive digitally?
- The OAIS-model
- Strategy for digital archiving in Denmark
  - Vision
  - Notification and approval
  - Frequent submissions
  - Access
  - Preservation planning
- Questions
Why archive digitally?

• The amount of information is growing rapidly – in digital form you can preserve more
• Both Iceland and Denmark are advanced countries in terms of e-Government – digital government creates digital archives!
• You may be able to print documents, but it makes no sense to print metadata, content from business systems etc.
• Retrieval, access, reuse of digitally created content must be handled digitally – otherwise you lose both authenticity and functionality.
The OAIS-model
Vision

• To ensure that digital records are preserved so as to maintain their authenticity and so that they can be found and reused
Born-digital records, challenges

• Logical preservation
• Physical preservation

• Possible strategies for logical preservation:
  – Emulation
  – Migration
Implementation

• Early identification and approval of it-systems
• Frequent submissions in archival formats
• Preservation planning and migration as necessary
Identification & appraisal/approval

• National authorities must notify the National Archives on all it-systems
• Appraisal at an early stage
• Systems with digital content that is to be preserved must go through a process of approval

Illustration: digitalbevaring.dk
Requirements for approval

- Data structure: relational database – or the potential to create one (All)
- Limited by organisation or subject (ERMS)
- Limited “archival period” (5 years) (ERMS)
- Instructions for the use of the system, including file formats (ERMS)
- Instructions for migration to archival format (All)
Mandatory Metadata?

- Only a few, because the agencies should use the metadata they need, not the metadata the National Archives can think of
- All metadata is transferred to the National Archives as tables in a relational structure
- Certain key metadata must be registered about files and documents
Mandatory Metadata

- Files: ID, title, classification according to file plan
- Documents: ID, title, date, sender/recipient (for letters), which file it belongs to
Evolving Concept of ”System”

Until recently: IT-system = a database and an application

Today: Service Oriented Architecture; loosely coupled systems etc.

→ what is the unit that can be handled by the archives in terms of appraisal, approval, transfer and dissemination?

→ Must be defined by the records creator at the time of notification – and responsibility for ”archival considerations” must be in place
Set-up for Submission of Digital Records

• Submissions to the National Archives must take place as non-system dependent SIPs, at times determined by the National Archives.

• ..usually every 5 years to avoid technological obsolescence. This means that current records can be submitted, while still in use by the records creator.

• All submissions are tested to ensure compliance with requirements.
Delimitation of SIP’s from RM-systems

• 1. Archival period: All files are closed, unfinished files must be reopened in a new archival period

• 2. Snapshot: All files and documents including metadata from a fixed period (e.g. 1.1.2007-31.12.2011).

• 3. Snapshot with closed files: A variation of model 2, with only closed files, but metadata from all documents and files from a fixed period
Formats for submission/preservation

Requirements are defined in an Executive Order on Submission Information Packages:

• relational databases
• XML (modified SIARD) & UTF-8
• TIFF, JPEG-2000, MP3, MPEG-2, MPEG-4, GML
• standardized documentation
• standardized structure
Standardized Mark-up of Certain Columns

- Identification of agency, if the system is used by more than one
- Document identification
- Storage form (paper, digital, irrelevant)
- Case file identification
- Case file title
- Document title
- Document date
- Sender/recipient
- Digital signature (i.e. information that has been extracted from a digital signature) [hardly ever used]
- Appraisal decisions
Structure and Content
Indices

- archiveIndex.xml
- docIndex.xml
- documentationIndex.xml
- fileIndex.xml
- tableIndex.xml
Test

- All submissions are tested using the test tool ADA
- The automated test is combined with a manual test
- Usually submissions are sent back to the records creator for corrections 2-3 times
Physical Preservation

- Distributed Digital Preservation
  - Several identical independent copies
  - Both optical and magnetic media
  - Several locations
- Currently about 100 TB born-digital data (in three copies), more than 4500 individual AIP’s.
Access: SOFIA

- AIPs from Electronic Records Managements Systems
  - have the same structure
  - solution: specific model for producing DIP resulting in a search form
  - usability: HIGH

- AIPs from other systems (business systems)
  - are very diverse
  - solution: general tool for making queries, general tool for making search forms
  - usability: LOW
Searching in DIPs of Records Management Systems

• Standardized interface
### Show only case file information

Double click on case file

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Group</th>
<th>Case Number</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90</td>
<td>2001-00-001</td>
<td>24-03-2001</td>
<td>Arne advising case</td>
</tr>
<tr>
<td>4</td>
<td>90</td>
<td>2001-00-001</td>
<td>31-10-2001</td>
<td>Birgit's advising case</td>
</tr>
<tr>
<td>8</td>
<td>90</td>
<td>2001-00-001</td>
<td>31-10-2001</td>
<td>Jane advising case</td>
</tr>
<tr>
<td>12</td>
<td>90</td>
<td>2001-00-001</td>
<td>05-12-2001</td>
<td>Marie Louise advising case</td>
</tr>
<tr>
<td>28</td>
<td>90</td>
<td>2001-00-001</td>
<td>05-12-2001</td>
<td>Peter's advising case</td>
</tr>
<tr>
<td>29</td>
<td>90</td>
<td>2001-00-001</td>
<td>31-10-2001</td>
<td>Lousie advising case</td>
</tr>
<tr>
<td>31</td>
<td>90</td>
<td>2001-00-001</td>
<td>31-10-2001</td>
<td>Nils advising case</td>
</tr>
<tr>
<td>32</td>
<td>90</td>
<td>2001-00-001</td>
<td>23-11-2001</td>
<td>Tore's advising case</td>
</tr>
<tr>
<td>53</td>
<td>90</td>
<td>2002-00-001</td>
<td>02-04-2002</td>
<td>Camilla advising case</td>
</tr>
<tr>
<td>54</td>
<td>90</td>
<td>2003-00-001</td>
<td>29-07-2003</td>
<td>Line's advising case</td>
</tr>
<tr>
<td>59</td>
<td>90</td>
<td>2003-00-001</td>
<td>09-05-2003</td>
<td>Winnie advising case</td>
</tr>
<tr>
<td>85</td>
<td>90</td>
<td>2003-00-001</td>
<td>02-06-2003</td>
<td>Anne's advising case</td>
</tr>
<tr>
<td>88</td>
<td>90</td>
<td>2004-00-001</td>
<td>07-12-2004</td>
<td>Simen advising case</td>
</tr>
</tbody>
</table>
Information of the document and link to the electronic document

Double click on document
Viewing the electronic document
Searching in DIPs of other digital records

Two options:

1. Use search form
2. Make your own query into the database
Preservation planning

• When should the preservation formats be changed?
• When should we migrate media or formats?
• What developments do we see in the use of IT among the authorities, and how may that affect us?
Questions?

E-mail: jds@sa.dk
Twitter: jdalsten

See also www.sa.dk
and www.digitalbevaring.dk
Risk assessment

• In digital preservation you will often find that there is no single correct answer
• Risk assessment is at the core of the development of a strategy
The CIA-Triad

- Confidentiality
- Integrity
- Availability
- Authenticity
Cost-benefit analysis