Evaluation of the

Format and Structure Conversion Project

December 2010
Table of Contents

1 Introduction ........................................................................................................................................ 7
2 Guide to Reading ................................................................................................................................. 8
3 Evaluation Format .............................................................................................................................. 9
   3.1 Assumptions .................................................................................................................................. 9
   3.2 Objective ...................................................................................................................................... 9
   3.3 Success criteria ............................................................................................................................. 9
   3.4 Methods ..................................................................................................................................... 10
      3.4.1 Evaluation methods .............................................................................................................. 10
      3.4.1.1 Method for evaluating if goals are reached (outcome evaluation) .................................. 10
      3.4.1.2 Impact evaluation method .............................................................................................. 10
      3.4.1.3 Methods for summative evaluation ................................................................................ 11
   3.4.2 Cost Calculation Method ....................................................................................................... 11
4 FSC Project Description ..................................................................................................................... 12
   4.1 Background ................................................................................................................................. 12
   4.2 The FSC Project ........................................................................................................................ 12
      4.2.1 Objective and Introduction ................................................................................................ 12
      4.2.2 The objects of the Project - Records .................................................................................. 13
      4.2.3 The Technical Objectives ................................................................................................ 14
      4.2.4 The Project .......................................................................................................................... 15
5 Summary of Evaluation ..................................................................................................................... 21
   5.1 Conclusions ................................................................................................................................. 21
      5.1.1 Achieving the formal goal .................................................................................................... 21
      5.1.1.1 Appropriation, budget and expenditure ......................................................................... 21
      5.1.1.2 Records .......................................................................................................................... 21
      5.1.2 The Evaluator’s Conclusions and Recommendations ........................................................ 22
      5.1.2.1 Future preservation policy and migrations .................................................................. 22
      5.1.2.2 Future Projects in General ............................................................................................. 26
      5.1.3 Draft to list of outstanding tasks related to the FSC project’s aims and the condition of the archival holdings ................................................................. 29
6 Evaluation of the FSC project ............................................................................................................. 30
   6.1 Prerequisites of the FSC Project ................................................................................................ 31
      6.1.1 Repackaging, registration classification and storage structure ........................................ 31
      6.1.1.1 Introduction and Evaluation Object ............................................................................... 31
      6.1.1.2 Essential Conclusions ................................................................................................... 31
      6.1.1.3 Description ..................................................................................................................... 32
      6.1.2 Preservation Standard and Documentation Standard ......................................................... 33
      6.1.2.1 Introduction and Object of Evaluation ........................................................................... 33
      6.1.2.2 Essential Conclusions .................................................................................................. 33
      6.1.2.3 Description ..................................................................................................................... 33
   6.2 Pre-project .................................................................................................................................. 35
      6.2.1 Introduction and Evaluation Object .................................................................................... 35
      6.2.2 Essential Conclusions ......................................................................................................... 35
      6.2.3 Description .......................................................................................................................... 36
      6.2.3.1 Establishing output - of formats, structures, storage structure and documentation ... 36
      6.2.3.2 Overview and examination of records ............................................................................ 36
      6.2.3.3 Migrating a representative segment of records, including system development ... 37
      6.2.3.4 Resources ....................................................................................................................... 38
   6.3 Project Planning ............................................................................................................................ 39
      6.3.1 Introduction and Evaluation Object .................................................................................... 39
      6.3.2 Essential Conclusions ......................................................................................................... 39
6.3.3 Description

6.3.3.1 Planning Basis

6.3.3.2 Preparation of the Project Plan

6.3.3.3 Performance Requirements in the Performance Contract the Ministry of Culture

6.3.3.4 Project Management

6.3.3.5 Result

6.3.3.6 Measurability

6.3.3.7 Fulfilling the requirements of the results contract – generally described

6.3.3.8 Project Management

6.3.3.9 Reporting

6.3.3.10 Measurability

6.3.3.11 ”Annex 4 Performance Requirements to the Format and Structure Conversion Project”

6.4 Knowledge Building and Transfer

6.4.1 Introduction and Evaluation Object

6.4.2 Essential Conclusions

6.4.3 Description

6.4.3.1 Knowledge requirements - IT professional knowledge about older records

6.4.3.2 Competency Development

6.4.3.3 Quotations from External vendors

6.4.3.4 Prospective use of knowledge and competencies from the FSC project

6.5 Defining the Migration Method

6.5.1 Introduction and Evaluation Object

6.5.2 Essential Conclusions

6.5.3 Description

6.5.3.1 Establishing the requirements to the migration tools, period A and B records

6.5.3.2 Determining the requirements to the migration tool, period C Records

6.5.3.3 Support of manual work – workflow description

6.5.3.4 Requirements for supporting manual work – migration principles

6.6 Preparation of Data

6.6.1 Introduction and Evaluation Object

6.6.2 Essential Conclusions

6.6.3 Description

6.6.3.1.1 End result and quality of end product

6.6.4 Use of Resources

6.7 Preparation of Documentation

6.7.1 Introduction and Evaluation Object

6.7.2 Essential Conclusions

6.7.3 Description

6.7.3.1 Documentation Standard

6.7.3.2 Collection and Systemisation of Documentation

6.7.3.3 Scanning and Registration

6.7.3.4 FSCs documentation on the migration itself

6.8 Infrastructure and Hardware

6.8.1 Introduction and Evaluation Object

6.8.2 Essential Conclusions

6.8.3 Description

6.8.3.1 Infrastructure

6.8.3.1.1 PCs

6.8.3.1.2 Migration Storage

6.8.3.1.3 Network

6.9 System Development
6.9.1 Introduction and Evaluation Object .......................................................... 72
6.9.2 Essential Conclusions .................................................................................. 72
6.9.3 Description .................................................................................................. 74
   6.9.3.1 KonvOld ............................................................................................... 74
   6.9.3.2 Pre-migration tools .............................................................................. 76
   6.9.3.3 Miscellaneous ...................................................................................... 77
   6.9.3.4 Konv2000 .......................................................................................... 78
   6.9.3.5 MARY ................................................................................................ 80
   6.9.3.6 DEA ................................................................................................... 81
6.9.4 Review of the common traits and results of software development .......... 82
6.10 Migration and Preparation of Metadata ...................................................... 84
   6.10.1 Introduction and Evaluation Object ....................................................... 84
   6.10.1.1 Period A and B Records - Introduction ............................................. 85
   6.10.1.2 Period A and B records – Essential Conclusions ............................. 87
   6.10.1.3 Periods A and B records – Description ............................................. 88
      6.10.1.3.1 Periods A and B records – Migration method ............................ 88
      6.10.1.3.2 Period A and B records – End Result and Quality of End Product 102
   6.10.1.3.3 Use of Resources ......................................................................... 107
   6.10.1.4 Period C Records - Introduction ....................................................... 108
   6.10.1.5 Period C Records – Essential Conclusions ....................................... 108
   6.10.1.6 Period C Records – Description ....................................................... 109
      6.10.1.6.1 Period C Records – End Result and Quality of End-Product .... 109
      6.10.1.6.2 Period C Records – Use of Resources ....................................... 110
6.11 Transfer to Archival Storage .......................................................................... 111
6.12 Audio/ Visual Records ................................................................................... 112
   6.12.1 Introduction and evaluation object ......................................................... 112
   6.12.2 Essential Conclusions .......................................................................... 112
   6.12.3 Description of Digitisation .................................................................... 113
      6.12.3.1 Audio ............................................................................................. 113
      6.12.3.2 Motion Pictures ............................................................................ 114
      6.12.3.3 External suppliers .......................................................................... 114
   6.12.4 Description of migration ....................................................................... 115
   6.12.5 Conclusions .......................................................................................... 116
6.13 Finances .......................................................................................................... 119
   6.13.1 Budget versus Actual Expenditure ......................................................... 119
   6.13.2 Actual time used .................................................................................... 122
7 Glossary .............................................................................................................. 124

Tables
Table 1 - Overview of records in the FSC project ............................................. 14
Table 2 - Allocation of time used on each record type (A, B and C) .................. 14
Table 3 - Hours used on FSC tasks and their interrelationship ......................... 18
Table 4 - Tools and materials used during the FSC project ................................ 19
Table 5 - Appropriation versus budget versus expenditure ............................. 21
Table 6 - FSC’s total expenses 2005-2008 ......................................................... 21
Table 7 - Overview of media migration project’s expenses ............................. 33
Table 8 - Overview of the expenses pertaining to the development of the preservation standard and the documentation standard ........................................ 34
Table 9 - Main variations in structure and format of the data before migration .... 37
Table 10 - Performance Requirements ............................................................... 45
Table 11 - Restructuring of competency development in the FSC project ...... 49
Table 12 - Example of migration status updating - 2007 ........................................ 54
Table 13 - Example of migration status updating - 2008 ........................................ 55
Table 14 - Reproduction of migration principles ......................................................... 56
Table 15 - Allocation of information package IDs ...................................................... 60
Table 16 - Total resources used for data preparation .................................................. 62
Table 17 - Criteria for choice of supplier for the scanning of the paper documentation .... 66
Table 18 - Overview of budget and actual usage when scanning the paper documentation . 67
Table 19 - Expenses for infrastructure and hardware .................................................. 70
Table 20 - Konv2000 task portfolio ......................................................................... 79
Table 21 - Resources used in-house for software development in the FSC project .......... 83
Table 22 - Overview of migrated records in FSC ......................................................... 84
Table 23 - Data basis for migration of period A and B records .................................... 85
Table 24 - Example of workflow description documents with versions ....................... 88
Table 25 - Arranging the records, as far as possible, in similar structures and systematic uniform series ................................................................. 89
Table 26 - Extent of digital descriptions in number and in elements .............................. 94
Table 27 - Specially developed data type migration routines and their uses .................. 96
Table 28 - Specially developed code page migration routines and their uses ................. 99
Table 29 - The migration tool’s (KonvOld) versioning ............................................... 99
Table 30 - Machine time for end migration ............................................................... 101
Table 31 - End product for period A and B records ..................................................... 103
Table 32 - Status of accepted errors in the end product ............................................ 104
Table 33 - Primary key error ..................................................................................... 104
Table 34 - Foreign key error ....................................................................................... 105
Table 35 - Code list error .......................................................................................... 105
Table 36 - Character error ........................................................................................ 105
Table 37 - Known errors .......................................................................................... 105
Table 38 - Lack of compliance with demands for documentation ............................... 106
Table 39 - Migration phases’ machine and manual time consumption ....................... 107
Table 40 - Overview of pre-migration and migration ................................................. 107
Table 41 - Time usage for pre-migration and migration ............................................. 108
Table 42 - Total number of hours used for the migration of period C records, including system development ................................................................. 110
Table 43 - Result of the migration of the audio/visual records ..................................... 112
Table 44 - Budget and actual expenditure for migrating of audiovisual records ........... 116
Table 45 - Result of migration of audio/visual records ............................................... 117
Table 46 - Resource and vendor discrepancies in the migration of audio/visual records ............................................................................................................... 117
Table 47 - Appropriation versus budget versus expenditure ..................................... 119
Table 48 - Appropriation versus expenditure ............................................................ 119
Table 49 - Budget versus actual expenditure ............................................................. 120
Table 50 - FSC tasks - time used and cross relations .................................................. 122
Table 51 - Glossary .................................................................................................. 124

Figures

Figure 1 - Data and workflow in FSC ........................................................................ 16
Figure 2 - FSCs migration process .............................................................................. 16
Figure 3 - Evaluation area ....................................................................................... 30
Figure 4 - Migration process FSCs pilot phase ......................................................... 38
Figure 5 - Example of a scanned documentation in TIFF ....................................... 63
Figure 6 - Example of a digital description ............................................................. 68
Figure 7 - MARY’s user interface ........................................................................... 81
Figure 8 - Machine migration of Period A and B records ........................................ 86
Figure 9 - Process of migrating records ................................................................. 89
Figure 10 - The digital description in MARY – Document affiliation to a specific record .... 91
Figure 11 - Description of records and data .................................................................. 92
Figure 12 - Digital description with record definition ....................................................... 93
Figure 13 - Digital description with mark-up of primary and foreign keys ...................... 94
Figure 14 - Example of erroneous reporting and a code page which resulted in incorrect character translation ........................................................................................................ 95
Figure 15 - The ensuing correction of the mistake by providing the correct code page .... 96
Figure 16 - FSC tasks’ reciprocal relations ..................................................................... 123
1 Introduction

In the Performance Contract 2009-12 for The Danish National Archives, The Ministry of Culture entered into agreement with the The Danish National Archives on performing a self-evaluation of the migration project1, Format and Structure Conversion Project (FSC), which was completed by The Danish National Archives between 2005 and 2008:

2010: Presentation of the self evaluation report on format and structure conversion2.

The Format and Structure Conversion Project (FSC Project) rendered the entire collection of digital records compliant with the Executive Order no. 342 of 11 March 2004 on information packages of preservation worthy data from electronic filing systems.

This Executive Order is almost outdated as it was replaced by a new order3 of 1 September 2010, and will finally be revoked on 1 July 2011.

A new “format and structure conversion” is hence expected. Since the migration process is the central activity – from an economical perspective - in the logical preservation4, an evaluation of the experiences gained by the Danish National Archives from the FSC Project would serve as a valuable data basis for the planning of future preservations.

This report constitutes a self-evaluation, which is the first phase of a project5 aimed at developing strategy proposals for logical preservation and implementing concrete planning of future migrations within the framework of the Performance Contract 2009-12:

2012: A revised plan for migrations including clarification of the tools to be developed6.

---

1 The term “migration” is used instead of “conversion”, cf. ISO 14721:2003, Space data and information transfer systems – Open archival information system – Reference model.
3 Annex 2 – Executive Order no. 1007 of 20 August 2010 on information packages.
4 In digital preservation logical – or functional – preservation ensures that preserved data can be interpreted in the future.
5 Preservation Planning Project (PPP)
2 Guide to Reading

This report is intended for three different user groups (The Ministry of Culture, The Danish National Archives and national and international experts). Therefore, parts may be irrelevant to one group, and a selective reading would hence be recommended.

- **Description of Methodology**

  The evaluation and calculation methods are briefly described.

- **Description of the FSC project**

  The FSC project is briefly described with a view to providing a cursory understanding of the project.

- **Summary of the evaluation**

  The report’s most essential conclusions and recommendations are presented in this chapter.

- **Evaluation of the FSC project**

  This chapter gives a detailed description and an evaluation of the FSC project. This chapter is very comprehensive and is intended to substantiate the general conclusions of the report.

Each section follows this structure:

- **Introduction and evaluation object**

  This section introduces the object of evaluation with a brief definition of the focus areas of the evaluation.

- **Essential Conclusions**

  The essential conclusions of the object collated.

- **Description**

  The object is described and evaluated in detail.

- **Glossary**

  A list of technical vocabulary and abbreviations is provided at the end of the document.

- **Appendix**

  Reference is made to a number of appendixes, which are in separate documents. These documents are not translated into English and are only available on demand.
3 Evaluation Format

3.1 Assumptions

The following assumptions are decisive for the evaluation:

Self-evaluation
The evaluation of the FSC project is a self-evaluation, which is a well known and recognized evaluation method. The advantages of the method are a keen understanding of the object being evaluated. The disadvantages are a reduced objectivity when compared to an evaluation performed by a third party. The report was prepared by an internal project group comprising persons who were active in the FSC project, along with others who were subsequently appointed to the project.

Retrospective evaluation
The initial FSC project description did not include any clause regarding evaluating the project. This means that there were no clear objectives for what and how to evaluate. Consequently, documentation and time-keeping for the FSC project were not adjusted to suit an evaluation.

The Project manager’s absence
The FSC project manager resigned his position about 8 months prior to completion of the project. Hence the evaluator had not been made privy to the entire projects insights.

3.2 Objective

The objective of the evaluation is twofold:

Knowledge gathering
The Danish National Archives needs to be prepared for future migrations; hence, knowledge gathering at this stage is crucial if migration is to be conducted at the most cost effective and secure way. Therefore, a systematic evaluation of lessons learned from the FSC project is very useful to The Danish National Archives’ future preservation planning, cf. The Preservation Planning Project.

Knowledge sharing
The evaluation of the FSC project has made it possible for The Danish National Archives to proliferate the acquired knowledge on migration, which at this time is a rather unique knowledge. Many other preservation institutions, in the process of building digital archiving facilities, have shown keen interest in the lessons learned from the FSC project; hence the decision to have the report translated into English.

3.3 Success criteria

The success criteria for the FSC project evaluation are closely linked to the objective of the evaluation, namely that:

- The lessons learned result in more efficient migrations.
- The evaluation constitutes a value for other preservation institutions facing the same challenges as we did.

Obviously the success criteria cannot be measured at the time when the report is published, but will first be apparent later.
Besides these functional success criteria, the evaluation should also:

- Fulfil the objectives of the Performance Contract

### 3.4 Methods

Here is a short description of the methods utilised in the evaluation report. Please refer to the appendixes for details on each method.

#### 3.4.1 Evaluation methods

The following part describes the general and the specific evaluation methods used for this evaluation.

The evaluation objective and methods of achieving the desired outcome are as follows:

- Evaluate if the FSC project’s central objectives have been achieved
  - By evaluating whether goals were reached (Outcome evaluation)
- Evaluate which impacts the FSC project has had
  - By impact evaluation
- Evaluate which of the FSC project’s lessons learned should be used for planning future migrations
  - By summative evaluation
- Publish the FSC project results nationally and internationally
  - By editing the evaluation project and translating it into English

#### 3.4.1.1 Method for evaluating if goals are reached (outcome evaluation)

Evaluating if goals are reached requires that it should be possible by the end of the project to measure whether the objectives established at the onset of the project have been achieved.

The methodical basis for the evaluation of goals reached is, therefore, a quantitative data accumulation of the project’s own objectives and results. The elements that constitute the data collection are grouped in a number of basic indicators related to the specific objectives, activities and results of the project.

#### 3.4.1.2 Impact evaluation method

The objective of impact evaluation is to evaluate the projects intended and unintended effects.

Impact evaluation is hence, to a large degree, based on qualitative measurements, that are not only based on the project’s own goals but also on goals defined by the evaluator.

In order to be able to gauge a project’s impact, we must look at the assumptions the project manager had at the beginning of the project about its impact and what short and long term effects – besides the quantifiable – the project has had.

---

1. A more detailed description of the method is found in appendix 3 – Methods for evaluating the FSC project.
3.4.1.3 **Methods for summative evaluation**

A summative evaluation is an evaluation that collates the lessons learned with a view to future uses.

The methodological approach for the summative evaluations is based on both the quantitative and qualitative data collection.

We performed a systematic mapping of the FSC project to determine the extent of reaching goals and to evaluate the effect. We analysed the elements of the project that can be used prospectively.

3.4.2 **Cost Calculation Method**

The FSC project’s budget was based on a norm that calculated one man-year to equal 1,297 time registration hours.\(^8\)

In order to be able to maintain a comparison between budget and actual expenditure, this evaluation will make use of FSCs own norm.

The average annual salary for year 2005 and the ensuing *actual* hourly rate - based on the above actual man-year – are calculated for the three profiles who were employed to perform the project.

The weighted average actual hourly rate, which the employees were paid whilst working on the project, is calculated based on these figures and in accordance with the project planning division of labour.

When presenting the price of performing a task in an evaluation report, the following elements should be present whenever possible:

- Time consumption in hours and/or man-years
- Expenses in monetary units

When beneficial or necessary, the expenses should be expressed as *total expenses*, i.e. expenses in monetary units plus man-hours converted to a monetary unit by using a suitable wage level or a weighted average wage.

See Annex 4 – Memo on man-year and hourly rates – for general documentation pertaining to the utilised method for expense calculations and for an example of the *total expenses*.

---

\(^8\) This is a *net man-year*, i.e. all other than actual work hours have been deducted e.g. lunch breaks, vacation time, hours used for department meetings and education. Hence there is a difference between an actual man-year of 1,540 time registered hours (korrekt?) and the calculation normally used by The Danish National Archives.
4 FSC Project Description

4.1 Background

Since 1973, The Danish National Archives has received public digital records. Initially, The Danish National Archives only accepted electronic databases, but from the 1980s, it also accepted electronic filing systems. At the end of the 1990s, it also became possible to receive electronic document management systems.

The older digital records existed in numerous formats and structures. This was because, until the end of the 1990s, the preservation standards available for digital archiving were not sufficiently defined. There were no format or structure descriptions available in a machine-readable format. As a result, it was difficult to test whether the digital records complied with the accompanied hardcopy and structure descriptions; this in turn led to the test not always being conducted and when done, not always sufficient in extent.

Furthermore, until the end of the 1980s, there were no comprehensive requirements to the structure or format of the digital records submitted, nor were there any guidelines as to how this should be documented.

Hence, it has been decided that all digital records should be migrated to the preservation standard, which coincides with the submission standard applicable at that time. The standard describes which data structure and formats the records should be preserved in.

4.2 The FSC Project

The ensuing passages provide an abstract of the FSC project, giving a general understanding of the project. The project is described in detail in the evaluation part, cf. Chapter 6,

---

9 “Electronic databases” is The Danish National Archives’ term for (professional) databases. “Electronic filing systems” denotes databases with reference to hard-copy paper based case files.

10 Annex 5 – Executive Order no. 342 of 11 March 2004 on information packages worthy of preservation data from electronic archive systems.
4.2.1 Objective and Introduction

The FSC project was a transformation\(^{11}\) type migration project, with the overall objective of:

- Saving data from obsolete information-bearing structures and formats, and from decayed physical media (audio-visual records)
- Standardising data in such a way to make it feasible to enable future automated migrations and standardised accessibility to the records.

The FSC project took 4 years to complete – from 2005-2008 – and included migration of all formats and structures that did not live up to the requirements set forth in Executive Order no. 342 of 11\(^{th}\) March 2004. This entailed that all hierarchical databases were migrated to relational ones; that all code pages were migrated to ISO 8859-1 (Latin 1); that all packed fields were unpacked; that all variable-length records were changed to non-variable; that all documentation was scanned and documented in accordance with the requirements, and that metadata were created.

The project cost just over 30 ‘FSC’ man-years and the aggregate expense for purchase of software, hardware and external services amounted to well over 1 million DKK. The project had a steady staff head-count of between 10 and 15 employees. The project also had a steering committee and a reference group responsible for handling production and specialist questions. The steering group referred to The Danish National Archives’ management.

4.2.2 The objects of the Project - Records

The Danish National Archives’ holdings\(^{12}\) of non-standardised records, which were to be migrated, originated from three periods:

- **Period A**: submitted before the 1998 rules\(^{13}\) came into effect. These records were the most difficult and complicated to migrate because of their complex structures and unique formats.
- **Period B**: submitted in compliance with the 1998 rules. Although these records were technically easier to migrate than those from period A, they were much bigger and included many code values and fields which had to be manually keyed in. Hence, the migration of this period’s information package demanded many resources as well.
- **Period C**: submitted in compliance with the 2000 circular\(^{14}\) or later\(^{15}\). These records encompassed 942 information packages, but the number was constantly increasing as the...

---

\(^{11}\) Cf. OAIS terminology (Open Archival Information System (ISO 14721:2003, s. 1-13)): *A Digital Migration in which there is an alteration to the Content Information or Preservation Description Information of an Archival Information Package. For example, changing ASCII codes to UNICODE in a text document being preserved is a Transformation.*

\(^{12}\) Note that the project was conducted by The Danish National Archives, and that the records that were migrated were solely records from public administration and courts of law. The FSC project was thus not entrusted with the task of migrating The Danish National Archives other digital collections, including research data from The Danish Data Archives (DDA, Dansk Data Arkiv).


\(^{14}\) Appendix 6 – Circular no. 4 of 14 January 2000 on Public Authorities submission of electronic archive systems to The Danish National Archives.

\(^{15}\) Also records submitted in accordance with Executive Order no. 342 and Circular no. 25 of 2002 on Public Authorities submission of electronic archive systems to The Danish National Archives (cf. annex 9) belonged to this category. The reason the submissions from Executive Order no. 342 should also be migrated to a preservation standard, which was identical with their submission standard, was that the test tool had undergone several development steps and therefore tested the submissions differently (versioning). This problem is explained later in this report.
period also included applicable submission standards. The task of migrating that period’s records was substantially easier than that of the preceding periods, since there was no large discrepancy between the 2 preservation standards (circular no. 4 and Executive Order no. 342), thus enabling a high degree of automated migration.

The records were categorized in information packages. An information package corresponds to a specific submission for a given period from an IT system, e.g. in the form of data from a registry for a 1-year period, or from an electronic filing system for a 5-year period.

The content object of the migration is illustrated in a directory, which is classified based on who created the archive and themes: in annex 29 – Public Authorities and Topics, a provenance and thematic list is shown of the records from periods A and B, besides the records from period C, which FSC had machine migrated.

The following table provides an overview of the records that were to be migrated.

<table>
<thead>
<tr>
<th>Record type</th>
<th>Submitted during the period</th>
<th>Content</th>
<th>Data size</th>
<th>Number of files</th>
<th>Number of information packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period B</td>
<td>1998-2000</td>
<td>Data files on CDs, disks and paper documentation</td>
<td>419 GB</td>
<td>8,078</td>
<td>641</td>
</tr>
<tr>
<td>Period C</td>
<td>2000-</td>
<td>Electronic filing systems and registries</td>
<td>1,187 GB</td>
<td>-</td>
<td>942</td>
</tr>
<tr>
<td>All Total</td>
<td>-</td>
<td>-</td>
<td>1,777 GB</td>
<td>11,187</td>
<td>2,233</td>
</tr>
</tbody>
</table>

It has not been possible to accurately allocate the time used on each period separately, since some of the tasks performed are to the benefit of all three record types (e.g. transfer of preservation media). It has thus been necessary to make an artificial breakdown of time. However, the following table shows an accurate picture of the time allocation:

<table>
<thead>
<tr>
<th>Record type</th>
<th>Time used</th>
<th>Man-years</th>
<th>% Allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>17,889</td>
<td>13.79</td>
<td>53</td>
</tr>
<tr>
<td>B</td>
<td>12,853</td>
<td>9.91</td>
<td>38</td>
</tr>
<tr>
<td>C</td>
<td>3,184</td>
<td>2.45</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>33,926</td>
<td>26.16</td>
<td>100</td>
</tr>
</tbody>
</table>

The time used here is not equivalent to all the time used in the FSC project, since only tasks are included that can be directly related to the handling or processing of the three record types. Therefore, the time used on migrating e.g. audio/visual records is not included in the calculation.

4.2.3 The Technical Objectives

The main objective of the project was to transform data into information packages that complied with the requirements defined in the Executive Order no. 342; which, basically, means:
• **Common format for data files**
  o All tables must have fixed record length\(^\text{16}\).
  o The content must be presented using a uniform code page: ISO 8859 Latin 1.
  o Fields in data files must be described using ISO standard data types such as NUM, REAL, STRING, DATE, etc.

• **Common Structure**
  o General information (documentation on the IT systems administrative function, structure and functionality)
  o Help tables (information on submission, including context information, reference information and descriptive information)
  o Documents
  o Tables
  o Metadata, description of information package’s tables with field description and the mark-up of relations between them

• **Common format for documents** (TIFF)
  o Digitising of paper documentation
  o Migrating of digital documents (e.g. from Word to TIFF)

### 4.2.4 The Project

Most of the project required specialized knowledge, which only existed in the National Archives Preservation & Disposal Department, and hence the tasks were performed internally. Few of the tasks were outsourced, e.g. scanning of paper documentation for the digital records and the digitisation of analogue audio/visual records (sound and film).

The migration of the records to the preservation standard, applicable at the time, was done with a migration system which was developed in-house. Each migrated information package was transferred to preservation media in compliance with the implementation of The Danish National Archives’ new media strategy of 2004\(^\text{17}\). All measures of the project’s progression were performed in the archive database DAISY\(^\text{18}\), while the registering of the records was performed partially in MARY\(^\text{19}\) and partially in the Danish Archive Information System (DAISY).

In order to ensure demands for safety, security and confidentiality, the work was performed on an existing closed records network (black net) in the National Archives IT workshop at the National Archives.

The infrastructure devised for the project had the following structure:

---

\(^{16}\) According to the Executive Order no. 342, it is optional to choose between fixed and variable record lengths. However the FSC project chose systematically fixed record lengths.


\(^{18}\) DAISY is The Danish National Archives digital registry in which all information on the creator of an archive and on records included in The Danish National Archives’ collections is lodged.

\(^{19}\) MARY (Metodes Arkivalie RegistreringssYstem) is The Danish National Archives preservation database used to monitor the digital collections.
The infrastructure formed the framework for a migration process, which can be simply illustrated as follows:

1. **Pre-project**
   The pre-project included the surveys that were prerequisite for estimating the budget of the actual project. The pre-project included the following phases, of which stages 3 and 4 were not fully accomplished:

   1. Determining output – of formats, structure, storage structure and documentation
   2. List of records – preliminary registration and accumulation of knowledge on the records
   3. Examining the records – structures and formats in the collection
   4. Migration of the records – development and pilot migrations
   5. Preparation of budget negotiation proposals and project plans

→ Read more in chapter 6.2,
Pre-project, page 35Fejl! Bogmærke er ikke defineret. and chapter 6.3,
2. **Restructuring the records to new preservation structure**

The first task in the FSC project was aimed at transferring the older records into new preservation structures with a view to getting a unique identification of the information packages. The information packages were also transferred to preservation media.

→ Read more in chapter 6.6, *Fejl! Henvisningskilde ikke fundet.*, page 59.

3. **Scanning**

The task included the scanning preparation and the scanning of ca. 400 archive boxes with paper documentation, which belonged to the old submissions. The older records were documented in hard copy format. The paper documentation was thus the prerequisite for analysing and migrating the older records.

→ Read more in chapter 6.7,
4. Registration
The task dealt with a basic and comprehensive registration and continuous updating and maintenance of data in MARY. Moreover, an accessibility of registration data in DAISY was made possible, which was necessary in order to enable users to search the records.

➔ Read more in chapter 6.6, Fejl! Henvisningskilde ikke fundet., page 59.

5. Development of digital descriptions
The digital description included the development of a standard for the description of the records. This description standard was used to key-in the information from e.g. paper documentation on records descriptions. This was necessary for enabling the records to live up to the preservation standard. The standard was developed in XML with its own schema, and was used for both achieving documentation on records in digital form. It was also used in order to get a digital form of the records’ documentation. The digital form would enable the migration application to test and migrate data based on this description.

➔ Read more in chapter 6.10,
6. Development of migration application and ancillary modules
This task included the system development of the programmes that would support the migration process. The programmes included KonvOld\textsuperscript{20}, which was the central migration application, besides a number of modules intended to handle the variations in code pages, data structures, etc.

→ Read more in chapter 6.9,

\textsuperscript{20} KONVerter OLD – CONVert OLD (records).
System Development, page 72.

7. Further development of test, CD-burning and registering systems
The task included the further development of a number of already existing systems. These included programmes used for testing (TEA)\textsuperscript{21}, preservation media transfer (DEA)\textsuperscript{22} and the export of automatically generated registration data to the preservation database, MARY.

\textbf{Read more in chapter 6.9,}

\textsuperscript{21} Test af Elektroniske Arkivalier - Test of Electronic Records
\textsuperscript{22} Distribution af Elektroniske Arkivalier - Distribution of Electronic Records (CD burning and transfer programme)
8. Migration and test
This task was the central task in the project. The migration of all records was performed in this phase, which stretched over almost the entire project. It was by far the most resource demanding part of the project. The task was divided into the individual tasks, which records had to go through before finally being digitally described, format and structure migrated, tested, proof read, and placed in preservation media along with their corresponding digitised paper-based documentation.

An important part of this task was to prepare 2 documents establishing the principles any migration would follow, and the routines any Convertor would use. The migration principles and workflow descriptions should ensure that the records were uniform and consistently handled, so they were migrated in a standardized manner.

➔ Read more in chapter 6.10,
Migration and Preparation of Metadata, page 84.

a. Migration of unique records
This task included the migration of a variety of special records, which proved to be extremely difficult to handle and not readily processed by the standard migration tool, KonvOld\textsuperscript{23}. The records could not be handled by the digital description format. Hence there was a need for the development of ad hoc tools and for the subsequent pre-migration. This required the use of experienced system developers with keen knowledge of the distinctive formats and structures. The unique records were prepared manually, one by one, in order to enable migration to proceed automatically. The records were initially migrated separately, but it turned out that ongoing migration was a better solution, as ad hoc programmes could thus be developed simultaneously with the test migrations.

\rightarrow Read more in chapter 6.10.1.1, Period A and B Records - Introduction, page 85.

b. Migration of newer records
The task included the machine migration of a number of newer information packages supplied in accordance with Circular No. 4.

\rightarrow Read more in chapter 6.10.1.4, Period C Records - Introduction, page 108.

c. Migration of analogue audio/visual records
Migration of analogue audio/visual records included the migration of the preservation worthy part of The Danish National Archives’ collection to the preservation standard applicable at the time.

\rightarrow Read more in chapter 6.12,

\textsuperscript{23} Cf. chapter 6.9.3.2, Pre-migration tools, page 71
The table below shows the resources used in the project calculated in hours, along with the percentage share used on each task. The calculations are unspecified and analysed in more detail in chapter 6.13,
Table 3 - Hours used on FSC tasks and their interrelationship

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Hours used</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Pre-project</td>
<td>3,192</td>
<td>7.45</td>
</tr>
<tr>
<td>1. Restructuring of records to the new preservation structure</td>
<td>1,782</td>
<td>4.16</td>
</tr>
<tr>
<td>2. Preparation and scanning</td>
<td>1,296</td>
<td>4.16</td>
</tr>
<tr>
<td>3. Registration</td>
<td>315</td>
<td>0.74</td>
</tr>
<tr>
<td>4. Development of digital description format, KonvOld and ancillary programme modules</td>
<td>3,724</td>
<td>8.69</td>
</tr>
<tr>
<td>5. Further development and test, burning and registration systems</td>
<td>995</td>
<td>2.32</td>
</tr>
<tr>
<td>6. Migration method, including devising workflows and migration principles</td>
<td>2,279</td>
<td>5.32</td>
</tr>
<tr>
<td>7. Migration incl. creation of information packages and test</td>
<td>18,535</td>
<td>43.25</td>
</tr>
<tr>
<td>8. Migration of unique records</td>
<td>1,583</td>
<td>3.69</td>
</tr>
<tr>
<td>9. Migration of Circular 2000 related records</td>
<td>224</td>
<td>0.52</td>
</tr>
<tr>
<td>10. Migration of analogue and audio/visual records</td>
<td>3,347</td>
<td>7.81</td>
</tr>
<tr>
<td>11. Project management, planning, etc.</td>
<td>3,998</td>
<td>9.33</td>
</tr>
<tr>
<td>12. Knowledge building and training</td>
<td>1,583</td>
<td>3.69</td>
</tr>
<tr>
<td><strong>Total for pre-project and the actual project</strong></td>
<td><strong>42,853</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

A wide range of tools were used for performing the tasks. These were:

Table 4 - Tools and materials used during the FSC project

<table>
<thead>
<tr>
<th>Project Management /Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sharepoint</strong></td>
</tr>
<tr>
<td>Forum on green net&lt;sup&gt;24&lt;/sup&gt;, which was the portal to all the FSCs steering and help documents e.g. migrating principles, procedures, guidelines, statistics, etc.</td>
</tr>
<tr>
<td><strong>information package ID/Quality assurance overview</strong></td>
</tr>
<tr>
<td>Each converter had a list detailing completed, in progress, and future tasks. Typically in the form of MS Excel worksheets.</td>
</tr>
<tr>
<td><strong>DAISY</strong></td>
</tr>
<tr>
<td>Registering of project progression.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preliminary Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DAISY</strong></td>
</tr>
<tr>
<td>General information on information package ID.</td>
</tr>
<tr>
<td><strong>Archive boxes for the individual information package IDs</strong></td>
</tr>
<tr>
<td>The physical documentation.</td>
</tr>
<tr>
<td><strong>MARY – Registration of documentation</strong></td>
</tr>
<tr>
<td>Documentation for information package ID in digital form. Access via MARY.</td>
</tr>
<tr>
<td><strong>Filing system case-files on submitting the specific systems</strong></td>
</tr>
<tr>
<td>In rare instances, information from filing system case-files from The Danish National Archives’ loft were used.</td>
</tr>
</tbody>
</table>

<sup>24</sup> Green net is an open network used for the exchange of ordinary information e.g. file drives and intranet. The opposite is a black net, which is a physically restricted network in which records are processed in accordance with applicable rules on confidentiality and security. The black net was used for migration in the FSC project.
### Data analysis

<table>
<thead>
<tr>
<th><strong>Tape dump</strong></th>
<th>Help for building the data structure. View in &quot;raw&quot; files from records before migrating. Overview of record structures, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CheckVariables</strong></td>
<td>Newer tools that provided an even better overview of the record structures. Stronger in certain areas than Tapedump, but did not have the same array of facilities.</td>
</tr>
<tr>
<td><strong>Code pages</strong></td>
<td>Finds any given characters in data.</td>
</tr>
</tbody>
</table>

### Design and Construction
Preparing digital description and migration. From raw data to information package

| **MARY** | - Documentation registration  
- Gateway to KonvOld.  
- Gateway to TEA test  
- Transfer to KORAID |
| **KonvOld** | In-house developed migrating programmes:  
- Validation of XML file created on the basis of the digital description of information package ID  
- Launch of Infopath, typical starting point for digital description  
- Validation of files, records and fields in the digital description  
- Migrating of parts of or the entire information package ID  
- Adding documentation |
| **InfoPath** | Forms tool: used for typing in the digital description. Unfortunately not suitable for large data quantities and complex relations. |
| **TextPad** | Text editor which is a good alternative to InfoPath; fantastic editor functions and solid and strong when handling big and complex descriptions. |
| **Migrating principles** | Up-dated description of the principles and rules applicable for the accrual of information packages. Often used during this phase for searches of code pages and data types. Included a clear description of these forms, content and areas covered. |
| **Create XML code** | Generates repeated XML codes for digital descriptions. |
| **MAVDA** | The migration tool for audio visual records. Information packages of this type of records were generated via this tool. |
| **Access-db (Heindb)** | Basic data for MAVDA. Basic data for audio/visual records. |

### Test

<p>| <strong>TEA</strong> | Additional test of digital description (full test of data and relations). Moreover, offered a different perspective of the information package. |</p>
<table>
<thead>
<tr>
<th>Various Tools and Aids</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XMLSPY</strong></td>
</tr>
<tr>
<td><strong>TextPad</strong></td>
</tr>
</tbody>
</table>

### 5 Summary of Evaluation

The summary addresses the following:

- Achieving the formal goals
- Evaluator’s general, comprehensive conclusions and recommendation
- Proposal to list the FSC project’s outstanding items and the collections status

#### 5.1 Conclusions

##### 5.1.1 Achieving the formal goal

The FSC project’s appropriation was adhered to, but the project management had slightly under-budgeted the project (4.5 %):

**Table 5 - Appropriation versus budget versus expenditure**

<table>
<thead>
<tr>
<th></th>
<th>Appropriation25</th>
<th>Budget26</th>
<th>Expenditure27</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DKK</strong></td>
<td>14,900,000.00</td>
<td>13,945,562.00</td>
<td>14,603,316.40</td>
</tr>
</tbody>
</table>

Note that DKK 350,000 from running costs was transferred to the salaries account.

**Expenditure (DKK)**

The table below shows a statement created 2 years after finalising the project. The table gives a tentative estimate on the project’s costs. Please note that the calculation does not include the hours the project management used after the project manager resigned, nor does it include the hours used by in-house IT.

**Table 6 - FSCs total expenses 2005-2008**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct wages</strong></td>
<td>929,003.67</td>
<td>2,598,842.75</td>
<td>2,915,078.09</td>
<td>3,129,736.85</td>
<td><strong>9,572,661.36</strong></td>
</tr>
</tbody>
</table>

25 See Annex 32 – Performances Contract 2005-08_annex_to_demands_14
26 Annex 25 – The FSC projects budget
<table>
<thead>
<tr>
<th>Overhead expenses</th>
<th>450,193.18</th>
<th>1,174,480.52</th>
<th>1,358,800.32</th>
<th>1,142,987.01</th>
<th><strong>4,126,461.04</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Other direct expenses</td>
<td>75,000.00</td>
<td>201,511.00</td>
<td>258,360.00</td>
<td>369,323.00</td>
<td><strong>904,194.00</strong></td>
</tr>
<tr>
<td>Total expenses</td>
<td><strong>1,454,196.85</strong></td>
<td><strong>3,974,834.27</strong></td>
<td><strong>4,532,238.41</strong></td>
<td><strong>4,642,046.87</strong></td>
<td><strong>14,603,316.40</strong></td>
</tr>
</tbody>
</table>

Consumption (hours)
The project planning had estimated using 39,053 hours or 30.11 man-years. The final number of hours used was 42,853 hours or 33.04 man-years. The difference is 3,800 hours, 2.9 man-years or just under 9% which is partially compensated for in the monetary amount used (see above).

5.1.1.2 Records

Number

The project’s main goal was achieved since all records were migrated, National Archive’s film collection, which was not even digitised. This project plan, cf. 6.12,
Quality
The goal of making future migrations easier was achieved; but a fully automated migration of the collection is not possible: a high number of records - but not a large amount of data - deviated from their original preservation standard before the FSC project migrated them. This meant that some of those records also deviated from the preservation standard after migration. Therefore, we prioritised maintaining the authenticity of the data instead of modifying them and stating how they were modified. The partial lack of standardisation of the records was, hence, not due to poor migration, but to a preservation policy prioritisation chosen based on the bad quality of the input.

5.1.2 The Evaluator’s Conclusions and Recommendations

After a thorough analysis of the FSC project, the Evaluator reached the following general conclusions taken from a broad review of the evaluated object; cf. Chapter 6,
The conclusions are substantiated in more detail later in the chapter with reference to documentation, annexes, etc.

The conclusions are of a summative nature and provide the basis for a review of:

- Future preservation policies and migrations
- Future projects in general

The Evaluator’s main conclusion is that the FSC project was an unprecedented project of very high standard, both for planning, execution and outcome.

The following conclusions, which question a number of issues, do not dispute the overall conclusion.

5.1.2.1 Future preservation policy and migrations

Resource demanding projects

Almost 50 man-years is a ball park figure of how much time it took to complete the FSC project (circa 33 man-years) and the project’s crucial prerequisites, the repackaging (circa 8 man-years) and the establishing of a preservation standard (circa 10 man-years). The figures should be interpreted with caution since not all assumptions of the calculation are fully taken into account, cf. Chapter 6.1,
Prerequisites of the FSC Project, page 31 and chapter 6.13,
The demand to stay within budget and the time constraint made it necessary to adjust quality.

The FSC project budget was mostly adhered to. However, the budget was in fact an appropriation, and hence should be adhered to. There are several examples where the budget was met at the expense of other factors such as product quality, performance, and quantity of migrated records.

The pre-project did not achieve all its stated objectives, e.g. an in-depth records and the development of prototype programmes, cf. chapter 6.2,
Digitisation of film, budgeted at DKK 700,000, was scrapped; cf. chapter 6.12,
Audio/ Visual Records, page 112. This decision was considered prudent by the project, as it was estimated that the analogue media the films were stored on was not at a risk preservation-wise. No documentation exists on the review of the media’s shape or standard. The recommendation to place the media in optimal preservation conditions is not yet complied with.

Too few resources were allocated to hardware. If the IT infrastructure had been in place on time, the project could have been completed several months earlier. Much frustration could have been avoided if e.g. the machine-processing of the migration itself had been satisfactory. Cf. Chapter 6.8,
Completing the handling of period C records was postponed until after the end of the project; cf. chapter 6.10.1.4, Period C Records - Introduction, page 108.

There were vague plans to user-test the migrated information packages, which was never done systematically. In the mean time, knowledge on the use of the records has been collected, and in general, The Danish National Archives’ Access functional entity has been very satisfied with the migrations. A log has been produced including inappropriateness and errors that have been detected; cf. Annex 7 - FSC Error List.

We never managed to find the sufficient funds needed to get a satisfactory solution to the primary and foreign key errors, which existed in some of the original databases and which had costly consequences for the accessibility of the information packages. Cf. 6.10.1.3.1, Fejl! Henvisningskilde ikke fundet., page 102.

Full preservation is very costly, but can be necessary

The preservation policy principle is expensive, but necessary to preserve our heritage

The Danish National Archives comply with the principle of never disposing of material, if it already has been deemed worthy of preservation once. Therefore, there is always the risk of using a lot of resources to preserve few damaged or faulty records at the expense of the majority of records. Examples:

The recovery project, which migrated damaged magnetic tapes, cost on average DKK 50,291 to migrate just one magnetic tape, while the price for migrating a magnetic tape of good quality was DKK 340 (factor 148). Cf. Chapter 6.1.1, Repackaging, registration classification and storage structure, page 31.

o Unique records. It took 70 times longer time to migrate older, non-standardised records (period A) than newer, standardised ones (period C). Hence, it took on average 0.23 hours to migrate one period C record, 13.72 hours for one for one period B record and 16.67 hours for one period A record. Other methods of calculation methods show the same result:

o Measured per file, it took ca. 158 times longer to migrate the older records compared to the newer C records (3.16 hours versus 0.02 hours)

o Per GB, it took 253.5 times longer to migrate one older record than a newer one (63.36 hours versus 0.25 hours).

Cf. Chapter 6.10,
Standardisation of data and tools is worthwhile

Standardisation of data is a prerequisite to ensure that digital preservation is economically sound, since the complexity and deviations are difficult to handle. The FSC project’s main objective was to ensure that the entire collection of The Danish National Archives was standardised.

A complete standardisation would not only ensure the ease of developing for future accessibility and migration tools, but would also allow automated migrations.

- **Standardisation is expensive, but a good investment.** Standardisation of data is hence an investment, which can result in a modest amount of resources for migration and standardisation of the records (cf. Period C records). There are, however, still examples of records (data and documentation) that cannot be automated now. Therefore, The Danish National Archives still have outstanding items including uniform document names in Geninfo\(^28\) as well as exemptions from the preservation standard. See item below: The goal is almost reached.

The goal is almost reached. The FSC project produced, in general, that complied with the preservation standard. However, some of them because they got exemption for a number of errors, which e.g. were prior to migration and was therefore not corrected. Cf. Chapter 6.10,

---

\(^28\) Geninfo (general information) is a document belonging to each information package. The document includes information on identifications (media and library identifications) besides a description of which documents would be included in the submission (i.e. submission records, user manuals, legal information, screen dumps of user interfaces, E/R diagrams, etc.) along with where the documentation of the submissions are.
• Migration and Preparation of Metadata, page 84.
• **System development of non-standardised records is expensive.**
  o Development of tools needed to handle the older records (period A and B), accounted for over 80% of the total system development costs.
  o 175 programmes were developed to handle the 167 records with unique structures and formats.

Programmes become cheaper to develop. **Konv2000, developed for namely period C records, took 184 hours to develop. For comparison, develop KonvOld, which was used for period A and B records’ formats. Cf. Chapter 6.9,**
KonvOld’s flexibility allowed for easier addition of modules and thereby handling many formats. Only 167 out of 1,291 records (just under 13 migrated before they were suitable as data input to KonvOld. The could be used directly as standard input. KonvOld handled the 87 % of records. This was only possible due to the ongoing, iterative expansion of KonvOld. cf. chapter 6.9,
A standardised, system-independent format was chosen for preparing the descriptions: An XML schema. This made it possible to use alternatives used for reading and maintaining the digital descriptions, cf. Chapter 6.9,
Two preservation databases are one too many. **Having to do (double)** MARY and DAISY required much manual resources - it took ca. 1 important to note that it might not have been economically feasible to systems within the scope of the FSC project. Cf. Chapter 6.6, Fejl! Henvisningskilde ikke fundet.
• Preparation of Data, page 59.

Software development method depends on knowledge of data

• **When data are not known, it is best to use iterative development.** The migration application, KonvOld, was developed iteratively. The main reason for this was that the records to be handled were of very different structure and format and were not known beforehand. The iterative development ensured the possibility of automated processing of 87% of the records, which was a huge advantage to the project.

• **When we assume we know the data**, the tool should be fully developed before it is used. The test application, TEA too underwent iterative development and existed in several consecutive versions, which to varying degrees met the requirements associated with the applicable preservation standard. This had an adverse consequence: the records were received, tested and approved (with errors) by various versions of the same tool, depending on when the submission actually occurred. Cf. the points below. It is unknown whether it is possible to avoid versioning of test tools.


Quality control is expensive but should be appropriate in extent and quality

• **It is costly to perform many random checks, but risky (and potentially more costly) to perform too few.** Random checks demand many resources but are necessary to perform on a large scale. There have been too many instances of rash conclusions – cf. the insufficient random checks of the magnetic tapes and of the quality of period C records, chapter 6.1.1, Repackaging, registration classification and storage structure, page 31 and chapter 6.10.1.4, Period C Records - Introduction, page 108.

• **Test tools must test exactly what is described in the preservation standard**, otherwise, the records would not be tested in compliance with the standard, which would make it impossible to perform future automated migrations and accessibility. As a minimum, it is required to know which version of the test tool the individual information package is tested with, and this version must be well documented in order to justify to what degree it lives up to the preservation standard, cf. chapter 6.10.1.4, Period C Records - Introduction, page 108.

Performance – software and hardware should efficiently be able to process a given amount of data

The technical infrastructure suffered from a number of shortcomings and late decisions that cost many resources.

• **Hardware lacked power.** The migration of records could have been performed much faster. Although, a late decision on the purchase of new machinery helped to meet the deadline, it was still not sufficient to make up for lost time. Cf. chapter 6.8,

29 Demands specification based on the preservation standard (cf. Executive Order 342).

30 Note: it must be emphasized here that what is included is what can be considered reasonable to machine test.
Infrastructure and Hardware, page 70.

- **Tools performed poorly in certain contexts.**
  - When using InfoPath for the digital descriptions, the speed of uploading and editing was very slow. This was especially the case for large records containing many tables, fields and long code lists.
  - KonvOld had the same problem of controlling the migration of large data quantities. The reason was that KonvOld did not use a database when controlling output data. Cf. Chapter 6.9.3.1, KonvOld, page 74 and chapter 6.10.1.3, Periods A and B records – Description, page 88

**Quality – the better the input data, the cheaper the migration**

It cost a lot of resources to improve on the poor quality of the records prior to migration, which emphasises the need for keen supervision and test when receiving records.

- **1.6 % of magnetic tapes cost 75 % of the total project amount** in the recovery project, because the tapes were in so poor shape that only an external specialist was able to restore the data, cf. chapter 6.1.1, Repackaging, registration classification and storage structure, p. 31.

- The quality of the input data were poor, and in some instances irreparable, which to a certain extent is reflected by the output data, as not all of these live up to their preservation standard. This meant that for some information packages, there was a need for exemptions from the rules, cf. chapter 6.6, **Fejl! Henvisningskilde ikke fundet.**, p. 59.

The quality of the documentation was not satisfactory and did not always complete description of the records. This meant that many resources the formats, structures and content of the records in order to optimise its digital form, cf. chapter 6.7,
• Preparation of Documentation, page 63.

Accessibility of the Period A and B records made easier by thorough, testing via KonvOId and TEA as well as manual quality control, cf. chapter 6.10,
• Migration and Preparation of Metadata, page 84.

5.1.2.2 Future Projects in General

Project planning – Tight control and loose methods

The FSC project was well planned. Nevertheless, there was always a clear understanding of the need for an ongoing learning curve. This enabled an improved decision making basis, but meant that the project ran the risk of deviating from the chartered course. This, however, did not occur.

The knowledge policy was efficient

The FSC project created a favourable environment for knowledge sharing; however, some of this knowledge was lost, cf. chapter 6.4, Fejl! Henvisningskilde ikke fundet., page 47.

• Migration principles and work-flow descriptions. Since the records were very different in type, it was necessary to accumulate and share a great deal of knowledge. This knowledge was preserved through meticulous and continuous written documentation of the aforementioned working procedures. The documents were constantly updated and discussed at the weekly project meetings.

• Quick access to information. Several factors were of great value to the daily work within the project, namely: access to the majority of the project’s information (intranet), the two documents mentioned above, interactive access to the records’ scanned documents via MARY and the digital description.

• Temporary employment and loss of specialized knowledge. Many of the project staff members gained both IT professional and professional archival expertise on digital preservation. This knowledge was largely lost when the employment contracts expired, but qua standardization of the records, that knowledge is no longer necessary for The Danish National Archives.

• Vulnerability of the project. Even though this did not give rise to major problems, and the vulnerabilities were partially addressed in the project planning risk assessment, the project remained vulnerable in two areas:
  o Terms of employment. The project risked seeing the temporary employees leave prematurely when nearing the end of the project, which would result in loss of knowledge at a critical time.
  o Staffing. The project operated with a skeleton staff. However, there was no contingency for illness or vacations, etc.

Focus on methodology and daily work

The documents describing work-flows and migration principles were invaluable in the daily work. They ensured a uniform, high quality and efficient processing of the records. Cf. chapter 6.5,
Defining the Migration Method, page 52.

- **It was not possible to determine the migration methodology beforehand,** since the feasibility studies did not sufficiently document the nature of the records. The three documents, which constituted the migration methodology were hence “dynamic” documents that reflected a process in constant change.

- **This created an explicit need for knowledge sharing,** which was formalised at the weekly meetings and in the collaboration forums in which the convertors worked together two at a time. All this led to updated, user-friendly and accessible documentation of methodology.

**Sustained precision**

The large amount of routine work and resource demanding manual tasks (over 6,000 hours) constituted a bit of a challenge, with much that could have gone wrong and time that could have been wasted. Based on the project staff feedback and on the machine-controlled data, we can ascertain that the tasks were handled to a very high level of sustained precision. The reason for this can be found in:

- **Control.** It was possible to machine control the production of some tasks. In addition, the convertors worked together in pairs. Cf. chapter 6.6, *Fejl! Henvisningskilde ikke fundet.*, page 59.

- **Professional pride.** Interview with the temporary project staff showed that this pride was decisive in maintaining the high quality of the manual work performed.

- **Delegation of responsibility** by the aid of serial allocation. The FSC project allocated to the convertors their “own” series of records, for which they were responsible. This sharing of responsibility partially explains the diligence with which the work was performed. Cf. chapter 6.10.1.1, *Period A and B Records*, page 85.

- **Passion and enthusiasm.** The organisation managed to create and maintain a high level of motivation amongst the temporary staff. Whether this was due to the project manager, the project staff, their common situation (temporary project employment) or a fourth reason, is difficult to assess. However, it was said that there was a very good chemistry amongst the staff and a good social environment.

- **The staffing committee succeeded** in hiring IT professionals who had earlier worked with projects that required the same level of structured work as in this project, cf. chapter 6.4, *Fejl! Henvisningskilde ikke fundet.*, page 47.

**Insufficient focus on the FSC project’s ancillary activities**

In general, there was too little focus on the activities that improved on, documented and quality assured the project. Two examples illustrate this:

- **Insufficient documentation and time registration**
  - Despite having a great deal of documentation on the FSC project, it is neither systematic nor archived in such a way that makes it suitable for evaluation. The Evaluator found it very challenging to navigate the documentation, especially when needing to find certain key fact reports on e.g. economic or quality issues.
  - Time registration was not sufficiently detailed; hence, it has been impossible to evaluate certain activities in relation to their use of resources.

With projects of the magnitude of the FSC, it would be prudent to project **before** starting it, in order to be able to avoid such deficiencies can however be debated whether it is cost-effective to provided a more hours used. Cf. chapter 6.3,
Project Planning, page 38, and other chapters in general.
- **Feasibility studies.** Several instances of insufficient feasibility studies which, if done, could have provided potential savings and efficiencies.
  - The condition of the magnetic tapes was already examined in 1995 through a random check. The tapes’ condition was found to be suitable, which it was probably not after all. Therefore, a repackaging of the media was not initiated at that time, which in turn resulted in the huge expenses incurred to salvage the tapes 10 years later. Cf. chapter 6.1,
The pre-project did not have enough time to examine a suitable sample of data to be migrated, which meant that it was not possible to produce a specification for KonvOld, which had to be developed simultaneously migration process. The fact that this turned out to be a very could not have been predicted. Cf. chapter 6.2,
Preparation for the system development was perhaps inadequate. **If there been set aside more time for surveying the market and trying out or open source tools, it might have been possible for the FSC project to resources on system development. Cf. chapter 6.9,**
Lengthy decision-making processes

- In certain instances, the chain of command was bureaucratic and impeded the quick implementation of decisions; cf. Annex 28 – *Memo re. Acute operational problems in FSC.*
  - Change management. Internal improvements in work processes gave rise to a greater need for making backups than originally planned. In spite of arduous meetings and negotiations, we did not succeed in getting more resources for backup and hence the FSC project resorted to doing the backups of the production data on loose media.
  - Purchase of additional machines. A need arose for more machine power than originally anticipated. The decision on which machines to buy took a long time, which delayed the completion of the project.
  - Optimal configuration of workstations (images) was never done by The Danish National Archives’ operations department and was left to be completed by the FSC staff.

- Fundamental decisions regarding migration methods (waivers and deviations from preservation standard) had an approximately 14-day turnaround time, which was both acceptable and necessary.

Outsourcing

Outsourcing of tasks does not entail freeing up of resources to internal pertaining to the task. The major share of the expenses for scanning of which was outsourced, was taken from sub-tasks performed in-house (55 %). Chapter 6.7,
• Preparation of Documentation, page 63.
• **Alternatives to in-house development.** The extent of outsourcing tool development should be considered. The obvious advantage of outsourcing is profitability, while the disadvantage is loss of control, difficulty of integration in the preservation environment, and efforts made in order to find a suitable vendor for such highly specialised tools. Cf. chapter 6.9.2, Essential Conclusions, page 72.

**Insufficient communication among the various projects**

Despite a common steering group for the FSC project and the TGP project\(^{31}\), the lack of communication had costly consequences.

- Migration of period A and B records allowed (after approved exemptions) the production of information packages, which included errors in relations, i.e. duplicates in primary keys and lack of foreign keys. We decided not to use a machine-based, technical correction of the errors, e.g. by designating “dummy” key values. That solution would be detrimental to the data’s authenticity but would have curtailed the development time of the accessibility system, SOFIA, by just under 1 year. Cf. chapter 6.10.1.1 Period A and B Records, page 85.

**During the process of buying hardware, the IT department had a fall project staff.** A better cooperation would have ensured a more performance than what was used earlier in the project, which would migration. Cf. chapter 6.3,
There was inadequate integration between the components for registration (Access department with DAISY) and accessibility (TGP). Possible to automatically import data from KonvOld to DAISY. Cf. chapter 6.3,
Perhaps the lack of communication can be explained by the fact that the project was not properly aligned with the rest of the organisation. Another reason could be that there was not enough time allocated to meetings amongst the various projects’ members. In the long run, it is recommended to completely integrate the preservation environment, as is directed by e.g. the OAIS model in which digital preservation – from data collection to logical preservation to accessibility – is considered as one organisational unit.

5.1.3 Draft to list of outstanding tasks related to the FSC project’s aims and the condition of the archival holdings.

A list will be prepared detailing which outstanding tasks the FSC project has, and what is required in order to enable fully automated migration of the collection.

Preparing the list is delegated to the Preservation Planning Project’s initial phases in 2011.
6 Evaluation of the FSC project

The evaluation of the FSC project has identified and evaluated the following areas of FSC project:

Figure 3 - Evaluation area

1. Prerequisites of the FSC Project
2. Pre-project
3. Project Planning
4. Knowledge Building and Transfer
5. Defining the Migration Method
6. Preparation of Data
7. Preparation of Documentation
8. Infrastructure and Hardware
9. System Development
10. Migration and Preparation of Metadata
11. Transfer to Archival Storage
12. Finances

Note that the sequence is functional and even though a chronological presentation was aimed at, it was neither possible nor expedient since many processes were performed iteratively.

Before giving a detailed evaluation of the FSC project, a number of presumptions for the project are described.
6.1 Prerequisites of the FSC Project

A media migration of the type *repackaging*\(^{32}\) and an *establishment of a preservation standard* were the two most important prerequisites for completing the FSC project on budget. They did not form part of the FSC project, but are briefly evaluated since they were of great importance to the project.

These two projects constituted a workload of 8 and 10 man-years respectively. This calculation is tentative since it has not been possible to make a precise estimate of the expenses for the establishment of the preservation standard. This was not possible because it took more than 15 calendar years to perform and the timesheets belonging to the task were not systematic.

Note that the audio/visual records that were migrated as part of the here; cf. 6.12,

\(^{32}\) OAIS terminology
6.1.1 Repackaging, registration classification and storage structure

6.1.1.1 Introduction and Evaluation Object

A repackaging denotes the migration of data from one type of physical media to another and presumes the existence or the creation of a registration classification and a storage structure.

The migration can e.g. occur from magnetic tape to CD R, as was the case with the repackaging project.

6.1.1.2 Essential Conclusions

The essential conclusions are that:

- **The budget was exceeded by 65 %.** This was due to that:
  - Some of the physical media had decayed so much that it was very expensive to salvage their data.
- **1.6 % of media were damaged.** This was because:
  - Deficient monitoring of the media’s physical status. Either the monitoring of the media’s physical condition was not sufficient or perhaps the warning signals were ignored.
  - Bad quality of media. Some of the media showed to be unsuitable for long time preservation because of production errors or poor quality.
- **1.6 % of the media cost 75 % of the projects total expenses.** It cost DKK 1,760,168 or DKK 50,291 per tape to restore the 35 defective magnetic tapes, while the remaining 2,165 tapes cost DKK 736,723 or DKK 340 per tape. That this was accepted is due to:
  - The data were worthy of preservation. Notwithstanding the principle of not disposing of records once they’ve been deemed worthy of preservation, it was necessary in the repackaging project to make a judgement call on preservation and on discarding content of the 2,200 magnetic tapes. All were found to be preservation worthy.
  - That one of the objectives of the repackaging project was a full migration of the entire tape collection.

---

33 In 1995, a report on the media’s general condition was written, *Test of Archived Magnetic Tapes*. The report was based on a random check control which showed not to be sufficiently comprehensive. Cf. journal no. 2000-262-0003-(34).
The essential **recommendations** were to:

- **The budget can better be kept on track if the feasibility study is done properly.** It must be noted, however, that a feasibility study will often be performed using random samples and that problems will not always be detected when it is only 1.6 % of the material that is defect.
- **The condition of the physical media must be carefully monitored on an on-going basis,** to ensure that the media migration can be initiated on time.
- **The physical media’s quality must always be good when purchased.**
- **On ingest, data should always be distributed on more than two media – and preferably from several different producers in order to avoid manufacturing defects.**

### 6.1.1.3 Description

The repackaging ran from 2001 to 2006 and processed magnetic tapes for the period 1973-1997. The goal of the repackaging was to move data from the magnetic tapes to CD-R and from there to a preservation RAID. It had gradually started to become more difficult to support the tape recorders and hence the accessibility to the magnetic tapes, the oldest of which were nearing their obsolescence. Furthermore, in some instances the magnetic tapes were in very poor condition, which was especially the case for tapes dating to the period 1980-88.

Part of the repackaging project dealt with the optimisation of the storage structure the data would be placed in. The structure was initially defined during the preparation of an earlier preservation standard, Circular no. 4.

Concurrent with the repackaging, a systematic registering method was devised which standardised the registration of the records. The registration methodology was a further development of the one already used described in Executive Order nr. 342. The registration methodology was implemented with the disk burn tool, DEA. The data in the registration methodology were stored by automating the burning of the CD-R media, simultaneously with the transfer of data to the preservation RAID.

The repackaging transferred the 2,200 magnetic tapes to CD-R creating 650 information packages (corresponding to period A records).

A few of the magnetic tapes (35) were challenging to migrate as they would not transfer error free. Therefore, in the years 2004-2006, a sub-project on the repackaging project, the ‘Recovery Project’, was initiated in order to try and recover data from the 35 magnetic tapes.

Both the repackaging and the recovery projects were successful as all data were recovered with the exception of a very small sample of the 35 poor-quality tapes (reportedly only 2 records). The project planning had budgeted with an expense of DKK 131,000 and a time usage of 3,222 hours, which is equivalent to a total expense of DKK 908,790.

The repackaging and the recovery project, excluding setting up the storage project and registration methodology, cost in total ca. DKK 550,000 and ca. 6.2 man-years, which corresponds to DKK 2,646,900 in total expenses.

The project thus ran 65% over budget.

---

35 Workforce were all IT professionals, for whom the hourly rate was calculated (DKK 241.40), cf. Annex 4 – Memo on man-years and hourly rates.
36 The evaluator did not find statements for these tasks.
The reason for the high cost of the project is mainly attributable to the recovery project, which took up ca. \( \frac{3}{4} \) of the total expenses.

Below is an overview of the media migration project’s expenses\(^{37}\).

### Table 7 - Overview of media migration project’s expenses

<table>
<thead>
<tr>
<th>Project</th>
<th>Time used (hours)</th>
<th>Expenses (DKK)</th>
<th>No. of migrated magnetic tapes</th>
<th>DK per migrated magnetic tape</th>
<th>Total actual expenses</th>
<th>Total budgeted expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repackaging</td>
<td>2,016.25</td>
<td>250,000</td>
<td>2,165</td>
<td>340</td>
<td>736,723</td>
<td>-</td>
</tr>
<tr>
<td>Recovery project</td>
<td>6,048.75</td>
<td>300,000</td>
<td>35</td>
<td>50,291</td>
<td>1,760,168</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8,065</strong></td>
<td><strong>550,000</strong></td>
<td><strong>2,200</strong></td>
<td><strong>661,75</strong></td>
<td><strong>2,496,891</strong></td>
<td><strong>908,790</strong></td>
</tr>
</tbody>
</table>

### 6.1.2 Preservation Standard and Documentation Standard

#### 6.1.2.1 Introduction and Object of Evaluation

Before starting the FSC project, standards were established for both preservation and documentation. Since neither were part of the FSC project, we have only included a few general conclusions in the following passage.

#### 6.1.2.2 Essential Conclusions

The most important conclusions are, that:

- **It is very costly to develop a preservation standard from scratch**, if the future objective is to gain a high level of automated migrations. The evaluator concludes, based on estimates and time registrations, that it cost The Danish National Archives circa 13,000 hours (or ca. 10 man-years) to compose the first preservation standard (Circular no. 4)\(^{38}\).
- **A preservation standard use-value can therefore e.g. be measured on to what degree its quality (degree of automation) offers cost savings in the long term.**

#### 6.1.2.3 Description

We chose the preservation standard from 2000. The requirements for this standard are stipulated the Executive Order no. 342 of 11 March 2004 on information packages of preservation worthy data from electronic archive systems\(^{39}\).

The reason why we chose this standard was that it would save money using an already developed standard that included test tools. Furthermore, using a standard would facilitate the development of both migrations tools and accessibility tools.

It was deemed that the standard would not be suitable for use with future fully automated migrations, but it was a first step. It was considered too costly and time consuming to develop a more advanced standard at that time. This could be risky for the records as many of them were in


\(^{38}\) Note: the numbers are estimates due to the insufficient documentation and time registration for the creation of the preservation standard.

\(^{39}\) Electronic archive systems are currently referred to as it-systems.
critical shape and the knowledge about them would disappear with the retirement of our older co-workers.

When we started the FSC project, we prepared a documentation standard\(^{40}\) based on the requirements in the preservation standard.

The general requirements to the documentation standard were defined as follows:

- There must not be any loss of information during the processing.
- The documentation must be as user-friendly as a hardcopy one.
- The processing must comply with the requirements and recommendations applicable for the information packages stated in Executive Order no. 342.

The development of the preservation standard\(^{41}\) cost almost 10 man-years. The documentation standard cost a fraction of the total expenses, namely 0.1 %. Unfortunately, it has not possible to get more detailed numbers than those presented below; hence it is difficult to make a varied evaluation of this particular area.

<table>
<thead>
<tr>
<th>Table 8 - Overview of the expenses pertaining to the development of the preservation standard and the documentation standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Documentation standardisation project</td>
</tr>
<tr>
<td>Establishing a preservation standard</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>


\(^{41}\) The submission standard was merely a further development of the two Circulars, dating back from 2000 and 2002 (Circular no. 4 of 14 January 2000 and Circular no. 25 of 8. March 2002 (cf. annex 9 - Circular no. 25 on public authorities submission of electronic archive systems to The Danish National Archives). The development expenses for complying with the two Circular and Executive order no. 342 are included in order to give an accurate view of what the price of development was for the preservation standard. Costs are approximate and are based on estimates by staff working with the development.
6.2 Pre-project

6.2.1 Introduction and Evaluation Object

The aim of the pre-project was to examine the records, which were the view to estimating the required resources and the degree of complexity was not concerned with the examination of the audio/visual records, cf. chapter 6.12,
The pre-project is evaluated briefly with emphasis on:

- Resource consumption
- The quality of the pre-project, including its significance for the FSC project

### 6.2.2 Essential Conclusions

**Aims of pre-project**

- **The pre-project completely achieved two out of four of the defined aims**, i.e.:
  1. Establishing the data output
  2. Overview of records
- **The pre-project partially managed to**:
  3. Examine the records
- **The pre-project did not manage to**:
  4. Migrate a representative cross-section of the records, including system development.

The reason for this was that the records were not uniform. Therefore, there was no time to either develop an analysis tool or migration tool with sufficient functionalities for a pilot migration of a representative segment.

The consequence of this was that a detailed analysis of the collection and the system development and pilot migration were first performed during the of the FSC project.

It is difficult to say whether the lack of achieving all goals had a negative or positive effect on the FSC project’s resource usage:

Firstly, it could be argued that an in-depth analysis of the collection and prototyping of the software could have avoided unforeseeable and costly surprises in the migration process.

On the contrary, a very small cluster of extremely heterogeneous records, which were difficult to handle, required a very high level of iterative development of the migration tool, KonvOld. In other words, it was practically impossible, already in the pre-project phase, to complete the development of a migration tool that could handle the records.

**The pre-project’s quality and budget**

The pre-project should counteract the FSC project’s difficulties and form the bases for preparing a project plan.

The pre-project’s share of the FSC project as a whole amounted to 7.45 %, which is considered quite normal for a pre-project.

The budgetary estimates for the FSC project were accurate (see chapter 6.3).
• Project Planning, page 38 and chapter 6.13,
This was despite the fact that a more detailed survey of the structures and formats of the older records was not performed. It is important to emphasise that throughout the whole project we tried to adjust the task to the budget (it was always the intention to stick to the budget); e.g. the digitization of the films was excluded.

- **The FSC project made efficient.** The precise resource estimation must be considered as proof that the pre-project’s efficiency aims were reached.

- **The pre-project was within budget,** because:
  - Several of the project’s key persons had an in-depth knowledge of the collection of The Danish National Archives.
  - Some of the time consuming and difficult to evaluate tasks were already completed, e.g. a defined preservation standard.
  - The budget was adhered to which e.g. was the reason why task no. 4 in the pre-project – development of prototype of the migration tool - was not completed.

### 6.2.3 Description

The pre-project included the FSC projects pilot project and was intended to investigate and deal with a number of issues concerning the format and structure conversion, i.e.:

- Establishing output: of formats, structures, storage structure and documentation
- List of records: tentative registration and accumulating knowledge of the records
- Examination of the records: structure and format of the collection
- Migration of sample of the records, including system development
- Preparing a budget and project plans.

The project ran from February to November 2005.

#### 6.2.3.1 Establishing output - of formats, structures, storage structure and documentation

Some of the crucial prequisites for performing a migration project in the calibre of the FSCs were already met from the beginning of the project. These include:

- Completed repackaging including the establishment of registration methodology and storage structure
- Set up preservation standard including documentation standards for records content.

These prerequisites are already described in chapter 3, *Fejl! Henvisningskilde ikke fundet*. page 31.

#### 6.2.3.2 Overview and examination of records

During the pre-project, we prepared a list of records found in the collection.

The objective of this registration was to select a representative sample of the records with a view to performing a more in-depth analysis of them and to doing a pilot migration.

When the registration was completed, the older digital records format and structure needed to be described in a standardised, machine-readable manner, to be able to perform machine migrations (automated) to the preservation format. With this in mind, we made a rudimentary registration of the records in which we included the main structures and the forms:
Table 9 - Main variations in structure and format of the data before migration

<table>
<thead>
<tr>
<th>Locating, returning</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code value number</td>
<td>303,980</td>
</tr>
<tr>
<td>Field number</td>
<td>124,100</td>
</tr>
</tbody>
</table>

Structure modification

<table>
<thead>
<tr>
<th></th>
<th>fixed</th>
<th>variable</th>
<th>Not examined</th>
<th>unknown</th>
<th>admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>File structure (number)</td>
<td>192</td>
<td>24</td>
<td>0</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>File structure (per submission in %)</td>
<td>82.5</td>
<td>10.3</td>
<td>0.0</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>Record structure (number)</td>
<td>157</td>
<td>60</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Record structure (per submission in %)</td>
<td>67.4</td>
<td>25.8</td>
<td>4.7</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Structure complexity (number)</td>
<td>168</td>
<td>39</td>
<td>16</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Structure complexity</td>
<td>350</td>
<td>82</td>
<td>32</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Number of record types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Migrating of packed fields

<table>
<thead>
<tr>
<th>Total</th>
<th>unknown</th>
<th>admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. number of packed fields</td>
<td>19,875</td>
<td></td>
</tr>
</tbody>
</table>

Date formats

<table>
<thead>
<tr>
<th>Total</th>
<th>unknown</th>
<th>admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1. number of record types</td>
<td>2,688</td>
<td></td>
</tr>
</tbody>
</table>

The task established the template for the digital description that would later be used in the project for documenting the records' original file structure. It was also a precursor for machine migration, since the migration programmes would be able to migrate records based on the machine-readable description of the records' structure and format.

6.2.3.3 Migrating a representative segment of records, including system development

The idea was that a prototype with central functionalities should perform a pilot migration with a view to estimating time used, errors in records, etc.

The first action point was to determine what documentation was necessary. The documentation should cover several aims, e.g. both to document that the migration was performed and to make possible future identification of the details of each migration, since machine reversibility was not the aim.

The development of the migration tools and conducting the pilot migration never took place. Instead, few selected records were analysed in depth, and by doing so, we were able to get an impression of which functionalities the migration tool should include and which steps the migration process should consist of.

The migration process was outlined as follows:
6.2.3.4  Resources

We did not divide the pre-project into several time registration posts. Neither have we performed a separate registration of what expenses were needed for e.g. external consultants and equipment.

We had budgeted with a time usage of ca. 2,430 hours, which is equivalent to 1.87 man-years. The factual time usage was 2,514 hours. The difference between the two numbers was negligible, when considering the scope of the project (84 hours, 0.06 man-years or 3.46 %).
6.3 Project Planning

6.3.1 Introduction and Evaluation Object

The following part will analyse the project planning with focus on comparing the desired outcomes with the actual results.

6.3.2 Essential Conclusions

- The project planning was based on a pre-project, which meant that many risks in the main project were eliminated.

- The project’s main objective was achieved since all records were migrated – almost – within the designated period, salary and running costs. The objective of making future migration easier has not been entirely achieved since a large number of records - albeit not in data quantity - deviated from the required standard. This was not due to poor migrations, but because the records were initially more defective than expected and we prioritised keeping the defective data instead of adjusting them and document the adjustments.

- The project was managed well by a steering group who were competent and had the necessary decision making power. The project had a thorough planning with extensive phases, time plans, work plans, time norms and quality norms (migration principles). The project manager provided weekly/monthly work plans to the project team and followed up on the performance of each task. The running expenses and the salaries were carefully registered, including mutual accounting when borrowing or lending out staff members to perform tasks, outside the scope of the project, for other parts of the organisation.

- The time-keeping was good, albeit not as detailed as we would have liked. A more detailed time-keeping per task/per record would have helped detect bottle-neck situations (and would have made the evaluation easier).

- While the project was running, we changed plans in several areas, the pertinent ones were:
  - DKK 350,000 for running costs was transferred to salaries. The excess amount was due to the smaller than expected expenses; moreover, with lack of additional funding for operation costs, it was easier to get more staff than to acquire new hardware.

The digitisation of film was abandoned for economic and technical chapter 6.12,
o Audio/Visual Records, p. 112.
o Migration of unique records was postponed in order to be able to improve on the
o Transfer to preservation media was postponed until the end of the process, which
made it necessary to make backup copies.

6.3.3 Description

6.3.3.1 Planning Basis

The planning of the FSC project was based on estimates.
One estimate was done in 2004 when the project applied for financing via the Performance Contract 2005-2008 with the Ministry of Culture. The second, and more accurate one, was done during the pre-project in 2005

Signing the Performance Contract

In 2004, we were asked to enter into a Performance Contract for 2005-2008 with the Ministry of Culture. The Performance Contract included much of The Danish National Archives’ resources including those for annual recurring tasks. Hence, it was crucial to define the FSC project in the scope of a Performance Contract with required resources. In 2004, it was estimated what resources the project would need and suggested a total of 30 man-years would be necessary to complete the project.

Pre-project

Part of the Performance Contract required that the project should be ready by end of October 2005, and on this basis, the performance requirements for the project had to be renegotiated. The pre-project is analysed in another chapter.

6.3.3.2 Preparation of the Project Plan

The project plan was designed based on an adjusted Prince2 model. The project plan included a detailed registration of the consumed resources and salaries, including outsourcing staff to other parts of the organisation. The time-keeping tool – or the use of such – was not sufficient.

6.3.3.3 Performance Requirements in the Performance Contract the Ministry of Culture

When planning the project, it was a requirement that we adhered to the Performance Requirements that were the premise for receiving the funds. The requirements were made available in annex 4 of the Performance Contract and are general requirements, cf. chapter 6.3.3.11, ”Annex 4 Performance Requirements to the Format and Structure Conversion Project”, page 45.

The actions are listed, the timeframes are provided, the quality of the end-product is described and the indirect productivity is stated.

Activities

- Restructuring of records to a new preservation standard
- Preparing to scan and scanning
- Registering information packages in DAISY
- Development of a digital description format
- Development of a migration application and ancillary programme modules

---

42 Cf. annex 25 - *FSC project’s budget*.
• Further development of test, CD-burning and registration systems
  o Migration
  o Migration of unique records
  o Migration of Circular 2000 records
  o Migration of IP records

**Productivity**
Not stated.

**Quality**
The preservation standard as described in the Executive Order no. 342 of 11 March 2004.

### 6.3.3.4 Project Management

Based on an adjusted Prince2 model, a project management group was named with a steering group and a project group including the persons most involved in the project, and led by a project manager.

**Steering group**
The Chairman of the steering group was the Preservation and Disposal Department’s Archive Manager, who thus was both a client and a supplier to the project.

A common steering group for the FSC and TGP (accessibility) projects was chosen because of the dependence of the TGP project on the FSC project.

**Project group**
The project manager handled the leadership role of the project group, which included many staff members employed to perform the project.

**Reporting**
It was decided to prepare the following reports:

- Monthly deviation reports to the steering group.
- Half-year reports to The Danish National Archives’ Management
- Annual control reports to The Ministry of Culture

The reports to The Danish National Archives’ Management and The Ministry of Culture should include a) results achieved and b) deviations.

**End Report**
At the completion of the project end of year 2008, a final report was produced that detailed the project’s progress regarding:

- Results achieved
- Possible results not achieved and why they were not achieved
- Submission of financial statements

### 6.3.3.5 Result

The project defined its primary result as follows:
• All digital records should be brought to the same preservation standard, which entailed that all paper documentation was scanned and placed in digital form along with appertaining information package.
• The Danish National Archives’ analogue audio/visual preservation-worthy records should be preserved (by digitisation).
• All digital records should be registered in DAISY, by which registrations would be made available for The Danish National Archives’ internet users.

6.3.3.6 Measurability

When planning the project, we were very aware of the necessity of establishing measurable actions throughout the project phases:

1. The first phases concerned the re-structuring of the records and the scanning of paper documentation.
2. The next stage was the rest of the migration, which was measurable. The prerequisite was the development of appropriate measurement tools in the preservation database DAISY. All the mentioned tasks were measurable; and migration progression was measurable.
3. It was difficult to quantify the on-going expansion of the migration applications – which was necessary for the new structures and formats. This part, however, only required a minor portion of resources needed for the entire project.

Likewise, the yields achieved from the project management and the purchase of software/hardware was not measurable, albeit an integral part of completing the project.

6.3.3.7 Fulfiling the requirements of the results contract – generally described

In general, the requirements of the results contract, which were essential aims of the project, were fulfilled. The following provides a short description:

Restructuring the records to new preservation structures
The goal was reached in 2006. The time consumed was more than expected and was impacted by the job change of the employee responsible for the project.

Scanning preparation and scanning
The goal was reached in 2006.

Registering of information packages in DAISY
The goal was reached in 2006, but some of the registrations were first ready in 2008.

Development of digital description format
The goal was reached in 2006. The format had to be reviewed several times more than expected.

Development of migration application and ancillary programme modules
The goal was reached in 2006. However, the documentation of the central migration programme, KonvOld, was first reached in 2007. Likewise, KonvOld required continual upgrading. This was needed to handle new structures, to deal with errors and to optimize speed.
Further development of test, CD-burning and registration systems
The goal was almost reached in 2006. However, the development of DEA (burning tool) and TEA (test tool) to ensure the conducted TEA test prior to the transfer to the preservation media, including documentation, was not completed. We chose instead to transfer the preservation media at the end of the project. This meant that the migrated records were temporarily stored on RAIDs and security copied. At the end of the project, all records were finally stored on their designated preservation media.

Migration
Migration, including the creation of information packages and tests, had as a goal to migrate 80% of the collection of information packages in 2007. We needed to migrate 1,291 information packages all in all. The plan to fully migrate 80% of the information packages was because the project should have completed the migration of all records by mid 2008, when many of the tenures of the convertors (IT staff) would expire. We did not reach that goal. By the end of 2007, we had migrated 56% of the information packages including all electronic filing systems, with an additional 5% under migration. 96.8% of all filing systems (334 out of 345) and 41.6% of registries (394 out of 946) were migrated.

We did not reach the goals for a number of reasons:
- It was not possible to substitute all the staff who resigned. This was because of the internal job rotation of staff, with the new staff needing time to be acquainted with the job functions required for migration.
- Since planning, the number of information packages grew by 40. This was because after completing the restructuring process, more submissions turned up that needed to be assigned to information packages.
- It was initially planned to wait with the migration of the more difficult records, known as the unique records, until 2008. This job was more time consuming than regular migrations. Since the unique records group needed the attention of our most experienced workforce, we decided to bring the job forward due to the risk of losing the staff that was close to the 2008 termination date of their interim contracts with the project. Furthermore, the most experienced project employee in 2008 was obliged to retire due to age (70).
- It took a long time to get the production up and running at the required speed, especially due to the need for clarification of the migration principles. The production was first up to speed in the spring of 2007. Since then productivity was satisfactory.
- The programme we chose (MS InfoPath), for typing in the digital description, showed to be slower than expected, despite the tests we did before choosing it. We made a survey of alternative programmes available on the market and concluded that there was no better alternative to InfoPath. Moreover, an investigation of the newer versions of the programme showed it to be even slower than the one we had.
- During 2007, we also considered if newer and faster PCs for the migration would speed up the production capacity, since the programmes should be able to run faster. The lack of clarification of servicing the black net, internal work load of IT staff, and the investment stop, made us conclude that buying new and specialised hardware would be too time consuming both in terms of calendar and work time.

In the autumn of 2007, the project chose instead to transfer DKK 350,000 from the expense account to salaries. It took much longer time to run production due to the lack of agreement with in-house IT regarding the servicing of black net coupled with the lack of recourses in the IT department required to perform tasks.
In the autumn of 2008, we had completed 1,126 out of 1,291 information packages, totalling 87%.

Migration of unique records
At the beginning of 2008, we had only managed to migrate 57 unique records. The plan was to have some of our best convertors to migrate the unique records from start to finish. There were not many of these unique records. However, quite a few of the simpler records were so difficult to handle that they could not be dealt with directly in the migration programme. These records required special handling, also known as pre-migration, before a normal migration could be accomplished. The very ancient and unique records were hence fewer than expected, while more of the newer, simpler ones contained too many errors and deviations to be suitable for standard migration.

Migration of Circular 2000 records
This task was started and completed in 2008. However, it showed that there were a number of errors in these records, which were not caused by migration, but by the fact that the test tool did not perform in accordance with specifications in Circular no. 4 of 2000.

Migration of IP Records
The digitisation of audio records was started in 2007 and completed in Q1 2008. The preliminary registrations were completed in 2007. The sub-supplier, Hein IT, had digitized ca. 50 % of the ca. 1,700 media in 2007. There was also a number of inaccessible media, which we did a great effort to digitise. Likewise, we decided to give special status to an array of media with museum character. Video records were digitized at Moving Media in Dublin, Ireland. All material was digitized, and transformed into information packages by MAVDA (Masse production of Information packages in Digital Records).

We decided not to digitise films, but preserve them analogue, primarily for economic reasons, as the films were not technically obsolete (material) nor technologically (format).

Productivity
No specific demands were met, as none was set. Indirectly there were demands since a given portion of records should be migrated within a certain budget and up to certain quality standard. It was estimated that the average productivity should be ca. 14 hours per information package, which was the actual time used (real time: 13,955 hours per information package).

Quality
The migration principles were designed including documentation on how the migration programme KonvOld functioned.

All records were tested with TEA and adhered to the quality standard acceptable by TEA, the tool used when receiving public records. Some records were first tested at the end of the project, making it difficult to correct errors.

We adjusted the quality level in order to adhere to the time and quantity demands and maintain data authenticity. Besides that, there were errors that were irreversible (e.g. lack of codes).

Rules of Order 342 were met with exemptions for the following errors:

- Data type DATE as 6 digit field in the format YYMMDD
- Prefixed blanks in data type STRING
- Up to 5 % duplicates in primary keys in a table
- Up to 30 % lack of foreign key values in foreign keys Tables without relations to other tables
• Lack of code name for code values was filled in with the standard text 'Code value unknown to The Danish National Archives regarding Format and Structure Conversion 2005-08'.
• Characters that could not be translated to a valid character in the ISO 8859 Latin 1, were replaced with an inverted question mark, ‘¿’.
• For primary and foreign key error, higher error limits were accepted, if they had a logical explanation for relationships between tables. If, for example, a relation to an empty table was correct, it was chosen to label the relationship, although it resulted in 100 % foreign key error.
• XML format documents with general information\(^{44}\) (migration log files, the digital description, schemas the digital description, pre-migration documentation). Executive Order 342 did not permit submissions in the XML format.

6.3.3.8 Project Management

The project management functioned very well in 2006 and 2007, aided by a project secretary. Regular project group meetings were held, in which e.g. migration principles were developed. Relations and inter-dependencies to the basic organisation and other projects were discussed in the steering group meetings.

The project manager went on maternity leave in Q1, 2008 and later resigned his position, which meant that during 2/3 of 2008 the project was managed by the chairman of the steering group. At that stage of the project, the project staff was professionally competent to carry on by themselves, but still needed management and focus on the project completion which was made difficult as the staff knew that their contracts would soon be up.

6.3.3.9 Reporting

The required reports to the Steering Committee, National Archives’ Management and the Ministry of Culture were written, but the reporting frequency decreased as the project progressed. Eventually they were included as part of the joint steering committee meetings.

Records of salary funding and equipment were kept up to date through self-developed spreadsheets because the institution's solutions were not adequate. These spreadsheets should increasingly have been accompanied by textual explanations and compared over time.

The report that was to mark the end of the project was very brief.

All salary funding was used, with surplus money from running costs partially channelled to salaries.

A minor outstanding amount between the project and the organisation was subsequently settled.

6.3.3.10 Measurability

There were not many demands to measurability and we largely succeeded in complying with them. The demands to time registration were more stringent than the norm of the organisation, but could have been even more detailed which would have made it easier to control the project.

Measuring per assignment per information package would have been an advantage to project control.

\(^{44}\) Geninfo.
### 6.3.3.11 "Annex 4 Performance Requirements to the Format and Structure Conversion Project"

Table 10 - Performance Requirements

<table>
<thead>
<tr>
<th>Performance Requirements</th>
<th>Deadline</th>
<th>Weight 2005</th>
<th>Weight 2006</th>
<th>Weight 2007</th>
<th>Weight 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restructuring of records to a new preservation structure</td>
<td>2006</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Preparing to scan and scanning</td>
<td>2007</td>
<td>-</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Registering of information packages in DAISY</td>
<td>2008</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Development of digital description format</td>
<td>2006</td>
<td>-</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Development of migration application and associated programme modules</td>
<td>2006</td>
<td>-</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Further development of test, burning and registration systems</td>
<td>2006</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Migration. For each year the following demands are applicable for records suitable for migration: 2006: 15 % 2007: 80 % 2008: 100 %</td>
<td>Annual</td>
<td>-</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Migration of unique records</td>
<td>2008</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Migration of Circular 2000 records</td>
<td>2008</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Migration of audio/visual records</td>
<td>2008</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Productivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration occurs in accordance with applicable preservation standard as described in Order no. 342 of 11 March 2004.</td>
<td>Annually</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Goals total weighting</td>
<td></td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Anticipated resource needs in million DKK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2005: 2.0  
2006: 4.3  
2007: 4.3  
2008: 4.3
6.4 Knowledge Building and Transfer

6.4.1 Introduction and Evaluation Object

The aim of the FSC project’s *knowledge building* was to enable the organisation to make a complete migration of the Danish National Archives’ collection of digital records. This was especially relevant for records of an older date, which constituted a knowledge - and thereby a resource – challenge.⁴⁵

There was no specific intention at the beginning of the FSC project to *transfer* knowledge.

However, from the perspective of evaluation, it would be of value to investigate if the FSC project’s initiatives were sufficient and if the knowledge accumulated could be used prospectively.

6.4.2 Essential Conclusions

- **The FSC project’s knowledge building was a success**
  - The FSC project managed to acquire specific knowledge required for the successful completion of the migration, because the necessary knowledge pertaining to older records was still available in the workforce.
- **The Danish National Archives retained partially (unnecessary) knowledge.**
  - The Danish National Archives retained only part of the acquired knowledge, because many of the project’s staff was temporary employed and would leave at the end of the project. Furthermore, resources were used to educate and train staff members within a highly specialized area. When their contracts elapsed, The Danish National Archives lost its accumulated knowledge.
  - That knowledge is no longer regarded as necessary, because the FSC project reached its end goals, which was the migration of the entire collection of records, and the older records structure and formats no longer exist in The Danish National Archives’ collection. The probability of receiving records based on the old structure and formats is considered rather unlikely, with the exception of certain private records.
  - Other competencies, e.g. pertaining to the migration process itself and to obtaining quotes from vendors, were lost, apart from what is sporadically available in the form of written documentation.
- **Sensible use of resources**
  - Hours spent on knowledge building complied with budget, since ca. 7 % fewer hours was used.
  - The knowledge-building portion the FSC project’s total number of hours used constituted 4 %.
- **Had there been better planning at the beginning of the project, it would have been possible to transfer knowledge more systematically and have more of it in writing.**
  - In-house. In total many resources were used to impart knowledge, but it would have been nice with much more well-documented facts about the digital collection, including descriptions of their state (by comprehensive quality check of both data and documentation), registering in the archive database and sufficient information on the migration process itself.

---

⁴⁵ Note that the following mainly focuses on the IT specialist competencies needed by the project, as it was those competencies that were assumed to require most resources and – perhaps more importantly – belonged to the critical mass required to achieve the aims of the project.
A project akin to the FSC project, both concerning the scope, thoroughness, and novelty, ought to be made available for all. For this specific reason, this report is translated into English and will be published on the internet.

### 6.4.3 Description

The FSC project identified the need for employee profiles to cover five different areas of expertise:

- Project manager
- Archivist
- System developer (analysis of records, development of tools)
- System assistant (convertor)
- Student worker

These profiles were responsible for all the FSCs tasks. The Danish National Archives had already almost all the profile types, but had special need of conveying the knowledge about the formats and structures in which the older records were described.

From the onset, it was understood that in certain instances it would be more cost-effective to acquire external knowledge and staff in the form of consultants. This was especially deemed prudent for the scanning of paper documentation and the migration of audio/visual records.

#### 6.4.3.1 Knowledge requirements - IT professional knowledge about older records

Before the project start, the organisation already had a group of key persons who had keen knowledge about digital preservation, and specifically knowledge about the collections that were the object of the forthcoming migration. These staff members formed a strong team for the project planning and tool development tasks.

Furthermore, one staff member had a unique knowledge about the older records. The employee had worked in the organisation since 1973, and the thorough knowledge about the records made it possible both to map out the records and to specify what additional staff competencies would be needed in order to complete the project.

The main challenge for the FSC project was to procure the necessary IT professional knowledge pertaining to the older records. This knowledge would be used not only for analyzing the records, but also for handling and migrating them into the new preservation standard.

Basically, the FSC project had three requirements to be met: a retrospective, a current and prospective. The biggest challenge of the three was the first:

1. It was essential to procure knowledge on the records, dating back to the period 1970-‘90s, which included structures and formats from that time: e.g. older hierarchical and relational databases, COBOL, code pages (EBCIDIC versions), packed fields, e.g. COMP-3.

2. A requirement for knowledge of the tools that The Danish National Archives would use for the migration, including the programming language Borland Delphi and The Danish National Archives’ basic infrastructure (preservation databases, etc.)

3. There was a need for an understanding of the formats and structures which were described in the new preservation standard, and which the records should be migrated to, including SQL, XML, code page (Latin 1) and newer relational databases.
The combination of the existing knowledge with the knowledge acquired through hiring new staff worked as planned. During most of the project, the IT professional team included eight system assistants (convertors) and six system developers.

6.4.3.2 Competency Development

The two most important factors in the IT professional competency development in the FSC project were the introduction programme and peer-to-peer training, respectively.

Both aspects focused on learning how to use and maintain the help, registration and migration tools as well as Delphi.

This training constituted just over 70% (1,114 out of 1,583 hours) of the total expenses used for competency development.

The remaining hours were used on knowledge building including attending Delphi and XML courses.

<table>
<thead>
<tr>
<th>Competency development</th>
<th>Time used</th>
<th>Man-years</th>
<th>Proportion %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1,114</td>
<td>0.86</td>
<td>57.9</td>
</tr>
<tr>
<td>Knowledge building</td>
<td>469</td>
<td>0.36</td>
<td>42.1</td>
</tr>
<tr>
<td>Total</td>
<td>1,583</td>
<td>1.22</td>
<td>100</td>
</tr>
</tbody>
</table>

All in all the competency development used just under 4% of the FSC project’s total hours.

The project manager had budget using 1,700 hours. The actual time used was 1,583 hours, or ca. 7% fewer hours than assumed.

6.4.3.3 Quotations from External vendors

During the process of hiring external consultants to do jobs such as scanning of paper production and migration of audio/visual records, it should be mentioned\(^{46}\) that the process itself of getting quotes from the external vendors resulted in knowledge building, regarding:

- Getting quotations (surveying the market, assessing quotation, preparing contracts)
- Preparing quality criteria for scanning and migrating of audio/visual records\(^ {47}\)

6.4.3.4 Prospective use of knowledge and competencies from the FSC project

\(^{46}\) Both aspects are described in more detail in another part of this document, c.f. chapter 6.7,

Preparation of Documentation, p. 58 and chapter 6.12,

\(^{47}\) Annex 10 – Decision memo on quality of scanning and Annex 11 – Decision memo on changes to audio format.
At the end of the FSC project, three key events occurred:

- **Interim employment contracts expired**
  - The contracts of most of the FSC projects staff expired.
  - A few staff members were permanently employed.

- **Knowledge transfer**
  - Transfer of business. A number of meetings were held with the objective of transferring lessons learned during the FSC project.
  - Knowledge sharing. Much focus was put on sharing of knowledge especially with other projects that could learn and benefit for the FSC project.

- **Evaluation**
  - The performance contract targets for 2009-12 stated that the project would be self-evaluated in an evaluation report.

**Terms of Employment**

Much of the knowledge necessary for the completion of the FSC project disappeared with the staff that left when their contracts expired. The FSC project manager resigned shortly before the project was completed. Moreover, the employee who had worked for The Danish National Archives since 1973, and who had been a central figure in the project, retired.

Two of the system developers were permanently employed. It must be reiterated that not all the staff members of the FSC project were temporary staff. Furthermore, it is important to note that the knowledge essential for the project, e.g., knowledge of the older records’ formats and structures, was no longer needed, since the originals were now migrated. The data are still available, stored in the old formats and structures (A-editions).

A great deal of the knowledge accumulated in the period 2005 to 2008 – and which is still nascent – still exists at The Danish National Archives. Nevertheless, specifically knowledge areas seem to have been hit the hardest:

1. The employee, who worked with the audio/visual records, was not given permanent employment when the contract expired, and hence valuable knowledge was lost here too. The Danish National Archives still has an employee working outside the scope of FSC project who had assisted the expert mentioned in the audio/visual area.
2. All staff who were employed as system assistants, and who had worked with the migration process (operate hardware and software, deal with errors and creating logs, etc.), were all temporarily employed and did not get the option of extending their contracts. We could not extend their employment with the project because we received lower appropriations in the following Performance Contract period.

**Knowledge transfer**

The Danish National Archives’ section for Digital Archiving, (DIARK), which was responsible for the Archives’ digital preservation, took over the migrated data. To achieve this, a number of meetings were held to initiate the transfer and a memo, Transferring the FSC tasks to DIARK, was composed.

Another of DIARK’s tasks at that time was to devise a new preservation standard, the successor of Executive Order 342, namely Executive Order 1007. The knowledge FSC had about that period’s

---

49 Cf. Demands revision project, 2007-10, the main product was the preservation standard, Executive Order no. 1007 on information packages, see annex 2.
preservation standard, Executive Order 342, was highly valuable to DIARK. This knowledge was included in a document\textsuperscript{50}.

Concurrently with the FSC project, another project, the Accessibility Project (TGP), was developed with a view to ensuring accessibility to the migrated data. The two projects had an ongoing exchange of knowledge (e.g. both project managers took part in a coordination group, which also included a specialist consultant and the chief consultant from DIARK, along with the archives manager form the Preservation and Disposal Department, the section manager and a specialist consultant from DIARK). Moreover, at the end of the FSC project, several meetings and theme days were held to hand-over the project.

The objective of these meetings was to inform about the project. For example, a symposium for specialists within the field was held in which the knowledge accumulated in the FSC project was made available to all\textsuperscript{51}. In-house meetings and theme days were also held resulting in written and verbal transfer of knowledge.

**Evaluation**
The current self-evaluation report represents the Performance Contract’s self-evaluation requirement for the FSC project.

**Conclusion**
Overall, much time was spent on transferring the knowledge gained in the FSC project to key stakeholders.

The documents prepared for the process of transfer of data focused mainly on which outstanding issues remained at the end of the FSC projects.

It would have been nice to have had a bit more well-documented facts about the digital collection, including their condition (by comprehensive quality check of both data and documentation), the registers in the preservation databases and comprehensive information on the migration process itself.

A project of FSCs character, both in scope, thoroughness and novelty should be published. This report should, hence, be translated into English and published on the Internet.

\textsuperscript{50} Annex 13 – FSCs contribution to demands revision, 2008.

6.5 Defining the Migration Method

6.5.1 Introduction and Evaluation Object

Defining the migration method mainly included the following tasks:

- Defining requirements to migration tools, periods A and B records
- Support of manual work – work-flow descriptions
- Requirements to support of manual work – migration principles

The work required a great deal of written communication. The documents produced showed to be of great value.

6.5.2 Essential Conclusions

The documents describing the workflow process and the migration principles turned out to be invaluable for the daily work and ensured a high quality in the handling of the records.

- It was not possible to completely define the migration process beforehand, since the feasibility study did not sufficiently document the nature of the records. The three documents described in this chapter which formed the basis for the method, were, therefore, ”dynamic” documents that illustrated a process in constant development.
- This created an acute need for knowledge sharing, which was formalised during the weekly meetings and in the collaboration forums in which the convertors worked in pairs thus ensuring the production of up-to-date, user-friendly and easily accessible documentation of method.

6.5.3 Description

6.5.3.1 Establishing the requirements to the migration tools, period A and B records

The intention was to perform the migration as a partial machine process supported by an IT tool for the machine production of older digital records in accordance with the preservation format defined by Executive Order no. 342 of 11 March 2004.

The programme was to perform only a single information package at a time based on the information in the digital description of each specific information package. The digital description would be keyed-in manually. An automated technical control would be performed with supplementary visual control of the digital description prior to migration.

The migration should result in an information package in the 2004 format including:

- Metadata (a new XML file)
- Migrated data
- Scanned (digitised) archival documentation
- Log files with statistics and deviation reporting
- Help tables

---

54 Annex 20c – Conversion principles for e-records version 50.
It was important that we would be able to:

- Preserve the authentic content of the records whenever possible.
- Critically review documentation (e.g. correcting spelling mistakes or structural errors)

We had planned a simple (loose) integration with the preservation database, MARY, easing the programme start with pre-determined file names. We also used the XML form-editing tool, InfoPath, to provide support for an interactive control /error correction cycle of the digital description.

6.5.3.2 Determining the requirements to the migration tool, period C Records:

The migration should be limited to one machine improvement of the structure of the records received in accordance with Circular no. 4 of 14th January 2000.

A new programme was written for this task.

6.5.3.3 Support of manual work – workflow description

The manual work related to the migration was supported by comprehensive descriptions of workflow procedures.

Workflow descriptions were divided into 2 categories:

1. Workflows used daily by all convertors

   - The preparation of digital descriptions was one of the most important tools used in relation to the migration. The tool provided a detailed description of how a migration should be performed. Along with 6 pages of description, the narrative was regularly expanded with 10 pages of annexes describing the following phases:
     - Validation: which tools could best be applied.
     - Migration, trial migration and final migration, along with error searches with KonvOld

   - Introducing new data type and code pages - described the process of introducing new data types and code pages which were part of the user demands and the resulting work needed to be done in order to implement these changes. Registering as user demand, implementing in KonvOld and InfoPath forms and description in the "Migration Principles".

   - Find information for SKABER.TAB and ARKVER.TAB in DAISY - a detailed description of how to find information, which would be written in SKABER.TAB and ARKVER.TAB in the digital description. This applied to e.g. Submission type, Archive Date, Creator Date, System name.

   - Decision not to migrate an information package - describes the procedure for the person making the decision when an information package is presumed unsuitable for migration.

55 See annex 17 – Detailed demands specification and annex 18 – System specification.
56 Annex 19 – Example of workflow description – the Final conversion and copying of archival to conversion storage.
57 Definitions of various file names in the following two lists can be found in chapter 7, Fejl! Henvisningskilde ikke fundet, page 117.
58 The migration application KonvOld was updated with new functionalities as more e.g. new codepages were found in records.
59 For all idiosyncratic filenames, cf. chapter 7, Glossary, p. 124
- **Transfer records for handling as unique records** - if it could not be described digitally and processed by KonvOld

- **Move records to quality assurance folder** - rendered migrated records accessible to quality control procedures. Each time the work with an information package's digital description, validation, trial migration and TEA-tests was completed, the final migration data were copied with equivalent TEA-logs to the Quality Control folder which made the data accessible for the quality controller.

- **Quality control of digital descriptions and migrations** - on how we controlled the results of a migration in compliance with the applicable Executive Order and migration principles

- **Test of migrated records with TEA3** - the workflow included the points of test required and not the use of TEA3 itself. The primary test items were: ArkivView, MetadataView, TableView and DocumentView. TEA3 was used for both the migration and the subsequent quality control.

- **Final migration and copying of records to the migration storage** - described how migrated and quality controlled records should be copied to the migration storage, by using MARY, and by getting a status update via DAISY; depending on the type of error detected during the migration process.

- **Update product in DAISY** - described how the updating of status on the individual archives should be performed, and which status codes should be used. The registered status had a big impact on how to calculate the number of hours used and on how to monitor progress made.

2. **Used by a few persons for administration and completing the migration**

- **Scanning of new paper documents** - ensured that all paper documentation arrived at the designated information package ID. This was especially needed when additional paper documentation surfaced and was subsequently added to the case-file.

- **Re-organising** when splitting an information package - this was used when a convertor found that e.g. several archival year groups were submitted under the same information package ID. If a division into many information packages was made, it would require comprehensive registration in MARY and DAISY, besides perhaps also scanning of documents and setting up an accession document.

- **Regenerating the records in the regeneration programme, TGP02** - was used to find out if the migrated records were useable. A certain number of records were selected for regeneration.

- **The use of VNC machines for migration tasks** was intended to relieve the convertor’s local PCs - to perform time and resource demanding machine migration tasks by performing these on remotely controlled PC.

- **Migration status updating and calculation of number of hours used per information package ID** - was done every Thursday before the FSC meeting and on the first day of each month. Based on data-draws from DAISY on the status of the individual records, a summary was created showing the percentage distribution on the entire collection of records.

The table shows examples from different periods with a one-year interval:

<table>
<thead>
<tr>
<th>No. of IPs</th>
<th>Product Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 12 - Example of migration status updating - 2007**

Created on 11 December 2007
### Project Name

<table>
<thead>
<tr>
<th></th>
<th>Filing system (A)</th>
<th>Filing system (B)</th>
<th>Registry (A)</th>
<th>Registry (B)</th>
<th>Main totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waiting Digital description completed Migrated Proofed Unique records Digital description in process Main total</td>
<td>1 3 41</td>
<td>4 289 3 2 16</td>
<td>406 14 166 3 2 16</td>
<td>123 3 183 3 1 34</td>
<td>534 20 679 6 3 51</td>
</tr>
</tbody>
</table>

Table 13 - Example of migration status updating - 2008

Updated 11 December 2008

<table>
<thead>
<tr>
<th>Number of IPs</th>
<th>Project Name</th>
<th>Waiting</th>
<th>Digital description completed</th>
<th>Not to be migrated</th>
<th>Migrated</th>
<th>Proofed</th>
<th>Digital description in process</th>
<th>Main total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filing system (A)</td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>Filing system (B)</td>
<td>294</td>
<td></td>
<td>294</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>294</td>
</tr>
<tr>
<td>Registry (A)</td>
<td>5 7 2 591 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>607</td>
</tr>
<tr>
<td>Registry (B)</td>
<td>1 5 337 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>347</td>
</tr>
<tr>
<td>Main total</td>
<td>6 12 2 1267 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1293</td>
</tr>
</tbody>
</table>

6.5.3.4 Requirements for supporting manual work – migration principles

We created a work flow document that gave a step-by-step description of the migration process. We also made a reference document "Migration Principles" which detailed all relevant information regarding the method of typing in information on the records in the digital description. The document underwent comprehensive development of content and size (from version 1 with 7 pages to version 50 including 67 pages)\(^{60}\).

Each version of the document saw the addition of new information on available code pages, new data types and migration routines.

The document included information on the following topics:

- Points 1 – 5 describe the main principles for migration, preservation and archival structures, registration in DAISY and positioning of metadata.
- Point 6 is the largest and describes everything on the records including character types and structures.
- Point 7 includes information on the records documentation.

The following table illustrates point 6 and gives an understanding of how the records should look.

---

\(^{60}\) Annex 20a – Migration principles for e-records version 1; Annex 20b – Migration principles for e-records version 25; Annex 20c – Migration principles for e-records version 50.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information package ID</strong></td>
<td>Describes how the field should be completed and where to find the information.</td>
</tr>
<tr>
<td><strong>Code page</strong></td>
<td>Describes the various code pages in main and sub groups, as EBCDIC, ISO and ASCII. At the sub-level, there were many to choose from and this especially had a bearing on the use of special Danish characters ÆØÅ and other special characters which could be ascribed to the individual hardware supplier’s implementation.</td>
</tr>
<tr>
<td><strong>COLOPHON</strong></td>
<td>Included the general information to be included by the convertor. This included initials, start date of typing, and comments on the migration. This was made available for the execution of the quality control. After approval and before the final migration, irrelevant text was removed.</td>
</tr>
<tr>
<td><strong>ARKVER.TAB</strong></td>
<td>Fill in with information package ID, submission type, earlier submission versions, system name, and start / end date.</td>
</tr>
<tr>
<td><strong>SKABER.TAB</strong></td>
<td>Fill in with name of relevant public authority that submitted the data, and for each creator a start / end date is added.</td>
</tr>
<tr>
<td><strong>FILE</strong></td>
<td>This topic required close attention since the final mark-up of the information package was done here with a view to automatic migration via KonvOld. Each file needed a description of the following:</td>
</tr>
<tr>
<td></td>
<td>- Exact name of each file</td>
</tr>
<tr>
<td></td>
<td>- File type – whether the file type was fixed or variable. Fixed is used when all entries in the file have the same length. Variable length had two different representations. It was sometimes necessary to do a process of trial and error – e.g. by using Tapedump.</td>
</tr>
<tr>
<td></td>
<td>- Skip first and/or last record – in some instances, it was found that these records should not be included in the migrated information package. If we did not have this option, we would have needed to pre-migrate several more records.</td>
</tr>
<tr>
<td><strong>RECORDDEFINITI ON</strong></td>
<td>Describes migration output</td>
</tr>
<tr>
<td></td>
<td>- Record name – depending on the structure of the &quot;old&quot; file, one or more files could be created. If there were several field structures in the input file, a similar number should be created in the new files. The following rules had to be followed: names of maximum length of 8 characters; special characters cannot be used in the name; capital letters could not be used.</td>
</tr>
<tr>
<td></td>
<td>- Record length equivalent to the sum of the field lengths was indicated; for variable length, the 4 bytes that included the value of the record length should be subtracted.</td>
</tr>
<tr>
<td></td>
<td>- Record owner name – this field is only used when the structure is hierarchical and should be omitted for relational structures.</td>
</tr>
<tr>
<td></td>
<td>- Table info – filled in with a description of the table – infinite text length possible.</td>
</tr>
<tr>
<td></td>
<td>- Table Comments – the field can be filled in with comments crucial for clarification. It can also be used as an internal aid for the quality assurance of the migration. The information was saved in the</td>
</tr>
</tbody>
</table>
### FIELD

The individual fields in the table were described with name, data type, start position, length, field information, code definition, etc.

- **Field name** – the name of the field was described. There is no limit to length; both capitals and small letters could be used. However, special characters (incl. ÅÆØÅ) or blanks/hyphens could not be used. If the field did not include data content, the following names were used: filler, filler_1, filler_2, etc.

- **Data type** – indicated the migration routine used to translate the data content to the data types allowed in the migrated table. There were many possibilities, but the following options were: NUM, REAL, EXP, STRING, DATE, TIME and TIMESTAMP. Each of these was available in countless variants depending on whether the field content was a left or right aligned number of decimals, prefix, unique values in the string field or packed fields.

- **SP (Start Position)** – denotes the start position.

- **L (length)** – the original length prior to migration

- **Field info** – used to describe a field. The text could be of unlimited length and use any ISO characters.

- **G (Grouped field)** – marked if the field was a group field. Group fields were not part of the migration, as the submission demands did not support grouping.

- **M (Included)** – the field could be marked as not included in the migration. De-selecting a field should be done after careful consideration.

- **Code definition** – documentation or data content could indicate if the field included codes that should be described with code value and appertaining explanation.

### PRIMARY KEY

All tables needed a primary key and all entries in the file should have a unique key value. A primary key should be assessed on usability and possible error percentages. If none of the fields in the table were suitable for use as a primary key, an auto generated primary key should be added.

### FOREIGN KEY

Used to define relations between two tables in the same information package, in which one table included a primary key and the other a foreign key. A foreign key could only refer to a primary key in another table. Error rates should normally be under 30 %, and the key’s authenticity should be validated.

### HIERARCHICAL STRUCTURE

Not allowed, cf. Executive Order no. 342, and hence all hierarchical instances were to be migrated to relational structures. All files including entries with several record types and/or record lengths should be split.

- **Haverecordid** – must always be defined when a table is split into several record types. The description was done with the aid of three fields:
  - **Record owner name** – filled in with the name of the record owner, but was left blank if the record did not have an owner.
- Record ID field – was given as the field that shows the actual record type.
- Value – the equivalent value was given.

<table>
<thead>
<tr>
<th>CANDIDATE KEY</th>
<th>Cf. Executive Order no. 342, it is not permitted to use candidate keys (alternative keys). If a candidate key was already described in a file, it could be demarcated by redefining the file with a new record description and by designating a candidate key as a primary key.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL- search paths</td>
<td>There were SQL search paths in the information packages that were to be migrated. SQL search paths were not generated, as this would constitute accessibility.</td>
</tr>
</tbody>
</table>
6.6 Preparation of Data

6.6.1 Introduction and Evaluation object

Data were prepared for migration. The task included the following:

- **Restructuring of files**
  - By restructuring, we mean organising the data files and documents in the respective information packages.
  - Files belonging to an information package were physically placed in an independent folder on the migration raid, so as each folder could establish input for the migration tools.

- **Registering information about the information packages in databases MARY and DAISY**
  - DAISY contains registration data on all records including the digital ones which can be viewed by the public.
  - MARY contains more detailed and technical information about the information packages which were intended for internal use and needed for e.g. supporting the migration process.

- **Transfer to archival storage.**
  - Information packages from periods A, B and C were prepared in their A-edition\(^{61}\) for transfer to the archival storage.

The evaluation focuses on results and resource consumption.

6.6.2 Essential Conclusions

- **The quality of the data source was often poor, which is reflected to a certain degree in the output data.** Some of the records prepared for migration lacked primary keys, foreign keys, code descriptions, etc., which at times was impossible to remedy and resulted in the necessity of accepting (make exemptions for) a number of errors in the migrated information packages.

- **The result of the preparation was good**
  - Data were ready for migration when tasks were completed. Data were restructured and prepared for the migration phase, registered in MARY and DAISY and ready in edition A for transfer to archival storage.
  - The task of restructuring and registering records in MARY and DAISY was time consuming, manual tasks requiring much concentration both at the time of performance and at the quality control phase.
  - The quality of task performance was good. Few deficiencies and errors were discovered in the migration phase.

- **Use of resources**
  - The task took just under 1½ man-years
  - Circa 85 % went to the restructuring part
  - The task was over budgeted by almost 30 %

6.6.3 Description

\(^{61}\) The term "edition" was used to keep track of versions. The ‘A-edition’ was the original (input data for migration), while the ‘B-edition’ was the migrated information packages (output data) which complied with the specifications in the preservation standard, Executive Order no. 342.
Data basis
The data basis comprised data from the 3 periods: repackaged files from period A and files from periods B and C were placed in a common folder on the raid.

Information package numbers were allocated based on the following rules:

<table>
<thead>
<tr>
<th>Period</th>
<th>Allocated</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>00000000-00000499</td>
</tr>
<tr>
<td></td>
<td>00003000-00003999</td>
</tr>
<tr>
<td>B</td>
<td>00000500-00002999</td>
</tr>
<tr>
<td>C</td>
<td>00010000-</td>
</tr>
</tbody>
</table>

The submission list\(^{62}\) from paper documentation ensured allocation of the respective files to a specific information package.

Restructuring and registration in MARY
The workflow for each of the records for periods A and B included the following steps:

- Locating a submission list
- Allocation of information package ID
- Briefing via a MARY correction document that an information package ID should be created in MARY.
- Creating an information package folder on the migration raid besides copying the associated files in the folder.
- For period B information packages an additional division of files into the sub-folders corresponding to Media IDs (size of up to 700 MB)
- Copying of 'recovery' processed files from period A.
- Stating on common lists whether the specific information package ID was ready for digital description.
- Comparing files (content and size) after copying using the Beyond Compare programme

Restructuring of period C records was done with the aid of a programme, which copied submissions from the source destination into new folders while concurrently dividing the submissions by volume number, name and media ID name.

The restructuring was completed with an MD5 test to ensure that the copying did not cause loss of data.

Registration in DAISY
Registration of records from A and B period in DAISY was done after the restructuring of the records and after updating in MARY.

The registration was based on an updated data extraction from MARY, which included the following fields\(^{63}\):

- System name

\(^{62}\) A submission list is a list of records affected by a single submission.

\(^{63}\) Concepts are defined in chapter 7, Glossary, page 117.
The approach to the records in DAISY was the heuristic unit. The information packages in the heuristic unit were created and described with information package ID, dating and type.

6.6.3.1.1 **End result and quality of end product**

**Result**
Collection of information packages from A, B and C periods:

- Were restructured and prepared for the migration phase
- Were registered in MARY and DAISY
- Were ready in A-edition for transfer to the archival storage

**Quality**
The process of restructuring and registering records in MARY and DAISY was time consuming, manual work that required much concentration both during performance and at the quality control stage.
Apart from the automated process of controlling the quality of the copying, all other functions were performed manually.
Defects and errors were first discovered in the migration phase, but the volume was limited.
These included:

- Necessity of restructuring of a few records, and specifically breaking up of large submissions into several smaller ones
- Lack of coordination between MARY and DAISY data, which required making corrections in DAISY

### 6.6.4 Use of Resources

The total resource used for the above mentioned task was just under 1½ man-years (calculated in hours, unless otherwise stated):
<table>
<thead>
<tr>
<th>Item/year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>Total</th>
<th>Budget</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restructuring</td>
<td>285</td>
<td>1,107</td>
<td>13</td>
<td>3</td>
<td>1,407</td>
<td>1,760</td>
<td>353</td>
</tr>
<tr>
<td>Registering of IPs</td>
<td>-</td>
<td>221</td>
<td>27</td>
<td>-</td>
<td>248</td>
<td>571</td>
<td>323</td>
</tr>
<tr>
<td>Total hours</td>
<td>285</td>
<td>1,328</td>
<td>40</td>
<td>3</td>
<td>1,655</td>
<td>2,331</td>
<td>676</td>
</tr>
</tbody>
</table>
6.7 Preparation of Documentation

6.7.1 Introduction and Evaluation Object

Part of the FSC project was to re-create the lack of documentation and to standardise accumulation of the already existing documentation, which was more often than not, in paper-format.

The documentation was required to cover two main objectives that the current evaluation dealt with:

- Documentation of the migrated information packages
  - Creating a documentation standard
  - Collecting, creating and systematisation of documentation
  - Scanning and registering documentation
- Documentation of the migration

An example of a scanned documentation in TIFF (partly in Danish):

Figure 5 - Example of a scanned documentation in TIFF
6.7.2 Essential Conclusions

- Documentation of the migrated information packages
  - No time records exist on the specific work done for the collection and systematisation of the paper documentation. It is a general problem for the evaluator that time keeping for specific tasks was not done in the FSC project.
  - The quality of documentation was uneven and did not always sufficiently describe the structure of the records – this required a lot of resources.
  - Scanning and registering of documentation
    - The extensive and thorough work performed to get quotes from vendors and to enter into contractual agreements with external suppliers was successfully achieved.
    - Concurrently, a knowledge base was created on the method of choosing and contacting external suppliers. The fact that one of the temporary staff members of The Danish National Archives documented this knowledge in writing was an essential factor for the integrity of the documentation, since the knowledge would have been lost when the person left the job at the end of the project
    - The largest part of the costs went to the scanning preparation task (55 %), which indicates that outsourcing does not free up resources for use on internal activities related to the task.
    - The registration process was somewhat unsystematic and therefore often flawed. It was possible to remedy this during migration, but was not done because it was never stated in the workflow instructions. The consequence of this, which impeded machine handling, was twofold: inconsistent naming of documentation, which hindered machine handling of future migrations. The second consequence was connecting documents to the wrong records, which was never tested.

- Documentation of migration
  - The documentation prepared – especially via migration logs – is extremely useful for evaluation and for future migrations. This is because it is possible to do an automated extract of a comprehensive and informative data source while making it possible to map the collection. It is important here to note that the time registration (e.g. per information package) in the FSC project was not sufficiently detailed to be able to perform a comprehensive evaluation; moreover the C records’ migration log was not suitable for analysis.

6.7.3 Description

6.7.3.1 Documentation Standard

Cf. chapter 6.1.2, Preservation Standard and Documentation Standard, page 33, a documentation standard was already devised in compliance with the requirements in the preservation standard applicable at that time.

The evaluation will hence only provide a cursory description of the requirements.

The general requirements to the documentation standard were stated as follows:

- Processing must not give rise to loss of information.
- The documentation should be as easy to navigate as it would be if it were paper-based.
• Processing must comply with the recommendation applicable for the information package stated in Executive Order no. 342

The specific requirements can be divided into two categories: technical and archive specialist.

The technical standard required scanning to the bitmap format TIFF, version 6.0 baseline, cf. Executive Order no. 342.

The archive specialist standard focused on the possibility of using the data in the digital records, i.e. the intended outcome of the preservation. Therefore, the processing of documentation was not only considered from a preservation standard but also with a view to accessibility. Hence, it is not enough to be able to read the data - they must also be understood.

6.7.3.2 Collection and Systemisation of Documentation

The documentation was paper-based and placed in archive boxes in the loft of The Danish National Archives’ building. The labels indicated The Danish National Archives as the creator of the archives, and included a provenance number and a digital identification number of the IT system being documented. Sometimes, there were several archive boxes for a single system.

In certain instances, this could give rise to a bit of confusion requiring – resource demanding – systemisation.

After the systematisation was completed, the documentation was grouped as follows:

• **Documentation registration** prepared by The Danish National Archives included the names and numbers of the documents in the archive boxes.
• **Submission list** prepared by The Danish National Archives and included:
  • Information on the archive creator, the system name, submission time and sometimes a dating of the data and a file directory including file name and file description.
• **File registration** prepared by the public authority or the supplier. These descriptions included:
  o A short description of the files content and function in the system
  o The individual field name, data type and length.
  o Statement on whether the field was a primary key or a foreign key (with indication of related tables and fields)
  o Value and significance of field content, including code listings (field information and code information)
• **Miscellaneous**. The documents that could appear in the documentation were very different in nature and dependent on the decisions made in the concrete situation at the time of submission.
  • Furthermore, a reference in the documentation was made to some XML files for all the information packages, these files are:
    • Digital description of formats and structures for the loading in a migration tool (XML file)
    • Schema for the digital description (XSD file)
    • Pre-migration log (XML file)
    • Migration log (XML file)

Apart from the migration logs described later in this chapter, the files are described elsewhere, cf. chapter 6.5,
Defining the Migration Method, page 52.

There is time registration for specific work done on collecting and systemising the paper documentation.

6.7.3.3 Scanning and Registration

Objective
The objective with the scanning was to scan the older digital records’ paper documentation so as to connect them to future information packages, which would conform to the preservation standard.

Programme development, preparing for scanning and registering
The first task was to develop a programme, which could register the description of a document and the relation between a document and an information package. This was done by expanding the preservation database MARY with a documentation module – i.e. tables that could include the description and relations as well as the addition of a keying in window for MARY’s user interface. As part of the development, a workflow description was prepared for the scanning preparation.

The development process required 301 hours and was allocated 330 hours.

The scanning preparation entailed collecting all boxes containing documentation and registering all documents in each box, with a description in the document registration module. When all the documents in the boxes were described, an information package was allocated. The final task in the preparation was the printing out of a bar code, which was placed at the front of each document. The supplier to whom the scanning was outsourced could then use the bar code when assigning a name to the scanned document. One of the first steps of the scanning preparation was to define workflow guidelines. This was done based on lessons learned from scanning of the contents of the first boxes of documentation. The registration guidelines for registration were written to ensure a uniform registration process. Based on the initial experiences with the document registration module, the registration process was expanded with a couple of extra functions that eased the keying in process.

On a whole, the scanning preparation process was on schedule and took less time than anticipated (862 hours versus the budgeted 1,160 hours).

Outsourcing
It was decided that in-house scanning of documentation was not practical for the following reasons:

- Too much time spent when using the existing inefficient equipment
- Lack of resources due to the necessity of buying professional equipment

It was, therefore, decided to outsource the scanning which necessitated requesting quotes from suppliers. For this purpose, a requirement specification and a scanning guideline were compiled to be used when selecting a supplier.

The choice of supplier was based on weighting the following criteria:

**Table 17 - Criteria for choice of supplier for the scanning of the paper documentation**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quality of the trial scanning. The scanned TIFF files were compared with the paper version, page by page, in order to ensure that all pages were scanned and legible. The TIFF files from the various suppliers were compared. General impression of how the records were handled.

<table>
<thead>
<tr>
<th>Quality</th>
<th>The degree of documented compliance with the requirement specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The suppliers’ credible timeframe for completing the task</td>
</tr>
</tbody>
</table>

Price

| Price  | The total price the supplier charged for the task. |

Based on the above criteria, suppliers were given grades on a scale from 1 to 10 as per pre-defined rules.

After selecting the supplier, the necessary agreements were entered into, e.g. Confidentiality Agreement and a Contract.

The scanning went without a hitch. The price for the entire scanning including transportation was DKK 101,000 compared to the budget of DKK 200,000.

Quality Assurance

Circa 10% (40 of the 404 boxes) of the scanned documents and boxes were randomly selected for quality assurance. The quality check was done by comparing the physical document with the scanned one, which also included checking the format. This control was performed based on a description of the quality that was used during the trial scanning. A few small errors of cosmetic nature were found. It was also controlled that all registered documents actually did exist in the scanned edition. This control revealed a two-sided document that was not scanned. This document, along with a couple of documents with the worst flaws, was sent to re-scanning. It was decided not to quality control more documents. If additional errors requiring re-scanning were detected, the project would do the re-scanning themselves. The supplier was informed in November that the job was approved and completed.

The rather comprehensive work of obtaining quotes and entering into agreements with external suppliers yielded a very satisfactory result.

A method of selecting and working with external suppliers was also drafted. This knowledge was documented in writing. The importance of this can be seen from the fact that a great part of the work was done by a temporary staff member whose contract expired at the end of the FSC project. Had we not had written documentation, this knowledge would have been lost.

The scanning required resources equivalent to 1,162 hours and DKK 103,337.50. This was a good deal less than budgeted, cf. the table below.

The table below also shows that the major part of the internal expenses was used on scanning preparation (74.12 %), while the scanning task itself was just under half the budgeted amount (48.3 % of the budgeted amount). The likely reason for this is that the paper documentation was of better quality than expected which meant that the supplier did not need to use so much time on restructuring or improving the documents.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Budget</th>
<th>Usage</th>
<th>Difference</th>
<th>Task proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours</td>
<td>Hours</td>
<td>Hours / DKK</td>
<td>%</td>
</tr>
<tr>
<td>Development of programme for</td>
<td>330</td>
<td>301</td>
<td>29</td>
<td>8.7</td>
</tr>
</tbody>
</table>
Scanning preparation, selection of supplier, quality control and re-scanning

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
<th>DKK</th>
<th>DKK</th>
<th>%</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning (supplier)</td>
<td>1,160</td>
<td>200,000</td>
<td>103,337.50</td>
<td>298</td>
<td>48.3</td>
</tr>
</tbody>
</table>

6.7.3.4 FSC documentation on the migration itself

The digital description
The first documentation created for a record was keying in the record’s structure, as presented in the paper documentation, in an XML file (the digital description). This was the prerequisite to enable the migration tool KonvOld to read and test the input data.

The digital description also referred to the applicable migration principles for the specific migration.

It was considered of vital importance that the digitised form of the structure and metadata of the originals (A-editions) be preserved for posterity as documentation for the record’s original form.\(^{64}\)

Example of a digital description (Danish):

\(^{64}\) Form here denotes structure and format only, while the original media form (tape or CD-R) is not documented here.
Migration logs
There was also a need to document which changes data underwent as a result of the pre-migration and migration processes.

The pre-migration log includes details for when records needed technical adjustments in order to machine-migrate them using the migration tool. The log documents which technical problems and solutions were found, e.g. too many black spaces in a record.

The migration log describes all the phases of the machine migration and logs the errors the information packages may have in the final format. Comments typed into the original description were automatically transferred to the migration log.

The final documentation format should follow the requirements in the Executive Order no. 342, and are already mentioned, cf. annex 8 – Preparing Documentation for Submission of e-records.

Analysis and usefulness of documentation
It is possible to analyse the data both as input and as output.

The input data can be analysed using the digital description, which only describes the structure and form of the input data prior to migration.

Output data can be made the object of more in-depth analysis for which the pre-migration log, and especially the migration log, can be used.

The evaluator developed programmes to extract data with a view to analysing the records and specific evaluations. The data foundation is solid, meaning that it is systematic, machine readable and informative. The data extractions are, however, not complete. In order to perform an exhaustive analysis of the existing collection with a view to creating a status report, it is necessary to extract more data.

The migration logs have a high functional value for both the evaluation of the migration and of future migrations. It is important to note here that the migration of the FSC project lacks a more detailed time-keeping registration (e.g. per information package) which would enable a more comprehensive evaluation; moreover the C records migration log is not suitable for analysis.

---

65 Aside from the 56 initial migrations, which did not have systematic migration logs and hence created difficulties in relation to automated extracts
6.8 Infrastructure and Hardware

6.8.1 Introduction and Evaluation Object

This chapter deals with an analysis of the infrastructure and equipment available to the project for use in the migration process. Hence, the focus is on hardware – software is only mentioned here in relation to the analysis of hardware.

6.8.2 Essential Conclusions

- The equipment used for the migration was suitable for most migration tasks, but under dimensioned for the large data quantities, which required comprehensive control of keys and codes.
- Two stronger machines were purchased for the more demanding tasks, but several more machines were needed and should have been purchased earlier in the project.
- The agreement with the IT Department was not clear on operation support.
- The speed of the archiving net (black net) was not optimal. The lack of operation support meant that the problem was neither handled nor solved.
- Infrastructure and hardware expenses throughout the project were:

<table>
<thead>
<tr>
<th>Year</th>
<th>DKK</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>4,890.00</td>
</tr>
<tr>
<td>2007</td>
<td>14,136.05</td>
</tr>
<tr>
<td>2006</td>
<td>89,152.09</td>
</tr>
<tr>
<td>2005</td>
<td>102,502.63</td>
</tr>
<tr>
<td>Hardware Total 2005-08</td>
<td>210,680.77</td>
</tr>
</tbody>
</table>

6.8.3 Description

6.8.3.1 Infrastructure

The pre-project planned to use stronger standard Wintel PCs for the main project. These would be used in the closed ‘black net’ network. Records would be temporarily placed in migration storage until they were migrated. Then they would be transferred to the preservation media in the form of optic media and magnetic media in the form of hard disk in RAID.

6.8.3.1.1 PCs

HP Intel Pentium 4 machines were bought equipped with between 1.5 and 4 GB RAM.

Each employee used his/her own machine to perform the migration, besides using several remote controlled machines of the same type. This was because records with large data quantities required so long processing times that it was worthwhile to use one or two extra machines. The convertors could not handle more than three simultaneous migration tasks.

---

66 Annex 21 – Black net.
All machines used MS XP as operating system, and a standard image was developed. MS Infopath 2003 (and later 2007) only existed in a 32-bit version and it was, therefore, not possible to get a 64-bit version which could use RAM over the 3.5 GB level, with the programme needing a lot of RAM, it could have benefited a lot from this.

When we realised that records with large data quantities showed to be very time consuming to handle, we requested more powerful machines in the form of the Intel Core 2 Duo machines, which were first introduced on the market in 2007. However, only two machines were bought because at the time the remainder of the budget of ca. DKK 350,000 was allocated to the salaries. What is more, the purchase of the machines was delayed as the IT Department did not undertake to perform the usual service responsibility. Therefore, the project team had to install the two machines. We had not expected that within three years there would be a need or necessity of exchanging the machines. Moreover, we did not have a clear agreement with the IT Department on providing operational support.

6.8.3.1.2 Migration Storage

All records that were to be migrated were accessible via a RAID of hard disks. When the records were migrated, they were subsequently stored in migration storage in the form of a RAID of hard disks. The next step was to move the records to preservation media on optical discs and hard disks in RAID.

The storage period was estimated to be for a maximum period of 14 days per information package, but it proved useful for the production to delay the completion of each individual information package.

Hence, the records were not moved to the archival storage as planned and this led to lack of space in the migration storage. It was necessary to borrow more space from another department.

The available migration storage in the form of a RAID of hard disks was insufficient due to the large amount of partially migrated records. This made it necessary to make backup copies of this migration storage. IT Department was not prepared to supply this operation support and hence the project was obliged to create its own manual backup of the portable hard disk.

6.8.3.1.3 Network

All machines and storage units were connected in the black net, which was often quite slow. The lack of operation support meant that the technical reason for the slowness (network equipment, servers, client set-up…) was never clarified nor solved.
6.9 System Development

6.9.1 Introduction and Evaluation Object

Several programmes were developed during the FSC project with the aim of handling the tasks required for migration:

- **Development of the migration tool for period A and B records**
  - Development of the digital description and KonvOld
  - Development of pre-migration tools for the preparation of the records which could not readily be handled by KonvOld
  - Development of various tools to support the migration process
- **Development of the migration tool for period C records** (Konv2000)
- **Tailoring existing tools to support the migration process** (MARY, DEA)

The tools made it possible to machine migrate the records. If a record could not be automatically migrated, it was sent for pre-migration adjustment which rendered it suitable for machine migration.

The evaluation focuses on use of resources, software performance, quality of software, deviations from original development plans and future adjustments of software.

6.9.2 Essential Conclusions

- **A thorough investigation of technologies and programmes is worthwhile doing before adopting them.** A couple of examples illustrate the point:
  - Upgrading to a new version of the programming language showed to be simple and resulted in an improvement in programme performance speed by ca. 30 %
  - The keying in tool had a week performance when used with large data quantities
- **Performance**
  - The migration application should – especially with large data quantities – be able to save interim results and use these as an end-product. For example, it was possible in KonvOld to request partial tests of data, e.g. test of codes, primary keys or selected tables. Instead of requiring the end migration to test all aspects again, it should be possible to implement log files already produced from earlier test runs. The function used to create the final information package should thus be able to use interim results that have already been created as part of the end project as opposed to processing all the data all over again.
  - The migration application should be database based concerning control of data. The central migration application was not based on a database because the input data did not have uniform data structures and was hence not compatible with a database structure. The control of the output data (especially with large quantities) could have been performed much faster if the application had been related to a database.
- **Standardising of records saves money.** It is very expensive to handle non-standardised records. The task of performing the standardisation is essential if digital preservation should be cost-effective in the long term. Below are two examples of the consequences of handling non-standardised records.

68 For a list of the entire tool collection of the FSC project, see Fejl! Henvisningskilde ikke fundet., page 17.
The system development in the FSC project took 4 man-years, of which the development of tools used to handle older records (periods A and B) represented over 80%.

175 programmes were developed to handle 167 records that had special structures and formats.

**Iterative development is not always good and should be targeted.**
- The migration application, KonvOld, was developed iteratively mainly because the records that were to be handled were very different in structure and format. The iterative development made it possible to automate the processing of 87% of the records and showed to be exceptionally well suited.
- The test application, TEA, was also subjected to iterative development and existed in several consecutive versions with varying degrees of requirement fulfilment of the accompanying preservation standard. The consequence of this was unfortunate: when received, the records were tested and approved (with errors) by several versions of the same tool, depending on when the submission took place. Whether it is at all possible to avoid multiple versions of a test tool is a question yet to be answered.

**It is crucial to have a uniform test of records before developing the migration tools.**
- It was assumed that a number of the newer records (C records) were correctly tested when submitted. The application that migrated them was therefore not completely “requirement specified” to these records, but addressed only few of the structural problems. This turned out to be a false assumption, since the records were not uniformly tested when received and the subsequent manual error correction cost 225 hours.

**When the collection is heterogeneous and includes many unknown formats and structures, it is most efficient to develop the migration application concurrently with the execution of the migration.**
- It is not possible to develop 175 programmes for highly dissimilar records in a pre-project.
- It is more efficient to develop the programmes at the same pace as the records are migrated, as this makes it possible to develop the migration tools iteratively and thus have an on-going access to knowledge of the issues/problems as they arise.

**Only few of the programmes can be used prospectively**
- The migration applications. There are no recycling options since the programmes address specific issues that are obsolete when migration is completed.
- Preservation environment. It is obvious that there is a higher degree of reuse of programmes used in the preservation environment, including e.g. registering programmes, although these can be optimised / rationalised.

**Flexibility**
- KonvOld was a flexible application, which handled many input variations. Only 167 out 1,291 records (just under 13%) needed to be pre-migrated before they were suitable as data input for KonvOld. The remainder of the data could be used directly as standard input and KonvOld handled the migration of ca. 87% of the records automatically. This was only made possible through the constant optimization and extension of KonvOld.
- The above supposes the possibility – easily – to add new programme modules for the migration of the unique structures.

**Low performance with high amounts of data due to the lack of database integration.** KonvOld's processing speed was significantly slower when testing records containing large data quantities. The respective tables, key fields and code values were downloaded to the machine’s internal memory and sorting occurred internally. This meant that the search process of records occurred sequentially.
• **Alternatives to in-house development.** The cost-benefit of outsourcing part of the development of the tools should be considered. The obvious advantage is cost savings. The disadvantages are lack of control, difficulties of integration with preservation environments, and the difficulty of finding a supplier for these highly specialised tools. An alternative to in-house development is participating in Open Planets Foundations initiatives and projects.

### 6.9.3 Description

We decided to develop most of the migration tools ourselves. There were many pros and cons to this decision. The most important views are listed below:

- **Lack of market survey.** Before deciding to do the migration ourselves, we probably did not do a thorough enough study of what was already available on the market. It is quite possible that the FSC project could have saved resources by buying tools from third parties or resorting to use open source.

- **Hybrid solutions are cumbersome and demand many resources.** Integration of third party tools is demanding. The data, which the FSC project migrated, arrived in many different structures and formats, which meant it would have been necessary to integrate several different tools in order to solve the same issues.

The programmes we developed for the project were written in the programming language Pascal by using the development tool Borland Delphi 2006. This was a deviation from the original guidelines specifying the use of Borland Delphi 7. Correcting errors that were difficult to locate, demanded, therefore, the use of the error-detecting tool in Borland Delphi 2006.

Updating to the newer version turned out to banal and resulted in a 30% increase in speed of production. As part of the error correction, the programme was also tested properly after the update, which was made permanent.

The following describes the tools separately and offers conclusions on the common traits associated with the software development.

#### 6.9.3.1 KonvOld

The migration tool KonvOld was the central application used for the migration records from periods A and B. 69

The first 2 years of development proceeded based on the lessons learned from the pre-project phase and related to a small sample of the records.

It was necessary to budget with an ongoing need for recourses to continue the development of the system. This was because it was not possible to forecast all instances of data types, data sizes and their correlation without the thorough analysis, which was first done at the time of the actual migration.

**Flexibility**

The structures and formats of the records submitted prior to 1998 were heterogeneous and required much of KonvOld that processed all the records. KonvOld was developed to be flexible and versatile. 167 out of 1,291 (just under 13%) of the records were to be pre-migrated before they were deemed suitable as data input for KonvOld. The remainder of the data could be used directly

---

69 Period A and B records were records submitted prior to 1998. See also chapter 4.2.2, The objects of the Project - Records, page 12
as standard input. Thus, KonvOld handled the automated migration of 87% of the records. This was made possible by the continuous optimisation and extension of KonvOld\(^\text{70}\).

**Performance**

KonvOld's performance was satisfactory. However, the performance was compromised in two situations:

Firstly, it was because *InfoPath* was chosen as the tool to key-in the digital descriptions. The read and edit functions were very slow when handling large records with many tables, many fields and long code lists. Hence, we resorted to the use of alternative keying-in tools, e.g. XML Spy or the word processing program, TextPad; this was not without risk\(^\text{71}\).

Secondly, the performance speed of KonvOld was significantly reduced when migrating and testing large data quantities. This was because the programme was not supported by a database. It would not have been possible to take advantage of possible database support functions when migrating because the input data had varied structures, which were not compatible with the structure of the database. Affiliation with a database would have made more sense when performing tests on the migrated data, since tables, key fields and code values were read into the machine’s internal memory. Sorting was performed internally, while the record search could only be done sequentially.

**Output Quality**

KonvOld produced information packages, which in form and structure complied with the preservation standards requirements cf. Executive Order no. 342 of 11\(^\text{th}\) March 2004.

In conjunction with the migration, we built in a certain error tolerance when forming information packages. The acceptance of the errors was hence not KonvOlds fault but due to the poor quality of the input data.

The following error types were accepted and granted exemption:

- We decided to use an inverted '?' in places that included a character that could not be identified and translated to a legal value in the code page ISO 8859 Latin 1.
- Primary key without data content (here blank fields).
- Duplicate primary keys that did not exceed 5% of the total records in the table.
- Lack of foreign key values that did not exceed 30% of all field relations
- In instances of code values without code designations, the code designation was added: "Code value unknown to The Danish National Archives concerning Format and Structure Conversion 2005-08".

All errors were thoroughly described in the migration log file.

**Functionalities**

KonvOlds primary functions were:

- **Access Microsoft InfoPath**, the tool used to key-in the digital description of the records. Converters produced the digital description as a *well formed* XML document and pre validated towards an *XML Schema*.

\(^{70}\) KonvOld was developed to version no. 42.

\(^{71}\) The risk of using other keying-in tools than the screen-controlled InfoPath, was the lack of validation of the keyed-in data and eventually the prolongation of the migration process, since the errors were not detected at the time of keying-in the data. An application specially made for the task would have solved the problem.
• **Loading and validating the digital descriptions.** The digital description was the basis for the metadata file in the final information package and included references to the migration routines of character types and data types.

• **Integration to MARY, the preservation database,** which included information about the scanned documents constituting the documentation (general information) in the final information package. MARY supplied KonvOld with names, positions and descriptions of documents.

• **Loading and presentation of data on the screen.** The function interpreted the data via the digital description; translated characters by the aid of the designated code page migration routines, and migrated data types with the aid of the respective data migration routines. The function also validated the data in the translated form. Character and data type errors were immediately presented on the screen.

• **Validation of record integrity** consisting of:
  o Control of the primary keys in all the tables
  o Control of the foreign keys and relations between the tables
  o Control of code values and code designations in the code bearing fields

• **Writing an error log file.** The log file also included all data errors either in the form of a detail list or in the form of a total in which the errors were presented per table and error type.

• **Ordering of a migration production run with an indication of the permitted errors including their types and number.** The migration could not be carried out until all errors were defined and permitted.

• **End migration with creation of the completed information package** and writing of the final migration log file.

• **Transfer of the finished information package to a migration raid** with control of the copying function by using a MD5 check.

**Deviations from original plans**

In the planning phase, it was stated that the "Migration application is a prerequisite for semi-automatic migration of records. Hence it is corollary that it should occur before the migration commences".

KonvOld was completed in its basic edition before the migration process started. During the 2 years in which the migration process took place, 42 versions of the programme were published with a constant stream of developments and improvements.

The need for the on-going adjustments of the programme was not predicted. Therefore, so as not to obstruct the migration process, the changes were often implemented and tested directly in the production environment.

**Future adjustments**

The migration tool was, with minor modifications, used for the production of information packages for revenue-funded business and in a project that focused on retrieving private digital records.

**6.9.3.2 Pre-migration tools**

Pre-migration tools here mean small programmes designed to improve the input data to KonvOlds standard input format.

Pre-migration was usually done per record or per file. Each production was completed with a programme specifically designed for this particular objective.

In all, 175 programmes were developed and adjusted for the total number of 167 records that were pre-migrated. The result of the pre-migration was 1,162 new files which substituted the old ones (total of 792) at a quality that was good enough to be used with the standard migration tool KonvOld for automated migration.
Pre-migration can be divided into 2 types of data corrections in the individual files:

- **Correcting file format:**
  - Removal of superfluous bytes from files, which were originally submitted as FB format\(^{72}\), and where the statement of the number of records in the block was incorrect.
  - Removal of superfluous bytes e.g. binary zeros and various control characters, etc.
  - Change to the records with tables from variable to fixed number of elements
  - Extension of records with variable record length to fixed record length
  - Change of values in the record identification fields in order to ensure correct identification of record type.

- **Correction of file structure:**
  - Dividing a file with multiple record types into several files each containing one record type in order to make it possible to create correct data relations (mark out primary and foreign keys) in e.g. filing systems;
  - Dividing a file with very big unmanageable record lengths into several files, concurrently with the transfer of key fields and relations
  - Collecting of several files with the same structure and logic into one file
  - Physical removal of files with identical data in one record

The programme development was based on a ’skeleton’ programme that was reused for each application.

Documentation for each operation was done by writing a standardised log file. The file was later included in the records’ collated migration documentation.

**Deviations from plans**

*It was originally planned that the handling of the unique records performed at the end of the migration phases, cf., phases 1, 4, 5, 6 in the project planning, chapter 6.3,*

\(^{72}\) VBF (Variable Block Format) and FBF (Fixed Block Format).
We later found that it was more expedient to pre and end migrate the unique records concurrently with defining and solving problems, i.e. already in the migration phase.

**Future adjustments**
The programmes we developed could be re-used during the project for the pre-migration of data with similar errors.

We do not think that the programmes can be used for future tasks.

6.9.3.3 **Miscellaneous**

We designed a number of minor tools to support the migration of records from the A and B periods. These tools can be divided into four categories:

- **Machine aid for typing in the digital description**
  - A programme that could create field descriptions in the metadata file with the aid of duplication of information and change of field names;
  - A programme to re-calculate the start position of fields in the record description after changes to the field length;

- **Analysis of the content and structure of a file for records with bad or insufficient documentation**
  - A programme to calculate the number of records with varying length in a variable record length file including a search function for a record identification field
  - A programme to show records in a file with division by fields. Data could be shown after translation from several different code pages including hexadecimal representation

- **Correction of metadata and log files in already migrated records**

- During the migration process, small errors in the migrated records were detected. The errors occurred because of mistakes in the migration tool KonvOld or because of insufficient test of output. KonvOld errors and the test tool were regularly corrected. However, instead of re-migrating the records, a machine post-migration was chosen. For this process, we designed two programmes:
  - A programme to correct errors in the line-counter in the migration log file
  - A programme to correct various minor errors in XML files

- **Analysis of data quantities (input/output)**

- During the migration phase, we regularly performed an analysis of the growth in data quantities in order to be able to monitor, and continuously ensure, disk capacity on the migration raid and on the backup disk. For this task, we wrote a small programme used to evaluate disk space requirement during the processing of new data.

**Future adjustments**
The analysis programmes have been used for receiving digital records and for revenue yielding business and acted as a substitute for the TapeDump programme.
6.9.3.4 Konv2000

The migration tool Konv2000 was the central application for the migration of the period C records, which were records submitted after year 2000. These records met the then applicable preservation standards hence enabled fully automated migration.

The programme’s task was:

- **The Search Function provided an overview containing:**
  - Information packages submitted in formats complying with Circular no. 4 of 14 January 2000, and which could thus be machine migrated.
  - Information packages submitted in formats that partially comply with Circular no. 4 of 14 January 2000, and which required manual processing.
- **Machine migration of metadata and help tables** to the format applicable in Executive Order no. 342 of 11 March 2004
- **Making a production log file**

The following table shows which structural changes the application was programmed to perform:

<table>
<thead>
<tr>
<th>File</th>
<th>Circular no. 4 of 2000</th>
<th>Executive Order 342 of 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARKVER.TAB</td>
<td>System name field with length of 50 characters</td>
<td>System name field length with 256 characters.</td>
</tr>
<tr>
<td>CREATOR.TAB</td>
<td>Creates fields with length of 50 characters</td>
<td>Creates fields with length of 256 characters.</td>
</tr>
<tr>
<td>FILE MAP.TAB</td>
<td>Specification of metadata file name is A Vid.txt or A Vid.TXT</td>
<td>Specification of metadata file name is A Vid.XML</td>
</tr>
<tr>
<td>META DATA</td>
<td>Metadata file name is A Vid.txt or A Vid.TXT</td>
<td>Metadata file name is A Vid.XML</td>
</tr>
<tr>
<td></td>
<td>HOVED::= is not stated.</td>
<td>HOVED::= must be stated cf. Executive Order</td>
</tr>
<tr>
<td></td>
<td>&lt;/information package&gt; is not stated</td>
<td>&lt;/information package&gt; must be stated</td>
</tr>
<tr>
<td></td>
<td>Sql definition ::= stated without restriction with &lt; SQLDATA&gt; and &lt;/SQLDATA&gt;.</td>
<td>sqldefinition ::= stated with restriction &lt; SQLDATA&gt; and &lt;/SQLDATA&gt;.</td>
</tr>
<tr>
<td></td>
<td>The following characters are allowed in the text:</td>
<td>The following characters are not allowed in the text:</td>
</tr>
<tr>
<td></td>
<td>&lt;</td>
<td>&lt; must be represented by &lt;</td>
</tr>
<tr>
<td></td>
<td>&gt;</td>
<td>&gt; must be represented by &gt;</td>
</tr>
<tr>
<td></td>
<td>'</td>
<td>’ must be represented by '</td>
</tr>
<tr>
<td></td>
<td>&quot;</td>
<td>&quot; must be represented by &quot;</td>
</tr>
</tbody>
</table>
Performance
A perquisite for the development of Konv2000 was that migrated records had to live up to the preservation standard applicable before the start of the FSC project. Konv2000 was developed solely to address the need for bringing the records from the preservation standard, Circular no. 4 of 14 January 2000, to the standard required for the FSC project, as described in Executive Order no. 342.

Random checks on the data were performed after completing the migration. This showed that many of the information packages contained errors. The errors were e.g.:

- Incorrect marking out of primary and foreign keys
- Use of illegal characters when naming data files
- Incorrect use of merge functions
- Geninfo files include other types of files than TIFF
- TIFF files do not comply with compression specifications

The errors existed because the Danish National Archives’ test tool did not – upon reception of many submissions – live up to the requirement specification described in the applicable preservation standard.

The errors could not be machine corrected. A consultant was hired to review the records and render them in compliance with the level required by Executive Order no. 342. This task was performed outside the scope of the FSC project.

6.9.3.5 MARY
The production of the migrated records required adjustments to the existing database system, MARY. This was done in order to ensure that the migration tools (the digital description and KonvOld) could access the database online.

MARY contained:

- **Documentation (general information) about records:**
  - information package ID
  - System name
  - Provenance number
  - E-number
  - Participant
  - ID for Heuristic Unit (HE)
  - Information package’s start and end dates
  - Creating period start and end dates
  - Filing system number
  - Other

- **Information on data files to be migrated**
- **Information on the scanned paper documents to be part of the general information about the archive**

It was decided to develop a new module in MARY that made it possible to:

- Easily find records that should be migrated
- Show the necessary information about the archive
- Show the scanned documents related to the archive
- Show the name of the files that should be migrated
• Load the tool for typing in the digital description (InfoPath)
• Load of the migration programme KonvOld

The screen shot below shows MARY’s user interface.

**Figure 7 - MARY’s user interface**

**Performance**
All records from the periods A and B were migrated aided by access to the migration tools from MARY.

Of notable value was:
- Collection and access to the scanned documents, which ensured correct, automatic routing of documents to the respective records when the final information package was created.
- Option for online OCR reading of the scanned documents when preparing the digital description

**Output Quality**
The content of documentation (*general information*) on the archive partially typed in manually in MARY.

This resulted in a lack of convergence between DAISY and MARY data and required manual control of the information in both systems as well as an assessment of which of them would be the right one to use in the digital description.

6.9.3.6 DEA
Adjustments in DEA should benefit transfer to archival storage of both the digital records received regularly from The Danish National Archives, and the records that the FSC project produced.

The development was divided into three phases:

- **Further development of DEA in order to enable the programme to be executed on a standalone PC.**
  - The objective was to be able to initiate (CD) burning jobs simultaneously on LSK1 and LSK2 instead of waiting for DEA to finish testing.

- **Phase 2 comprised the following main tasks:**
  - **Scheduling in DEA.** The objective was to initiate (schedule) burning in order to have DEA transfer and burn multiple submissions, e.g. many small submissions during the night and at the weekend.
  - **Integration of DEA with the MARY database.** The objective was the direct collection of information regarding earlier versions and data packages in MARY instead of that from the filing system. This meant that it was possible to update the status of the transfers and perform CD burning in MARY rather than in an export file.

- **The final phase comprised the following main tasks:**
  - **Initiating copying / burning via TEA.** The objective was to make it possible for TEA3, after completing the test, to request copies to the storage media and to other media via DEA.
  - **Changes to DEA regarding the transition to DVD R media and the implementation of two new robots to handle CD/DVD**

During all the development phases, we worked to ensure the optimisation and correct implementation of the MD5 test routines.

We tested if using two or three drives with our burn robots (LSK) could increase the aggregate performance regarding loading and/or MD5 test.

Later on, the MD5 test was transferred to our validation tool (TEA3), which meant that the MD5 test could be performed together with the final approval of the test, and that MD5 test results could be saved in MARY.

**Performance**

Prior to the automation of DEA, the total number of hours used to copy the restructured records was 379. The second transfer occurred almost completely automatically, since the new above-mentioned version of DEA could be used.

**Future adjustments**

DEA and TEA are under constant development since the systems are part of the active parts of The Danish National Archives’ preservation environment.

### 6.9.4 Review of the common traits and results of software development

FSCs approach to software development was:

- **Pragmatic.** The major part of the development of tools for the migration was done in-house. This was because the records and preservation standard required tailor-made tools. When financially beneficial, we bought tools, cf. Fejl! Henvisningskilde ikke fundet., page 14.

---

73 Distribution of Electronic Records – burning and distribution to archival storage of the information packages.

74 LSK1 and 2 are robots that execute DEA's orders.
Iterative. The complex nature and variety of the records demanded iterative development to e.g.:

- Correction of errors and irregularities
- Addition of new code pages and data pre-migration routines
- Improvement of production speed
- Improvement in user-friendliness of the system
- Addition of several fields in the digital description

Resources
The system development in the FSC project took circa four man-years, of which the development of tools needed for handling the older records (periods A and B) represented over 80 % of the time used.

The remainder of the resources were used for the optimisation of the existing preservation environment (the preservation database, MARY, and the burn programme DEA – just under 6 % and 10 % respectively) and for the development of the tool used to migrate the newer records (Konv2000 – just under 3 %).

<table>
<thead>
<tr>
<th>Tool</th>
<th>Year 2005</th>
<th>Year 2006</th>
<th>Year 2007</th>
<th>Year 2008</th>
<th>Main numbers (hours)</th>
<th>Man-years</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>KonvOld</td>
<td>328</td>
<td>2,99</td>
<td>754</td>
<td>343</td>
<td>3,724</td>
<td>2.87</td>
<td>57.45</td>
</tr>
<tr>
<td>Pre-migration</td>
<td>0</td>
<td>199</td>
<td>751</td>
<td>633</td>
<td>1,583</td>
<td>1.22</td>
<td>24.42</td>
</tr>
<tr>
<td>Konv2000</td>
<td>0</td>
<td>184</td>
<td>0</td>
<td>0</td>
<td>184</td>
<td>0.14</td>
<td>2.84</td>
</tr>
<tr>
<td>Misc.55</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MARY</td>
<td>0</td>
<td>358</td>
<td>20</td>
<td>0</td>
<td>378</td>
<td>0.29</td>
<td>5.83</td>
</tr>
<tr>
<td>DEA</td>
<td>26</td>
<td>545</td>
<td>30</td>
<td>12</td>
<td>613</td>
<td>0.47</td>
<td>9.46</td>
</tr>
<tr>
<td>Total</td>
<td>354</td>
<td>3,585</td>
<td>1,555</td>
<td>988</td>
<td>6,482</td>
<td>4.99</td>
<td>100</td>
</tr>
</tbody>
</table>

Just under 80 % of the development resources were used in 2006 and 2007. The reason being that 2006 was the year in which KonvOlds development was completed. 2007 was the year when KonvOld completed the first migrations resulting in the necessity of developing minor programmes for the pre-migration process.

As indicated in the table above, the development of these programmes continued with full force in 2008. All in all the 175 pre-migration programmes cost just about 1 man-year (1,583 hours) or 24.42 % of the total expenses.

It could be argued that this expenditure could have been reduced if better pre-analysis of the records’ varied types had been performed. The project’s developers maintain that it would have been:

- More effective to develop the programmes simultaneously with trying to migrate the records, as this made it possible to develop iteratively, and by doing so, have closer contact with the problems as they arise.

55 System development was done on a continuous basis. The time used cannot be specified since it is part of the total development time in the time keeping registration.
• Impossible to develop all the programmes in the pre-project phase because the records were so varied in nature.
6.10 Migration and Preparation of Metadata

6.10.1 Introduction and Evaluation Object

The migration in the FSC project is categorised by OAIS as a digital migration of the type transformation.

The records which were to be migrated were divided into three periods A, B and C. Note that the time used in the table below only addresses the processing of the migration itself (including manual production and pre-migration), and thus does not include time used on adjacent tasks e.g. system development.

Table 22 - Overview of migrated records in FSC

<table>
<thead>
<tr>
<th>Record type</th>
<th>Submitted in the period</th>
<th>Content</th>
<th>Data quantity (GB)</th>
<th>No. of files</th>
<th>No. of IPs</th>
<th>Hours</th>
<th>Time per IP</th>
<th>Time per GB</th>
<th>Time per file</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1973-1998</td>
<td>Data files on tape and paper documentation</td>
<td>5,185</td>
<td>650</td>
<td>10,836</td>
<td>16.67 t</td>
<td>2.08 t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1998-2000</td>
<td>Data files on CD R and disks with paper documentation</td>
<td>690</td>
<td>8,617</td>
<td>7,207</td>
<td>11.24 t</td>
<td>49.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2000-</td>
<td>Digital filing systems and registries</td>
<td>1,481</td>
<td>-</td>
<td>942</td>
<td>0.23 t</td>
<td>0.25 t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td></td>
<td>2,171</td>
<td>13,802</td>
<td>2,233</td>
<td>18,267</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.38</td>
<td>24.84</td>
<td>-</td>
</tr>
</tbody>
</table>

The table above clearly shows that the migration of non-standardised records (period A) took much longer time than the migration of standardised ones (period C):

- It took 70 times longer time to migrate one information package from period A compared to one from period C (16.67 hours versus 0.23 hours).
- Per GB, it took ca. 200 times as long to migrate period A and B records compared to those of period C (49.44 hours versus 0.25 hours).

Records A and B were the oldest, with A being quite heterogeneous in format and structure. Therefore, it was necessary to migrate these separately from C records whose preservation standard was quite similar to the standard set for the entire migration project.

For the same reason, the two record types are dealt with separately in the following passages.

---

76 See also Annex 30 – Overview of input and output data.
Metadata were to be uniform for all three types and conformed to the specifications in Executive Order no. 342, page 10, which complied with the standard for XML version 1.0 and the notation described with the meta-language EBNF.

6.10.1.1 Period A and B Records - Introduction

Data basis
The basis for the data migration was:

- Data files belonging to the records received in the 1973 – 1998 (period A). The files were placed in the migration raid in one folder per record with names compatible with the relevant information package numbers.
- Data files belonging to records received in the period 1998 – 2000 (period B). The files were placed in the migration raid in one sub-folder per medium ID equivalent to their medium ID and one folder per record with a name compatible with the relevant information package numbers.
- Scanned documents belonging to records from periods A and B were placed in one comprehensive folder.
- The MARY database included the following for each record:
  - The typed-in master data for each submission
  - Affiliation to the respective scanned documents of the records
- DAISY included information on records from the date of submission including information package numbers
- Filing system case files

<table>
<thead>
<tr>
<th>Data Basis</th>
<th>Filing system</th>
<th>Registry</th>
<th>No. of files</th>
<th>Data amount in GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46</td>
<td>604</td>
<td>3,109</td>
<td>171</td>
</tr>
<tr>
<td>B</td>
<td>299</td>
<td>342</td>
<td>8,078</td>
<td>419</td>
</tr>
<tr>
<td>Scanned documentation</td>
<td>-</td>
<td>-</td>
<td>24,171</td>
<td>6.71</td>
</tr>
</tbody>
</table>

Tools
The tools used to handle the data were, as a rule:

- Tools for typing in the digital description (InfoPath forms and schema) in version 1.0
- Migration programme KonvOld in version 1.0
- Integration to the MARY database from KonvOld
- TEA32 – the validation tool
- Various tools e.g. XMLSpy, TapeDump, word processing programmes.

The tools are described in detail in chapters 6.9,
System Development, page 72 and in Table 4 – Tools and material used in the FSC project process, page 19.

**Method**
The migration method was mainly:

- “Migration principles” in its first version 1.0
- Partially described work procedures

Both are described in detail in chapter 6.5,
Defining the Migration Method, page 52.

The competency profiles that the project prepared regarding the migration were:

- Internally trained system assistants (convertors) who would do the migration work
- System developers who could perform the ongoing development of the tools
- Archivists who could contribute with professional knowledge
- A person to take responsibility for describing working procedures and migration principles

The evaluation deals only with periods A and B records, and focuses on:

- Result of migration
- Quality assurance of migration
- Future use of records in various contexts
  - Accessibility
  - Migrations

Migration of Period A and B records is illustrated as follows:

**Figure 8 - Machine migration of Period A and B records**

![Diagram of machine migration process](image)
## 6.10.1.2 Period A and B records – Essential Conclusions

- **Detailed guidelines, descriptions and routines that standardise all phases of the migration work are essential for the efficiency and quality of complex migrations**
  - Work process descriptions were devised for the convertors to use when migrating. These guidelines were updated on an ongoing basis and were considered invaluable when performing the daily work.
  - Migration principles constituted the convertor’s technical help tool for the correct mark-up of the records’ content. These tools were also continuously upgraded.
  - Records with similar structures were collected in clusters and allocated to two convertors for migration. One of them had the overall responsibility whilst the other acted as sparring partner and quality officer.

- **It is essential to use the iterative procedure when migrating records of complex and heterogeneous nature**
  - Repeated test runs of the records, followed up by corrections to the applications ensured a high level of standardisation of the records even though they were of varied structures.

- **Quality control by random check is a problem because it is costly to perform many random checks and too risky to perform too few.** Quality control suggests the option of automated testing
  - Information packages that were part of an organised series were not logically quality controlled\(^\text{77}\) if an approval already existed for records from the same series from which only the first and last submission were quality controlled.

- **Migrating of certain records demanded a lot of machine time (one took 101 hours to migrate) because:**
  - The hardware’s performance was inadequate.
  - KonvOld was slow to process large records containing many items in few tables.

- **The project was a success seen from the perspective of the information packages’ level, because:**
  - 100 % of the information packages were approved by the test programme, TEA
  - 100 % of the information packages had a uniform format and structure and hence lived up to the preservation standard, which ensured the possibility for future automated migration.
  - Only 2 submissions were not migrated since one was received in error and the other, containing only one file, could not be read.

- **The migration of A and B period records permitted (after approved exemptions) production of information packages with errors in relations, i.e. duplicates in primary keys and lack of foreign keys.** We did not consider as an alternative a form of machine, technical improvement of the errors, e.g. by supplying “dummy” key values. That solution would have harmed the data’s authenticity, but would have made future migrations easier and would have decreased the development time of the accessibility system, SOFIA, which was based on loading data into a relational database requiring uniform values in primary keys and complete mark-up of foreign keys. The migration of periods A and B records did not take into consideration the necessity for mark-up of ‘field functions’ (special information in electronic document management systems, (see Executive Order no. 342, page 12 and 13, Figure 3 and 4.). The consequence of this is that a part of the accessibility process of a filing system requires of a manual mark-up of these fields.

---
\(^{77}\) KonvOld quality controlled structurally, but the random check objective was quality control pertaining to intelligibility, e.g. not if a relation existed (KonvOld had already checked that), but whether it was logically correct.
• **Exemptions were given for a number of error types.** See chapter 6.10.1.3.1, Fejl! Henvisningskilde ikke fundet., page 102.

• **The quality of the general information (the documentation) in the completed information package is not always satisfactory.** The required information is available on all records but not in a homogenous form. Moreover, the description of documents is not sufficiently structured. This will make it more difficult to perform fully automated migrations in the future.

### 6.10.1.3 Periods A and B records – Description

#### 6.10.1.3.1 Periods A and B records – Migration method

The migration process was supported by 2 important document collections with the objective of providing guidelines and descriptions of all the phases of the work, namely work process description and migration principles.

The description of the work process included procedures for all phases of the migration process and a comprehensive description of the guidelines related to the exemptions.

During all the phases of the migration process, the convertors exchanged knowledge on specific issues pertaining to the migration.

Each week, all the group’s employees met to share knowledge and insights.

Each employee presented a status report on their work from the preceding week, mentioned problems, knowledge acquired and presented the solutions.

Problems and solutions where discussed in forum. The conclusions gave rise to:

- Changes to workflow
- Changes to migration principles
- Corrections to or devising of new functions for the migration tools

The project steering group made decisions - based on input from the working group - of a pragmatic nature, e.g. approval of deviation from Executive Order no. 342.

The table below shows a specimen of the workflow description documents with number and versions.

**Table 24 - Example of workflow description documents with versions**

<table>
<thead>
<tr>
<th>Document</th>
<th>No. of versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1._Digital_description.doc</td>
<td>22</td>
</tr>
<tr>
<td>1.a._Find_information_for_SKABER.TAB_and_ARKVER.TAB_in_DAISY.doc</td>
<td>5</td>
</tr>
<tr>
<td>1.b._Decision_making_about_non-migration.doc</td>
<td>1</td>
</tr>
<tr>
<td>1.b._Transfer_record_for_processing_as_unique_record version 1.doc</td>
<td>3</td>
</tr>
<tr>
<td>3._quality_control version 1.doc</td>
<td>6</td>
</tr>
<tr>
<td>4._Final_migration_and_copying_for_archival_storage version 3.doc</td>
<td>6</td>
</tr>
<tr>
<td>4._Migrate_record_with_KonvOld udg 1.doc</td>
<td>2</td>
</tr>
<tr>
<td>6._Test_migrated_record_with_TEA3 udg 1.doc</td>
<td>8</td>
</tr>
</tbody>
</table>

The illustrated phases below represent the migration phases of the individual records:
1. Allocation of records for migration

The records were arranged in series each containing similar submissions in the form of registries or filing systems, which were submitted in uniform structure and format and from the same supplier or public office.

The table below illustrated a selection of these series – in all 184:

**Table 25 - Arranging the records, as far as possible, in similar structures and systematic uniform series**

<table>
<thead>
<tr>
<th>Series</th>
<th>No. of information package IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment registry</td>
<td>14</td>
</tr>
<tr>
<td>ATPs employee registry (Supplementary Labour Market Fund)</td>
<td>18</td>
</tr>
<tr>
<td>Employment registry</td>
<td>12</td>
</tr>
<tr>
<td>Housing benefits registry</td>
<td>12</td>
</tr>
<tr>
<td>Budget registry</td>
<td>16</td>
</tr>
<tr>
<td>Pupil statistics</td>
<td>10</td>
</tr>
<tr>
<td>Traffic accident registry</td>
<td>15</td>
</tr>
<tr>
<td>Greenland Trade</td>
<td>17</td>
</tr>
<tr>
<td>Harbour system registry</td>
<td>68</td>
</tr>
<tr>
<td>Filing system SJS</td>
<td>1</td>
</tr>
<tr>
<td>Filing system SJS I</td>
<td>15</td>
</tr>
<tr>
<td>Filing system SJS II</td>
<td>5</td>
</tr>
<tr>
<td>Municipal budget</td>
<td>13</td>
</tr>
<tr>
<td>Kommunedata i/s</td>
<td>4</td>
</tr>
<tr>
<td>Social security registry</td>
<td>11</td>
</tr>
<tr>
<td>Control card registry</td>
<td>10</td>
</tr>
<tr>
<td>Crime registry</td>
<td>13</td>
</tr>
<tr>
<td>Teachers registry</td>
<td>11</td>
</tr>
<tr>
<td>Salary statistics</td>
<td>20</td>
</tr>
<tr>
<td>Name registry</td>
<td>16</td>
</tr>
<tr>
<td>NYTSCR - system</td>
<td>6</td>
</tr>
<tr>
<td>Retirements registry</td>
<td>12</td>
</tr>
<tr>
<td>Register</td>
<td>15</td>
</tr>
<tr>
<td>Accounts</td>
<td>40</td>
</tr>
</tbody>
</table>
The convertors were assigned specific information package IDs for migration. Each convertor worked with records that were related to their specific competencies and knowledge. This method of organised series not only ensured a better quality of the end product, but also shortened the processing time.

From the onset, we selected a person who had the responsibility for the quality control of the migrated records. The quality controller could also act as a sparring partner for the convertor when trying to solve problems. The benefit of this was that the controller at this point already had knowledge of the records.

The migration sequence of the records was also influenced by requests from another concurrent project, namely the TGP project\(^7^8\), which ordered a specific information package to test the project’s tools or for accessibility purposes.

This was, however, not coordinated beforehand, and the spontaneous requests slowed down the process.

2. **Retrieval of paper documentation** and comparing it with the scanned documents registered in the preservation database, MARY, and saved as files on the migration raid. The errors and deficiencies to the scans were reported and amended.

The following shows an example of how a document was affiliated to a specific record:

---

\(^7^8\) The accessibility project
See also chapter 6.7,
Preparation of Documentation page 63.

3. Collecting of general technical pertaining to the IPs

Before the convertors could initiate the keying-in of the digital descriptions, it was necessary to:

- Find information for general tables\(^79\) in MARY and in DAISY. In several instances, we found incompatibility or errors in both systems, which required clarification.
- Review the documentation and prepare for typing-in the digital description of:
  - The format (code page, number of tables, content of tables, code lists)
  - The structure (hierarchical /relational, primary keys, foreign keys)

The following table shows examples of record and data descriptions.

![Figure 11 - Description of records and data](image)

The documentation was submitted directly from a scanned file by using OCR reading via MARY. Quite often the documentation was not sufficient, making it necessary to perform an analysis of the data files by using various tools such as TapeDump, FSC_Debug, word-processing.

In order to collect knowledge about the records their respective filing systems were also used.

4. Entry of the digital description data

\(^{79}\) Information such as submission type, supplier dates, archive creator and system name are included in special tables e.g. SKABER.TAB, ARKVER.TAB, cf. Executive Order no. 342.
The digital description, along with the data files and the scanned documents, constituted the basis for creating the completed information packages in the migration tool KonvOld.

KonvOld retrieved the following information from the digital description:

- Content of the record (files, documents, code lists, general information (documentation))
- The structure of the record (relations in the form of mark-up of primary and foreign keys)
- The format of the input files with references to migration routines of code pages and data types.
- The format of the input files with references to the method used for the possible restructuring from hierarchical relational structures
- Content of metadata file

The digital description was typed in as an XML file in an InfoPath form supported by a schema. The keying-in tool performed online validations of all elements in the file.

It paid off well to work with the concept of organised series as it was possible to re-use (copy) digital descriptions already typed in from other records.

The document, ”Migration principles”, functioned as a reference manual that described each field in an XML file for three main points:

- **Rule**: the definition of the valid content of a field
- **Method**: how to retrieve the valid information
- **Areas of doubt**: how to handle possible deficiencies or doubt regarding the validated information

Working with InfoPath proved to be both slow and cumbersome when the records included several tables, many fields and long code lists. In such instances we, therefore, substituted InfoPath with XMLSpy or word processing programmes, e.g. TextPad and Notepad.

We developed special tools to type-in code lists.

To illustrate the keying-in method, the following screen dump is from the digital description with a record definition and a picture with mark-up of primary and foreign keys:

---

80 Cf. item 1. *Allocation of records for migration*, page 83.
**Figure 12 - Digital description with record definition**

<table>
<thead>
<tr>
<th>Nr</th>
<th>Feltnavn</th>
<th>Datatype</th>
<th>SP</th>
<th>L</th>
<th>Feltnote</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>jour, nr 1</td>
<td>NUM_002</td>
<td>1</td>
<td>1</td>
<td>Digital description på regnsommeren</td>
</tr>
<tr>
<td>2</td>
<td>form</td>
<td>OPRING_02</td>
<td>1</td>
<td>2</td>
<td>Form på skilt med skiltvæg</td>
</tr>
<tr>
<td>3</td>
<td>ant, db</td>
<td>NUM_002</td>
<td>1</td>
<td>3</td>
<td>Antal dæk, som faktureren dækker</td>
</tr>
<tr>
<td>4</td>
<td>note, nr</td>
<td>NUM_002</td>
<td>1</td>
<td>4</td>
<td>Noter/faktureren fra not/faktureringsbogen</td>
</tr>
<tr>
<td>5</td>
<td>dat, nr</td>
<td>OPRING_02</td>
<td>1</td>
<td>5</td>
<td>Datoer/faktureren på skiltvæg af sikringsplan.</td>
</tr>
</tbody>
</table>
The table below shows the extent of the digital descriptions and number of keyed-in elements:

**Table 26 - Extent of digital descriptions in number and in elements**

<table>
<thead>
<tr>
<th>Digital descriptions</th>
<th>Described tables</th>
<th>Described fields</th>
<th>Keyed code lists</th>
<th>Keyed code values with accompanying code names</th>
<th>Demarcat ed foreign keys</th>
<th>Character routines used</th>
<th>Data types used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,291</td>
<td>11,187</td>
<td>447,203</td>
<td>48,653</td>
<td>2,854,434</td>
<td>9,492</td>
<td>25</td>
</tr>
</tbody>
</table>

5. **Pilot migration and error handling**

The migration programme, KolvOld, was used for ongoing testing of the data description of the records. A test migration was used for:

- Processing designated files
- Test of designated subjects, e.g. primary keys, foreign keys or code lists.

A test migration resulted in a log file including a description of potential errors.

In general, there were three error types:
• Incorrect mark-up of the digital description
• A data error or a data discrepancy
• A data deficiency in the form of doublets in primary keys, lacking in foreign keys, lacking code value designation as well as characters in text fields which could not be translated

These are described thoroughly below.

A. Incorrect mark-up of the digital description: e.g. incorrect reporting of code page, record length or primary key. Errors can be rectified by correcting the digital description.

The following illustrated an erroneous reporting of code page which resulted in an incorrect character translation with the ensuing correction of the error by providing the correct code page:

Figure 14 - Example of erroneous reporting and a code page which resulted in incorrect character translation
B. A data error or a data discrepancy that required one or more of the following actions:

- Preparation of a new data type routine for e.g. a new data format that could be migrated to the DATE type

The table below presents how many special (compared to the standard) data type migration routines were developed and their uses:

Table 27 - Specially developed data type migration routines and their uses

<table>
<thead>
<tr>
<th>Data type routines</th>
<th>Applied to no. of information package IDs (IPIDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE 001 (standard)</td>
<td>451</td>
</tr>
<tr>
<td>DATE 003</td>
<td>1</td>
</tr>
<tr>
<td>DATE 004</td>
<td>8</td>
</tr>
<tr>
<td>DATE 005</td>
<td>195</td>
</tr>
<tr>
<td>DATE 006</td>
<td>82</td>
</tr>
<tr>
<td>DATE 009</td>
<td>5</td>
</tr>
<tr>
<td>DATE 011</td>
<td>5</td>
</tr>
<tr>
<td>DATE 012</td>
<td>14</td>
</tr>
<tr>
<td>DATE 013</td>
<td>21</td>
</tr>
<tr>
<td>DATE 014</td>
<td>29</td>
</tr>
<tr>
<td>DATE_015</td>
<td>1</td>
</tr>
<tr>
<td>DATE_016</td>
<td>4</td>
</tr>
<tr>
<td>DATE_017</td>
<td>1</td>
</tr>
<tr>
<td>DATE_018</td>
<td>96</td>
</tr>
<tr>
<td>DATE_019</td>
<td>73</td>
</tr>
<tr>
<td>DATE_020</td>
<td>18</td>
</tr>
<tr>
<td>DATE_021</td>
<td>7</td>
</tr>
<tr>
<td>DATE_022</td>
<td>1</td>
</tr>
<tr>
<td>DATE_023</td>
<td>6</td>
</tr>
<tr>
<td>DATE_024</td>
<td>7</td>
</tr>
<tr>
<td>DATE_025</td>
<td>6</td>
</tr>
<tr>
<td>EXP_001 (standard)</td>
<td>2</td>
</tr>
<tr>
<td>EXP_003</td>
<td>29</td>
</tr>
<tr>
<td>EXP_004</td>
<td>11</td>
</tr>
<tr>
<td>NUM_001 (standard)</td>
<td>709</td>
</tr>
<tr>
<td>NUM_002</td>
<td>321</td>
</tr>
<tr>
<td>NUM_004</td>
<td>132</td>
</tr>
<tr>
<td>NUM_005</td>
<td>163</td>
</tr>
<tr>
<td>NUM_006</td>
<td>1</td>
</tr>
<tr>
<td>NUM_007</td>
<td>49</td>
</tr>
<tr>
<td>NUM_008</td>
<td>126</td>
</tr>
<tr>
<td>NUM_009</td>
<td>226</td>
</tr>
<tr>
<td>NUM_010</td>
<td>8</td>
</tr>
<tr>
<td>NUM_011</td>
<td>39</td>
</tr>
<tr>
<td>NUM_012</td>
<td>147</td>
</tr>
<tr>
<td>NUM_013</td>
<td>35</td>
</tr>
<tr>
<td>NUM_014</td>
<td>16</td>
</tr>
<tr>
<td>NUM_015</td>
<td>6</td>
</tr>
<tr>
<td>NUM_016</td>
<td>6</td>
</tr>
<tr>
<td>NUM_017</td>
<td>6</td>
</tr>
<tr>
<td>NUM_018</td>
<td>6</td>
</tr>
<tr>
<td>NUM_019</td>
<td>6</td>
</tr>
<tr>
<td>NUM_020</td>
<td>5</td>
</tr>
<tr>
<td>NUM_022</td>
<td>24</td>
</tr>
<tr>
<td>NUM_023</td>
<td>21</td>
</tr>
<tr>
<td>NUM_024</td>
<td>7</td>
</tr>
<tr>
<td>REAL_001 (standard)</td>
<td>239</td>
</tr>
<tr>
<td>REAL_002</td>
<td>119</td>
</tr>
<tr>
<td>REAL_003D01</td>
<td>25</td>
</tr>
<tr>
<td>REAL_003D02</td>
<td>48</td>
</tr>
<tr>
<td>REAL_003D03</td>
<td>2</td>
</tr>
<tr>
<td>REAL_003D06</td>
<td>2</td>
</tr>
<tr>
<td>REAL_004D01</td>
<td>30</td>
</tr>
<tr>
<td>REAL_004D02</td>
<td>53</td>
</tr>
<tr>
<td>REAL_004D03</td>
<td>8</td>
</tr>
<tr>
<td>REAL_004D04</td>
<td>3</td>
</tr>
<tr>
<td>REAL_004D06</td>
<td>2</td>
</tr>
<tr>
<td>REAL_005</td>
<td>2</td>
</tr>
<tr>
<td>REAL_006D02</td>
<td>6</td>
</tr>
<tr>
<td>REAL_007D01</td>
<td>1</td>
</tr>
<tr>
<td>REAL_007D02</td>
<td>23</td>
</tr>
<tr>
<td>REAL_007D03</td>
<td>10</td>
</tr>
<tr>
<td>REAL_008D02</td>
<td>2</td>
</tr>
<tr>
<td>REAL_009</td>
<td>4</td>
</tr>
<tr>
<td>STRING_001 (standard)</td>
<td>1,245</td>
</tr>
<tr>
<td>STRING_003</td>
<td>154</td>
</tr>
<tr>
<td>STRING_004</td>
<td>63</td>
</tr>
<tr>
<td>STRING_006</td>
<td>54</td>
</tr>
<tr>
<td>STRING_007</td>
<td>9</td>
</tr>
<tr>
<td>STRING_008</td>
<td>54</td>
</tr>
<tr>
<td>STRING_009</td>
<td>186</td>
</tr>
<tr>
<td>STRING_010</td>
<td>6</td>
</tr>
<tr>
<td>STRING_011</td>
<td>1</td>
</tr>
<tr>
<td>STRING_012</td>
<td>1</td>
</tr>
<tr>
<td>STRING_013</td>
<td>2</td>
</tr>
<tr>
<td>STRING_014</td>
<td>1</td>
</tr>
<tr>
<td>STRING_015</td>
<td>371</td>
</tr>
<tr>
<td>STRING_016</td>
<td>1</td>
</tr>
<tr>
<td>STRING_017</td>
<td>9</td>
</tr>
<tr>
<td>STRING_019</td>
<td>17</td>
</tr>
<tr>
<td>STRING_020</td>
<td>37</td>
</tr>
<tr>
<td>STRING_021</td>
<td>1</td>
</tr>
<tr>
<td>STRING_SPACE</td>
<td>335</td>
</tr>
<tr>
<td>STRING_ZERO</td>
<td>60</td>
</tr>
<tr>
<td>TIME_001 (standard)</td>
<td>4</td>
</tr>
<tr>
<td>TIME_003</td>
<td>7</td>
</tr>
<tr>
<td>TIME_005</td>
<td>5</td>
</tr>
<tr>
<td>TIMESTAMP_001 (standard)</td>
<td>64</td>
</tr>
<tr>
<td>TIMESTAMP_003</td>
<td>2</td>
</tr>
<tr>
<td>TIMESTAMP_004</td>
<td>3</td>
</tr>
<tr>
<td>TIMESTAMP_005</td>
<td>35</td>
</tr>
</tbody>
</table>

**Total number: 92 routines**

- Preparing a new code page, e.g. a variant of EBCDIC with a presentation of Danish code pages.

The tables below present how many special (compared to the standard) *code page migration routines* were developed and their uses:
A pre-migration of one or more input files in order to render the format acceptable in KonvOld. In total, 167 records have been pre-migrated, allocated to 134 from the A period and the other 33 from the B period.

Further development of the respective functions in KonvOld, which could handle specific data discrepancies. The following table shows versions of KonvOld.

<table>
<thead>
<tr>
<th>Version</th>
<th>Applied to no. of IPIDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>KonvOld 1.0.0.0 – 1.0.0.6</td>
<td>56</td>
</tr>
<tr>
<td>KonvOld 1.0.0.7</td>
<td>74</td>
</tr>
<tr>
<td>KonvOld 1.0.0.8</td>
<td>4</td>
</tr>
<tr>
<td>KonvOld 1.0.0.9</td>
<td>17</td>
</tr>
<tr>
<td>KonvOld 1.0.1.0</td>
<td>38</td>
</tr>
<tr>
<td>KonvOld 1.0.1.1</td>
<td>11</td>
</tr>
<tr>
<td>KonvOld 1.0.1.10</td>
<td>6</td>
</tr>
<tr>
<td>KonvOld 1.0.1.11</td>
<td>41</td>
</tr>
<tr>
<td>KonvOld 1.0.1.12</td>
<td>18</td>
</tr>
<tr>
<td>KonvOld 1.0.1.13</td>
<td>34</td>
</tr>
<tr>
<td>KonvOld 1.0.1.14</td>
<td>34</td>
</tr>
<tr>
<td>KonvOld 1.0.1.15</td>
<td>47</td>
</tr>
<tr>
<td>KonvOld 1.0.1.16</td>
<td>30</td>
</tr>
<tr>
<td>KonvOld 1.0.1.17</td>
<td>43</td>
</tr>
<tr>
<td>KonvOld 1.0.1.18</td>
<td>50</td>
</tr>
<tr>
<td>KonvOld 1.0.1.19</td>
<td>43</td>
</tr>
<tr>
<td>KonvOld 1.0.1.2</td>
<td>34</td>
</tr>
<tr>
<td>KonvOld 1.0.1.20</td>
<td>31</td>
</tr>
<tr>
<td>KonvOld 1.0.1.21</td>
<td>68</td>
</tr>
<tr>
<td>KonvOld 1.0.1.22</td>
<td>1</td>
</tr>
<tr>
<td>KonvOld 1.0.1.23</td>
<td>16</td>
</tr>
<tr>
<td>KonvOld 1.0.1.24</td>
<td>53</td>
</tr>
<tr>
<td>KonvOld 1.0.1.25</td>
<td>53</td>
</tr>
<tr>
<td>KonvOld 1.0.1.26</td>
<td>41</td>
</tr>
<tr>
<td>KonvOld 1.0.1.27</td>
<td>21</td>
</tr>
<tr>
<td>KonvOld 1.0.1.28</td>
<td>55</td>
</tr>
<tr>
<td>KonvOld 1.0.1.29</td>
<td>27</td>
</tr>
<tr>
<td>KonvOld 1.0.1.3</td>
<td>4</td>
</tr>
<tr>
<td>KonvOld 1.0.1.30</td>
<td>147</td>
</tr>
<tr>
<td>KonvOld 1.0.1.31</td>
<td>3</td>
</tr>
<tr>
<td>KonvOld 1.0.1.32</td>
<td>46</td>
</tr>
<tr>
<td>KonvOld 1.0.1.33</td>
<td>59</td>
</tr>
<tr>
<td>KonvOld 1.0.1.4</td>
<td>1</td>
</tr>
<tr>
<td>KonvOld 1.0.1.5</td>
<td>14</td>
</tr>
<tr>
<td>KonvOld 1.0.1.6</td>
<td>24</td>
</tr>
<tr>
<td>KonvOld 1.0.1.7</td>
<td>15</td>
</tr>
<tr>
<td>KonvOld 1.0.1.8</td>
<td>29</td>
</tr>
<tr>
<td>KonvOld 1.0.1.9</td>
<td>3</td>
</tr>
</tbody>
</table>

**In total 43 versions**

C. A data deficiency in the form of doublets in primary keys, lacking in foreign keys, lacking code value designation as well as characters in text fields which could not be translated. These kinds of errors could not be corrected.

Test migration and correction of errors were conducted in several stages until we were sure that the only remaining mistakes were those that could not be corrected without resorting to interpretations; or if corrections would be detrimental to the record’s authenticity.

6. Finalising the migration

It was possible to perform the final machine end migration when all potential problems were solved; metadata were properly described in an XML file, and when all control runs in KonvOld were completed.

The end migration supplied a migration file including a description of all the phases in the machine migration and a statement of all errors that may have been carried over into the final information package.

The end migration was the only part of the migration process that was fully automated.

The following table shows a statistic of the machine time consumed. The times are rounded up to whole hours.
Table 30 - Machine time for end migration

<table>
<thead>
<tr>
<th>Machine migration time in hours</th>
<th>Number of IPIDs</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,097</td>
<td>1,097</td>
</tr>
<tr>
<td>2</td>
<td>54</td>
<td>108</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>80</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>144</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>75</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>57</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>4</td>
<td>124</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>128</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>34</td>
<td>4</td>
<td>136</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>46</td>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td>71</td>
<td>2</td>
<td>142</td>
</tr>
<tr>
<td>97</td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>98</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>101</td>
<td>1</td>
<td>101</td>
</tr>
</tbody>
</table>

| IP total 1,291 | Hours total 3,371 |

The records were migrated using equipment of similar type. Nearing the end of the project, we bought two new machines in order to shorten the processing time of some of the time-consuming migrations.

The average machine time used was 114 minutes, but the 0.25 quantile was 1 minute, 0.5 quantile was 5 minutes and 0.75 quantile was 29½ minutes.
The most time consuming migration constituted 10 % of the total migration, but took over 85 % of the total machine time.

The 2 % longest migration time took 50 % of the total machine time.

The 25 longest migrations had no pre-migrations, had only 1 or 5 input files and were in code page ISO 8859-1 or EBCDIC 277.

There was no correlation between number of pre-migrations, of input files or divergent code pages.

The general information (documentation) on the record, along with the scanned documents, included the migration file, the digital description, schema for the digital description and documentation of potential pre-migrations.

Before the records could be cleared for quality control, a control was performed with the control tool TEA3. The test results from TEA were compared with the content of the migration file from KonvOld.

Concordance between TEA and KonvOlds test results gave additional confirmation that the quality of the migration was satisfactory.

7. Quality assurance

All end migrated information packages were quality controlled by a different convertor than the one who performed the end migration.

The control had the following objective:

- Show potential errors, deficiencies or discrepancies from the migration principles in the metadata.
- Show possible typing errors, e.g. spelling mistakes.
- Show possible misinterpretations that could lead to incorrect logical definitions of data structures and table relationships.

The information packages included in an organised series were not quality controlled if an approval already existed of records from the same series. In such series, usually the first and last submission in a series was quality controlled.

During the control process, a number of errors were detected which proved to be of a generic type and occurred not only in the records we were quality controlling, but also in the already end migrated and approved records. The errors were only of structural character and existed in:

- The XML files in the form of illegal characters from other code pages than ISO8859 Latin 1; lack of mark-up of various elements; incorrect reference to forms
- Migration log file in form of incorrect line counter

The errors were all machine corrected in one process by using a chapter 6.9,
System Development.

When a record was migrated, quality controlled and possibly corrected, it was then transferred to a migration raid, registered in DAISY as migrated and was thus ready for the transfer and the final preservation.

6.10.1.3.2 Period A and B records – End Result and Quality of End Product

End Result
In December 2008, the migration of the records from periods A and B was completed and resulted in completed information packages that lived up to the preservation standard, Executive Order no. 342.

The information packages were transferred to archival storage in accordance with applicable rules that required the existence of three copies of each information package: two DVDs and a transfer to the State Library’s preservation RAIDs.

The end product distributed on information packages from both periods’ respective submission types was:

Table 31 - End product for period A and B records

<table>
<thead>
<tr>
<th>Period</th>
<th>Filing system</th>
<th>Registry</th>
<th>No. of IPs</th>
<th>No. of files</th>
<th>Added no. of scanned documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>46</td>
<td>604</td>
<td>650</td>
<td>5,185</td>
<td>10,110</td>
</tr>
<tr>
<td>B</td>
<td>299</td>
<td>342</td>
<td>641</td>
<td>8,617</td>
<td>14,061</td>
</tr>
<tr>
<td>Total</td>
<td>345</td>
<td>946</td>
<td>1,291</td>
<td>13,802</td>
<td>24,171</td>
</tr>
</tbody>
</table>

All told, two information packages from period A were not migrated:

- The one submission consisting of one file, originally submitted on tape, could not be read.
- The second one turned out to be an erroneous submission and was returned to Danish Data Archive (Dansk Data Arkiv), which was the designated receiver of the IP.

Quality of the End Product
The quality of the end product is evaluated based on the following 3 criteria:

1. All migrated records were approved by the test programme TEA
2. All records got a uniform format and structure. This has now made it possible to access all digital records by use of a standardized access tool (SOFIA). Furthermore, all records are ready for the next migration. Since all records are uniform in structure, we expect an automated (machine) migration using standardised tools.

---

81 Submission types are defined in Executive Order no. 342 as “1”, “2” and “3” for registry systems and for “A” and “B” for filing systems:

“1” is a one-off submission of data from a registry in which data are no longer corrected nor added;

“2” is a year’s submission of data from a registry in which data are no longer corrected nor added for the submitted years;

“3” is a snapshot submission of data that are continuously corrected or added to.

“A” is a submission of data from an electronic filing system or an electronic document management system, where no data additions or corrections are performed.

“B” is a submission of data from an electronic filing system or an electronic document management system where data are still being added.
3. **Compliance with rules in Executive Order no. 324.** The rules had exceptions for the following 9 errors:
   - Data type DATE as 6 digits in format YYMMDD
   - Preset blank fields in the data type STRING
   - Up to 5 % duplicates of primary keys in a table
   - Up to 30 % missing foreign key values in a foreign key
   - Tables not related to other tables
   - Missing code terms for code values were filled in with the standard text ‘Code Value unknown to The Danish National Archives in accordance with Format and Structure Conversion 2005-08’
   - Characters that could not be translated to a value character in ISO 8859 Latin 1 were substituted with an inverted question mark ‘¿’.
   - Error margins for primary and foreign keys were accepted at a higher error rate, if the error margin had a logical explanation for the relations between the tables. If e.g. a relation to a blank table was correct, we decided to demarcate the relation even though this resulted in a 100 % foreign key error.
   - XML/XSD format of documents with general information (migration log file, the digital description, forms for the digital description and the pre-migration documentation). Executive Order no. 342 did not allow for submission in XML or XSD schema.

**Status of known errors in end-product that were given dispensation:**
The following table shows known and accepted errors, detailed in the ensuing tables:

<table>
<thead>
<tr>
<th>IPs With primary key errors</th>
<th>Tables with primary key errors</th>
<th>IPs With foreign key errors</th>
<th>Foreign key with errors</th>
<th>IPs With code list errors</th>
<th>Code lists with errors</th>
<th>IPs With character errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 %</td>
<td>9 %</td>
<td>39.4 %</td>
<td>21.8 %</td>
<td>57.3 %</td>
<td>15.4 %</td>
<td>8.2 %</td>
</tr>
</tbody>
</table>

**Characteristics of known, authorized errors in end-product:**
   - Errors resulted from errors in input data, not from the migration, which authorized them. Error correction on future automated processing is therefore in principle not necessary.
   - Errors are of a technical type, i.e. duplicates of primary keys.
   - Errors can only be corrected manually, but this can hurt authenticity and requires devising methods, principles and documentation of performed error corrections.
   - Estimating the resources needed to correct errors is not possible until thorough analysis is performed. Furthermore, the analysis is also necessary if the extent of the errors is to be correctly evaluated. It is e.g. not know whether the errors affect unique, preservation worthy data or whether the corrupted data amount is minimal or substantial.

Below are tables that have formed the basis for the outline tables showed above.

**Table 33 - Primary key error**

<table>
<thead>
<tr>
<th>Description</th>
<th>Errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>452 out of a total of 1,291 IPs have one or more tables with errors in the primary key</td>
<td>35</td>
</tr>
<tr>
<td>12,599 have no errors in the primary key out of a total of 13,802 tables</td>
<td>91.2</td>
</tr>
</tbody>
</table>
753 have errors in primary keys in 1 % of entries out of a total of 13,802 tables

196 out of a total of 13,802 tables have errors in the primary key in 10 % to 100 % of entries; and 17 tables have errors in the primary key in 100 % of their entries.

<table>
<thead>
<tr>
<th>Description</th>
<th>Errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>509 out of a total of 1,291 IPs have a foreign key error in one or more tables</td>
<td>39.4</td>
</tr>
<tr>
<td>1,203 out of a total of 9,492 foreign keys are faulty</td>
<td>12.6</td>
</tr>
<tr>
<td>1,260 out of a total of 9,492 foreign keys have up to 1 % errors</td>
<td>13.2</td>
</tr>
<tr>
<td>547 out of a total of 9,492 foreign keys have between 1 % and 10 % errors</td>
<td>5.7</td>
</tr>
<tr>
<td>134 out of a total of 9,492 foreign keys have between 10 % and 30 % errors</td>
<td>1.4</td>
</tr>
<tr>
<td>130 out of a total of 9,492 foreign keys have over 30 % errors, of which 27 of them have 100 % errors</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Table 35 - Code list error

<table>
<thead>
<tr>
<th>Description</th>
<th>Errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>740 out of a total of 1,291 IPs have one or more code list with errors</td>
<td>57.3</td>
</tr>
<tr>
<td>1,979 out of a total of 13,802 tables have one or more code list with errors</td>
<td>14.3</td>
</tr>
<tr>
<td>48,652 fields with associated code lists were demarcated. Out of 48,652 fields (code lists) there were 7,507 fields with incomplete code lists.</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Table 36 - Character error

<table>
<thead>
<tr>
<th>Description</th>
<th>Errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>106 out of a total of 1,291 IPs, more than one has character errors</td>
<td>8.2</td>
</tr>
<tr>
<td>Total number of character errors is 22,101 distributed as follows:</td>
<td></td>
</tr>
<tr>
<td>9 IPs have over 100 character errors</td>
<td>0.7</td>
</tr>
<tr>
<td>33 have between 10 and 99 character errors</td>
<td>2.5</td>
</tr>
<tr>
<td>64 have under 10 character errors</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Status report on other errors in end-product

Aside from the exemptions from Executive Order no. 342, the following summarises the errors found since the completion of the FSC project.

1. Known, not corrected errors that occurred during production

IPs with errors were found after completing the production of the information packages.

The table below shows a description of the errors:

Table 37 - Known errors

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect submission type</td>
<td>Records were incorrectly designated as type “2” instead of type “3”. The error was due to a lack of understanding by the convertor of the term &quot;submission type&quot;.</td>
</tr>
</tbody>
</table>

Definition of submission type code ('Executive Order no. 342', page 22) is as follows:
1: Information package includes a one-off submission of data from a registry in which data are no longer corrected nor added.
2: Information package includes a year’s submission of data from a registry in which data are no longer corrected nor added for the submitted years
3: Information package that includes a snapshot submission of data that are continuously corrected or added to

The error was detected in February 2008 and was assessed as being of minor importance.

<table>
<thead>
<tr>
<th>Old migration files</th>
<th>In the period 18 December 2006 to 13 March 2007, 134 filing systems were migrated with the old version of the KonvOld log file. The migration file does not contain errors, but was disorganized and not very informative.</th>
</tr>
</thead>
<tbody>
<tr>
<td>XML files with incorrect mark-up of code definitions</td>
<td>2 information package IDs (00000136 and 00000137) had errors in their XML files during the process of mark-up of the code definitions: &lt;kodedef&gt; element missing and XML files cannot be paired in TEA and XMLSpy. The error can be corrected manually.</td>
</tr>
</tbody>
</table>

2. Known, uncorrected errors observed during accessibility (2009-)

Besides a number of know issues (described in chapter 6.10.1.3, Periods A and B records – Description88) which were dealt with by the accessibility procedures, small errors were detected in some information packages, such as incorrect designation of data type in the case file number in the electronic filing system (e.g. need for change from NUM to STRING and illegal characters (e.g. è, ã and ô or ü)). See list of detected errors in appendix 7 – FSC error list.

The quality of the general information (GENINFO documents) in the finished information packages

A machine read review of the GENINFO.TAB files in the 1,291 migrated records was performed.

The objective of the extraction was to gain an overview of the content and quality of the description of the general information.

It was found that all the information packages included documentation of the migration in the form of:

- The digital description
- Schema for the digital description
- Migration log file

All pre-migrated records include the required pre-migration documentation.

An additional demand to documentation required that all records, as a minimum, should include documentation on ”submission list’ and on “documentation registration”.

These demands have not been totally complied with:

Table 38 - Lack of compliance with demands for documentation

<table>
<thead>
<tr>
<th>Period</th>
<th>No. of information</th>
<th>No. of documentation’s registering documents</th>
<th>Number of submission lists</th>
</tr>
</thead>
</table>
It was also found that all information packages included, as a minimum, one document (apart from the migration documentation).

This indicates that the required information exists in all records, but not in a uniform style, while the document descriptions were not well structured.

6.10.1.3.3 Use of Resources

Use of resources in connection with the migrations is calculated as machine time in addition to the manual work the convertors performed in connection with the preparation of data, surveillance of machines and the performance of quality controls.

Another important condition is that the time consumed does not reflect the possible occurrence of trial migrations often performed on an information package.

The time used is thus only a reflection of the amount of time used on the final migration of an information package.

The table below shows the total time used in the migration phase divided by main points:

<table>
<thead>
<tr>
<th>Type</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>Total</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Definition of migration method</td>
<td>425</td>
<td>1,136</td>
<td>719</td>
<td>2,279</td>
<td></td>
<td>10.18 %</td>
</tr>
<tr>
<td>B. Migration, filing system (A period)</td>
<td>107</td>
<td>610</td>
<td>1,258</td>
<td>74</td>
<td>2,049</td>
<td>9.15 %</td>
</tr>
<tr>
<td>C. Migration, registry (A period)</td>
<td>6</td>
<td>189</td>
<td>1,227</td>
<td>5,783</td>
<td>7,205</td>
<td>32.17 %</td>
</tr>
<tr>
<td>D. Migration, filing system (B period)</td>
<td>11</td>
<td>592</td>
<td>1,834</td>
<td>56</td>
<td>2,493</td>
<td>11.13 %</td>
</tr>
<tr>
<td>E. Migration, registry (B period)</td>
<td>464</td>
<td>1,763</td>
<td>2,487</td>
<td>4,714</td>
<td></td>
<td>21.05 %</td>
</tr>
<tr>
<td>F. Pre-migration</td>
<td>199</td>
<td>751</td>
<td>633</td>
<td>1,583</td>
<td></td>
<td>7.07 %</td>
</tr>
<tr>
<td>G. Quality control</td>
<td>58</td>
<td>1,345</td>
<td>668</td>
<td>2,071</td>
<td></td>
<td>9.25 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>124</strong></td>
<td><strong>2,537</strong></td>
<td><strong>9,314</strong></td>
<td><strong>10,420</strong></td>
<td><strong>22,395</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>

In Table 30 - Machine time for end migration, page 14, the machine time usage is summed up to a total of 3,371 hours. When subtracting this figure from the above table’s time used for migration and pre-migration, we get the number of hours the convertors used in the migration phase (the total time used). The manual time used would thus be 18,044 – 3,371 = 14,673 hours or 11.3 man-years and constitutes 81.3 % of the total time used for the migration.

This figure should be taken with great deal of reservation as it presumes that we know - which we actually don’t - exactly how convertors register their time consumption (perhaps they perform other tasks whilst the machines are running, etc.?).

It is, thus, also possible to perform a number of calculations on how long time it takes to migrate records as regards data quantity, number of files, documents and structures.

Table 40 - Overview of pre-migration and migration

<table>
<thead>
<tr>
<th>Type</th>
<th>Time</th>
<th>No. of</th>
<th>No. of</th>
<th>No. of</th>
<th>Data</th>
<th>No. of</th>
<th>No. of</th>
</tr>
</thead>
</table>

82 Rows B, C, D, E and F in table 37 – Migration phases’ machine and manual time usage
83 Note: the number of structures is calculated in relation to how the FSC project itself handled the records, as far as possible structurally and semantically, cf. Fejl! Henvisningskilde ikke fundet, page. 83.
The above numbers result in the following calculations:

<table>
<thead>
<tr>
<th>Type</th>
<th>Time per registry</th>
<th>Time per filing system</th>
<th>Time per file</th>
<th>Time per GB</th>
<th>Time per structure</th>
<th>Timer per document</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration A period</td>
<td>-</td>
<td>-</td>
<td>1.94 hours</td>
<td>58.75</td>
<td>101.47</td>
<td>0.99</td>
</tr>
<tr>
<td>Migration B period</td>
<td>-</td>
<td>-</td>
<td>0.93</td>
<td>19.09</td>
<td>94.10</td>
<td>0.57</td>
</tr>
<tr>
<td>Total</td>
<td>23.67</td>
<td>64.91</td>
<td>1.62</td>
<td>37.96</td>
<td>121.71</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Due to the lack of machine capacity in the project (faster / stronger PCs), it is very probable that a lot of time was wasted. This time waste can be attributed to the use of InfoPath for the keying-in of XML descriptions of large records (many tables / fields). InfoPath functioned perfectly when used with smaller records, but when the description exceeded a certain size, the product was downright useless.

Some of these problems could perhaps have been avoided if we had, at an earlier stage in the process, selected larger records for migration. Thus, we would have been able to identify the problems and to implement alternate solutions at an earlier stage, rather than waiting to solve the more complex problems associated with the migration of larger records that were first identified in the last half year of the project.

As the process of migrating larger records went forward, InfoPath and a lack of sufficient machine capacity had aggravated the problems as each migration cycle included many test runs.

6.10.1.4 Period C Records - Introduction

Data Basis
The data basis comprised 942 information packages. The information packages were structured in compliance with the regulations set forth in Circular No. 4, and were collected in one folder.

Tools
The programme used for the migration of the records, Konv2000, was developed and tested in 2006 (see chapter 6.9.3.4, Konv2000, page 78).

Method
The method was defined in the requirements specifications for Konv2000 and comprised three phases:

- Examination of information packages and defining those that should be machine migrated.
- Performance of the machine migration production

---

85 Annex 15 – System Description
• Control of the results which included testing a random selection of migrated records using the test tool TEA3

6.10.1.5 Period C Records – Essential Conclusions

• **The migration was successful.** 205 information packages were migrated via Konv2000, and 3 manually.
• **212 information packages did not at input - as expected – live up to the preservation standard** because they had not been correctly tested when received. The ensuing manual review and error correction cost 225 hours, but was not considered as a mistake on the part of FSC project.

6.10.1.6 Period C Records – Description

The migration process was performed in accordance with the predefined migration method:

• **Examination** of records showed that out of the 942 reviewed information packages:
  o 205 records could be machine processed
  o 734 records complied with the requirements in Executive Order no. 342, and thus did not require processing
  o 2 records were in an unknown format, i.e. a format that was not defined by a preservation standard
  o 1 record included files in format for respectively Circular no. 4 and Executive Order no. 342
• **The migration produced** the selected 205 information packages as well as the final migration log file.
• **Testing** of a segment of the migrated records with TEA showed that the structural improvements of the records were correctly performed. It also showed that several of the records did not live up to the preservation standard. The testing process was therefore extended to include all 942 records, which showed that 212 of them had the same error. The reason was that the records, *when submitted*, were not correctly tested by the test tool because the tool’s first version did not live up to the preservation standard’s specifications. Since Konv2000 assumed that the initial testing was performed correctly, the programme was not configured to address anything other than the few structural changes that were the consequence of the shift from complying with Circular no. 4 to Executive Order no. 342, cf. Table 20 - Konv2000 task portfolioTable 1, page 79.

6.10.1.6.1 Period C Records – End Result and Quality of End-Product

**Result**
205 information packages were improved structurally in order to comply with the applicable rules of Executive Order no. 342.
3 information packages with unknown/partial 2004 format were manually migrated by correcting the respective files in a text editor.

**Quality**
The test of the migrated records showed that, despite achieving the desired structural improvements of the data, the migrated XML files could not be parsed in TEA3, as the records were riddled with various errors.

A cursory review of the errors showed that they could not be machine corrected and hence an improvement for the data quality could not be achieved by manual review.
After the completion of the FSC project, we decided to do the manual review of all 942 information packages. The result of the project is shown in table form with a listing of all information packages. 212 information packages were corrected, and all 942 were successfully tested by TEA.

6.10.1.6.2 Period C Records – Use of Resources

The table below shows the total number of hours used for the migration of period C records, including system development.

Table 42 - Total number of hours used for the migration of period C records, including system development

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hours</td>
<td>0</td>
<td>184</td>
<td>0</td>
<td>41</td>
<td>224</td>
</tr>
</tbody>
</table>

In 2008, it took 41 hours to perform the migration and test of the migration results. The reprocessing of the 212 faulty records did not take up additional resources from the FSC project. In total, 225 hours were used to review and correct the errors.

---

86 Annex 22 - Period C records Project Result Overview
6.11 Transfer to Archival Storage

This part of the evaluation will only be descriptive as it does not include evaluation worthy data.

The transfer to the archival storage in the FSC project was conducted in two steps. The first transfer occurred at the end of the re-structuring of all the records (periods A, B and C). See also chapter 6.6, **Fejl! Henvisningskilde ikke fundet.**, page 58.

For records from periods A and B, the first transfer did not transfer finalised information packages, but only the repackaged and restructured input data to the final FSC migration. These editions were preserved as original editions of A-editions.

The records were subsequently end-migrated and transferred to the archival storage as B-editions complying with Executive Order no. 342.

Records from period C, which were received in preservation format equivalent to Circular 2000 or 2002, were initially transferred as original editions of A-editions.

Only a portion of them was subsequently end-migrated. These were – as in periods A and B records – transferred to the archival storage and labelled as edition B, migrated in compliance with Executive Order no. 342.

The first transfer entailed copying to a preservation RAID and to CDs.

The second transfer entailed the production of two sets of DVDs and the transfer of data to the State and University Library.

After the production of the information packages, the audio/visual records were transferred to the archival storage.

The first transfer consumed 379 hours, as it had to be performed manually. The DEA programme was developed for the second transfer. This programme enabled the automatic transfers, which meant that only ca. 10 hours were used for manual handling of burning (CDs) and transfer, surveillance and subsequent control.

It took 613 hours to develop DEA. In other words, it added an expense of 40 % to automate the process compared with the 379 manual hours used for the first transfer.

However, the programme is still being used and so the resources must be considered recouped many times over.
6.12 Audio/Visual Records

6.12.1 Introduction and evaluation object

The audio/visual material that was to be migrated in the FSC project was audio and motion pictures both in the form of videos and films.

The material was preserved on a plethora of formats and included ca. 1,900 analogue media, which were, generally speaking, in good condition. However, some of the media was mould infested.

It was estimated that the product would create circa 300 information packages.

6.12.2 Essential Conclusions

- The result of the migration of the audio/visual records was:

<table>
<thead>
<tr>
<th>Migration</th>
<th>No. of migrated records</th>
<th>Success rate as per the project’s assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio migration</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Video migration</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Film migration</td>
<td>0 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

- The outstanding items currently available are:
  - Audio. The audio migration was successful; however we have since found a couple of packing boxes with tapes that had not yet been digitised; this will be done (2011).
  - Video. All are digitised except one video tape. No other material has shown up since.
  - Film. The project cancelled plans for digitising films and postponed this to a future point in time. The reason for this decision was that there were not sufficiently suitable formats available for the long-term preservation of films. As far as we know, there are circa 3,000 film reels in The Danish National Archives’ collection. The FSC project considered the postponement of the digitisation to be a prudent decision regarding preservation, since it was deemed that the analogue media the films were preserved on were not in danger of being lost for posterity. However, we do not have any documented review of the state of the media, and the decision to place the media in more suitable preservation environment has not yet been acquiesced.

- Both the quality and the prices of the suppliers were reasonable.

- Resources
  - The total expenses are equal to 2.7 man-years and DKK 476,776.
  - The project was completed within budget (with a margin of circa 11 %), mainly due to the fact that the digitisation of the films was postponed and because the expenses for the digitisation alone had been estimated to DKK 714,000 which would have resulted in a budget overrun of circa 26 %.
  - The migration part of the project (the internal part) amounted to slightly over DKK 900,000 kroner, with 38.4 % of these expenses used on registering the

---

Note that the success rate of 100 % only applies to the project’s revised objective for the film migration. The FSC project had originally calculated to perform the film migration and the success rate would have been 0 %.
records in DAISY (almost 1 man-year), while the remaining 61.6 % was used on work related to project management, supplier management and processing in MAVDA (the tool for producing audiovisual information packages)  
  
- **The digitisation part (the extra part) amounted to ca. DKK 475,000**  
- **Hence, the division of resources between the two main tasks** is that the migration part constituted ca. 66 % of the projects expenses while the remaining 34 % went to expenses for external consultants (digitization).  
  
- **The price per minute for migration of audio and video** was respectively DKK 1.58 and 5.40 (close to factor 3).  
- **The project team changed their decisions often.** The reason for this was that the project’s pre-analysis was insufficient and the recommendations were changed frequently because of a more qualified decision-making basis. Two of the decisions can be considered controversial:  
  
- **When migrating audio:** the original reason for deciding to use external help for the digitisation process was that in-house staff was only able to digitise at a speed of 1:1. However, it showed that material digitized at higher speeds would not comply with the minimum recommendations of IASA. **This knowledge did not change the decision to use external suppliers.**  
  
- **When migrating video** the final decision was to choose MPEG2 instead of JPEG2K lossless, as in 2007 the latter did not yet have the desired market penetration. A digitisation to JPEG2K lossless was considered unpractical to accomplish because of the high level of digitisation expenses, need for space and because replaying the format would require special video servers. **The consequence of this decision was that the quality of the digitisation was poorer and the digitisation was incomplete, but still considered to be acceptable to the preservation standard.**

### 6.12.3 Description of Digitisation

#### 6.12.3.1 Audio

The analogue audio material comprised tape spool, cassette tapes, micro cassettes, 78 LP (records) and a few steel wire tapes. The major part of the audio media was tape spool and cassette tapes.

When the project was launched, it was decided to outsource the digitisation process despite the fact that The Danish National Archives actually possessed all equipment necessary to perform the digitisation process. However, digitisation would occur at a speed of 1:1 and would have taken 3,393 hours or 2.1 man-years to complete.

An external supplier would, however, be able to digitise at speeds of ca. 8x or 16x. This was considered a much cheaper solution.

It was originally decided to outsource the digitisation of all cassettes and tape spools to Dansk Blindebibliotek (DBB - Danish Library for the Blind). The price quoted was 585,200 and DKK 82,666 for storage, whilst Danmarks Radio (DR - Denmark’s Radio) would digitise more specialised formats (maximum DKK 50,000).

It was decided that sound digitisation would be transferred to WAVE 16 bit 44.1 kHz (CD quality), which was the maximum high speed at which DBB could digitize at. However, archival material in bad shape would be digitized to WAVE 24 bit 96 kHz with a view to possibly trying to restore the material. WAVE 16 bit was chosen despite its being of lower quality than that recommended by experts. Nevertheless, the chosen format was rated to be not only sufficient for the needs of The Danish National Archives, but also economically more feasible.

---

88 Cf. annex 23 – Time used.  
89 The International Association of Sound and Audiovisual Archives  
Later on the decision was changed to using digitisation with WAVE 48 kHz 24 bit - and not at high speed - as this would render the project compliant with IASAs lowest recommended quality. Hein IT stated in correspondence that the additional cost of digitising with WAVE 24 bit compared to 16 bit was insignificant. An increase in storage expenses would occur – but an additional expense of DKK 35,000 was estimated to be within budget.

6.12.3.2 Motion Pictures

The analogue material included VHS, Betamax, System2000, U-matic and cine films (super 8, 16 mm and 35mm). The major part of the records was cine films followed by VHS.

Video

It was temporarily decided that the preservation format for video should be JPEG2K lossless. However, there were a number of practical implications connected with migration to this format; hence, the decision was postponed to mid 2007\(^{91}\). The final choice was to go with MPEG2 MP@ML 15 Mbps cbr and MPEG layer II 384 kbps audio, instead of JPEG2K lossless, as the latter in 2007 still did not have the desired market penetration. A digitization to JPEG2K lossless was deemed impossible to perform due to the very high cost of digitisation, need for space and because replaying the format would require special video servers. The consequence of this was that the quality of the digitization was poor and the end product was compromised. This was deemed to be marginal and unavoidable.

Film

It was initially decided that the vendor VideoPro, or a similar external supplier, should perform the digitisation task, as we did not have the necessary equipment in-house. The estimated expense for this was DKK 714,350, with an additional DKK 196,811 for storage of the digital records. Later on in the process, it was decided that suitable formats for long time preservation of film did not exist and hence it was decided that film should not be digitised, but instead stored under best conditions in a climate regulated storage area\(^{92}\). It was estimated that the cheapest and most flexible solution would be to create own storage areas of ca. 30m2. Dansk Filminstitut (DFI – The Danish Film Institute) could provide an alternative solution but it was unclear if it was possible to reach an agreement with them, especially in terms of how flexible they would be for access rights to the storage area, etc.

It was decided that the task be handed over to a preservation specialist from the staff\(^{93}\).

As far as we know, there are about 3000 film reels in storage.

The FSC project deemed the postponement of the digitisation to be a prudent preservation decision, as it was estimated that the analogue media the films were preserved on was not in danger of obsolescence. There is however, no documented review of the medias’ condition or shape. The recommendation to place the media in optimal archival storage has not yet been complied with.

6.12.3.3 External suppliers

Audio

Hein IT, a sole proprietorship, with experience in sound restoration and who works in compliance with IASA TC-04 – and in some instances at even more stringent requirements, was asked to digitise sound\(^{94}\).

---

91 2005-232-000008, akt 51
92 2005-232-000008, akt 51
93 2005-232-000008, akt 51
94 The technical report form Hein IT on executing the project is available at 2005-232-00008, akt. 110
The external supplier generated the technical metadata.

Some of the tapes were "sweaty" and needed to be baked before the digitisation process could begin. For this purpose, an oven was rented from DBB for a “special” price.

Based on the quotation received for digitising tape spools, the average price was estimated to have been:

\[
\text{DKK 175,770} \div (620 \text{ tapes} \times 180 \text{ minutes}) = \text{DKK 1.575 per minute}
\]

Based on the quotation received for digitising cassette tapes, the average price was estimated to have been:

\[
\text{DKK 102,060 kr.} \div (720 \text{ tapes} \times 90 \text{ minutes}) = \text{DKK 1.575 kr. per minute.}
\]

A special quote was given for handling the specific challenges related to handling other media types, including cardboard records and e.g. damaged tapes and hence the price-per-minute cannot be compared.

The 204 cardboard records recorded by the mechanic Hjalmar Andersen during WWII required special handling.

**Video**

Moving Media in Ireland, who had experience from similar tasks, performed the digitisation process. They were estimated to be the cheapest and submitted a quote that was ca. 50 % lower than the one received from the other supplier (Decentia).²⁵

The major part of the migrated media was VHS tapes (249). In addition, there were 23 U-Matic, and a small collection of other formats. The price paid per tape was EUR 90, and EUR 175 EUR for 1 ½” and 1” reels.

The 204 cardboard records recorded by the mechanic Hjalmar Andersen during WWII required special handling.

The average price based on the invoices received for the digitisation of video was:

\[
\text{EUR 24,840} \div (\text{estimate of 34,290 minutes}) = \text{EUR 0.7244 per minute or DKK 5.40.}
\]

### 6.12.4 Description of migration

As part of the FSC project, the collection of audio/visual records should be migrated and preserved as digital information packages. We wanted to ensure efficiency in the process whilst ensuring a uniform quality and therefore we wanted partially to automate the process supported by a programme used for creating information packages of the material. The programme was specifically developed for FSC, with the future possibility of developing a version 2.0 that could be used for a broader audience outside the scope of the FSC project.

The programme should be able to receive input from:

- Extracts from DAISY – either as a CSV file or as a table in an SQL database;
- Documentation supplied from vendor as a CSV file registered by the digitised media’s barcode;
- WAV files named by the barcode and a numeration of the trace number;
- User keyed-in information that is not apparent from the extracts from DAISY.

---

²⁵The technical report from Moving Media on executing the project is available at 2007-232-00016, akt. 25.

²⁶2005-232-000008, akt 32, page 50

²⁷Recalculated by multiplication of 7.44
With this input, the programme should be able to generate an information package in compliance with Executive Order no. 342. It should be possible to test the version with TEA without getting any errors.

During the test, the programme achieved a migration speed of ca. 100 GB / hour when creating files to a local computer.

Test of the created information packages did not give rise to any remarks.

The budget for creating the information packages was as follows:

- Generating the migration documentation: 175 hours
- Creating the information package of the migrated records: 500 hours
- Test of information package: 175 hours
- Transfer of preservation media: 50 hours

It has not been possible to identify data about how much it cost to develop MAVDA. Therefore, it is not possible to assess how many hours have been used.

6.12.5 Conclusions

- There is no longer a need for specialist knowledge in SA regarding migration of audiovisual records.98
  - The collection is digitised
  - In the future, the supplying party will digitise new Audio / Visual materials before submitting them to the archive.
- However, private records and records that may be localised in the collection will continue to be digitised.
- The complete collection of digitised Audio/Visual materials fills 2.2 TB
- The completed information packages are registered in DAISY, and have been assigned information package IDs between 00030000 and 00039999.

Upon completion of the digitisation, we decided that the records should not be discarded as originally planned, but that they will be stored for posterity and in the event that future improved technologies make it possible to re-digitise the materials.

Resources
The table below shows the budgeted and actual expenses of the project:

<table>
<thead>
<tr>
<th>Table 44 - Budget and actual expenditure for migrating of audiovisual records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Budgeted expenses</td>
</tr>
<tr>
<td>Actual expenses</td>
</tr>
<tr>
<td>Difference</td>
</tr>
</tbody>
</table>

98 Note that per Oct. 2010, a small number of moving boxes including A/V materials have been found. This material has not been included in the digitization project, and hence there is a small digitisation task waiting.
99 Based on the Budget regarding FSC 1.5.4. Budget
100 Estimate based on the latest available budget draft
101 Annex 23 – Used hours.
102 Estimate based on the completed accounts for FSC 2005-2008
In addition to these expenses, DKK 35,000 was allocated for the expansion of the storage facility intended for storing the digitised data. We have not been able specifically to locate the actual expense for this entry in the accounts.

Of the actual hours used, 1,345 (38.4 %) hours were designated to registering the records in DAISY, while 2,157 (61.6 %) hours were used for the tasks of digitising and handling of the project. The resource usage is thus divided so as circa 66 % of the project’s expenses were used internally (internal hours) versus 34 % expenses to external consultant.

The Result
The result of the migration was as follows:

<table>
<thead>
<tr>
<th>Migration</th>
<th>Percentage of migrated records</th>
<th>Success rate as per the project’s assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio migration</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Video migration</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Film migration</td>
<td>0 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

Audio migration was successful, but we have since found a couple of moving boxes with tapes in them. These were submitted to digitisation as per 16 December 2010.

Discrepancies
The project mandate changed several times during the process, so the outcome was not quite as envisaged at the onset of the project. The reason for the change was that the knowledge acquired during the pre-project phase had decisive influence on a number of decisions. These changes include postponing the digitisation of films, revising digitised quality of sound, positioning close to the IASAs digitising standard and choosing other external vendors than those who were initially part of the project’s initial financial forecasts. The decision to work with Hein IT is a good example of this and shows the effect of choosing a cheaper solution and a higher methodology and digitisation quality than the project had originally intended.

The table below shows some of the discrepancies in level of expenses:

<table>
<thead>
<tr>
<th>Description</th>
<th>Vendor</th>
<th>DKK</th>
</tr>
</thead>
<tbody>
<tr>
<td>The project’s initial assumptions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio105</td>
<td>DBB</td>
<td>386,497</td>
</tr>
<tr>
<td>Video106</td>
<td></td>
<td>161,003</td>
</tr>
<tr>
<td>Film107</td>
<td></td>
<td>714,350</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1,261,850</td>
</tr>
<tr>
<td>Recommendation Changes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio, offer108</td>
<td>Hein IT</td>
<td>274,032</td>
</tr>
</tbody>
</table>

103 Annex 23 – Number of hours used.
104 Note that the success rate of 100 % only applies to the project’s revised goals for film migration. The FSC project had originally calculated for the migration of film; the success rate would have been set to 0%.
105 2005-232-000008, akt 58
106 2005-232-000008, akt 58
107 2005-232-000008, akt 3
Even rather late in the project, adjustments were made to the digitisation strategies. We had originally decided not to crop the black margins of digitized videos, as these margins could include picture relevant information. However, we changed our decision on this later and decided to go ahead with cropping which Moving Media performed, even though it was outside the scope of the original contract with them.

As regards films, the original decision was to digitise them together with the other audio/visual material. This decision was revised after receiving input from DFI and after internal deliberations. We decided to wait with the digitisation of film until we were able to produce a better quality product using future standards. It was decided that The Danish National Archives would build a storage facility for the film and the quoted price was ca. DKK 210,000. As per 1 November 2010, the status is that the films have not yet been digitised and the decision to improve the film storage facility has not yet been followed up on, hence the films are stored in The Danish National Archives’ regular storage rooms.

Regarding the handling of A/V material after the end of the FSC project, the project group recommended in January 2008 to:

New A/V records should be collected and registered at regular intervals, or when a sufficient amount has been collected, they should be sent to an external vendor for digitisation depending on the media type\textsuperscript{109}.

The following is a summary of the effects / consequences:

- It is necessary to allocate resources to registering, preservation-worthy assessment, contact to suppliers and storage of information packages, which can be done by use of MAVDA, which also needs to be maintained.
- Money must be allocated towards digitisation performed by vendors. The amounts depend on the number of media.
- It will cost work hours from DIARK (the Department for Digital Archiving)
- Procedures must be defined for the reception and registration
- Procedures for submitting of A/V records from units
- All units must be aware of continuing the work of exhibiting A/V records
- The Danish National Archives are made aware of the fact that the A/V records are stored in secure facilities and thus ensure ongoing accessibility to the materials\textsuperscript{110}.

\textsuperscript{108} 2005-232-000008, akt 62
\textsuperscript{109} Annex 24 – Decision proposal for on-going digitisation and cf. journal no. 2009-015893.
\textsuperscript{110} Ibid. 6, page 9.
6.13 Finances

The final chapter in the evaluation concerns the finances of the project and analyses the following:

- Budget versus actual expenditure
- Actual expenditure

6.13.1 Budget versus Actual Expenditure

All budgeted allocations to salaries were used and the excess running costs were partially transferred to salary resources. A minor outstanding between the project and the organisation was subsequently squared. The table below shows appropriations versus budget versus expenditure:

<table>
<thead>
<tr>
<th>Appropriation</th>
<th>Budget</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>DKK</td>
<td>14,900,000.00</td>
<td>13,945,562.00</td>
</tr>
</tbody>
</table>

The next table is an overview of the detailed expenditure as per 1 August 2008 (not available for 31 December 2008, the project’s actual conclusion).

<table>
<thead>
<tr>
<th>Appropriation for</th>
<th>Amount</th>
<th>Used per 1 Aug 08</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning</td>
<td>kr. 200,000.00</td>
<td>kr. 103,337.50</td>
<td>kr. 96,662.50</td>
</tr>
<tr>
<td>Digitising of audio and video</td>
<td>kr. 432,500.00</td>
<td>kr. 575,785.41</td>
<td>kr. -143,285.41</td>
</tr>
<tr>
<td>Hardware &amp; software</td>
<td>kr. 710,200.00</td>
<td>kr. 344,081.12</td>
<td>kr. 366,118.88</td>
</tr>
<tr>
<td>Storage media</td>
<td>kr. 115,000.00</td>
<td>kr.</td>
<td>kr. 115,000.00</td>
</tr>
<tr>
<td>Total</td>
<td>kr. 1,457,700.00</td>
<td>kr. 1,023,204.03</td>
<td>kr. 434,495.97</td>
</tr>
</tbody>
</table>

Note that:

- Kr. (DKK) 350,000 has not yet been transferred from operating account to the salary account.
- We budgeted with DKK 115,000 for storage media, but with the declining prices, the budget was lowered to DKK70,000.

The table below shows a detailed budget versus the actual usage\[114\], as well as the difference between them.

---

\[111\] See Annex 32 – Performance Results Contract 2005-08 Annex_to_demand_14
\[112\] Annex 25 - FSC projects budget
\[113\] See also annex 31 – FSC’s aggregate expenses 2005-2008.
\[114\] The budget and an overview of the actual expenditure do not use the same task division, and it has therefore been necessary to perform a mapping which should be considered to be adequate. Documentation for the budget, actual expenditure and mapping can be found in annex 25 - FSC project’s budget and Annex 23 – Hours used.
Table 49 - Budget versus actual expenditure

<table>
<thead>
<tr>
<th></th>
<th>Budget</th>
<th>Usage</th>
<th>Difference</th>
<th>Usage as per budget</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(hours)</td>
<td>(allocate DKK)</td>
<td>(hours)</td>
<td>(DKK)</td>
</tr>
<tr>
<td>0. Pre-project for the Format and Structure Conversion</td>
<td>2,430</td>
<td>3,192</td>
<td>-762</td>
<td>-31.36</td>
</tr>
<tr>
<td>1. Restructuring of records in new preservation structure</td>
<td>3,060</td>
<td>1,782</td>
<td>1,278</td>
<td>41.76</td>
</tr>
<tr>
<td>2. Scan preparation and scanning</td>
<td>1,425</td>
<td>200,000</td>
<td>1,296</td>
<td>103,338</td>
</tr>
<tr>
<td>3. Registering</td>
<td>571</td>
<td>315</td>
<td>256</td>
<td>44.83</td>
</tr>
<tr>
<td>4. Developing a digital description format, KonvOld and ancillary programme modules</td>
<td>2,275</td>
<td>3,724</td>
<td>-1,449</td>
<td>-63.69</td>
</tr>
<tr>
<td>5. Further development of test, CD burning and registration systems</td>
<td>1,325</td>
<td>995</td>
<td>330</td>
<td>24.91</td>
</tr>
<tr>
<td>6. Pre-migration</td>
<td>4,285</td>
<td>1,583</td>
<td>2,702</td>
<td>63.06</td>
</tr>
<tr>
<td>7. Migration including creating information packages and test</td>
<td>12,107</td>
<td>20,814</td>
<td>-8,707</td>
<td>-71.92</td>
</tr>
<tr>
<td>8. Migration of Circular 4 Records</td>
<td>250</td>
<td>224</td>
<td>26</td>
<td>10.40</td>
</tr>
<tr>
<td>9. Migration of analogue and audio visual records</td>
<td>4,075</td>
<td>432,500</td>
<td>3,347</td>
<td>575,785</td>
</tr>
<tr>
<td>10. Project management and planning, etc.</td>
<td>6,150</td>
<td>3,998</td>
<td>2,152</td>
<td>34.99</td>
</tr>
<tr>
<td>10.a Learning and training</td>
<td>1,100</td>
<td>1,583</td>
<td>-483</td>
<td>-43.91</td>
</tr>
<tr>
<td>11. Equipment</td>
<td>710,200</td>
<td>344,081</td>
<td>366,119</td>
<td>51.55</td>
</tr>
<tr>
<td>Total hours</td>
<td>39,053</td>
<td>1,342,700</td>
<td>42,853</td>
<td>1,023,204</td>
</tr>
<tr>
<td>Total man-years</td>
<td>30.11</td>
<td>33.04</td>
<td>-2.92</td>
<td></td>
</tr>
</tbody>
</table>
In total, we used 3,800 hours more than budgeted and DKK 319,496 below budget, which can be converted to an approximate deficit of DKK 530,000 - 2,369 hours or 1.8 man-years.

The numbers are calculated based on the weighted, average actual hourly rate that the staff received during the project, namely DKK 223.76, cf. annex 4 – Memo on man-years and hourly rates.

When looking at the individual entries in the table 47 – Budget versus actual usage – the largest percentile or hourly positive deviations from the budget concern:

- **Restructuring of records to a new preservation structure** ended up 1,278 hours cheaper than expected. It is difficult to know why, but one hypothesis can be presented:
  - Simple incorrect estimation in the project planning due to fewer restructurings than anticipated, which was not properly investigated in the pre-project.

- **Scanning preparation and scanning cost** DKK 96,663 or almost 50 % cheaper, since the project management estimates were overly cautious, which also led to finding a cheaper supplier than originally expected.

- **Registering** landed at 256 hours cheaper than expected due to the following reasons:
  - Better and more efficient tools than expected for use of conjoining A/Vs and files from the scanned documents.
  - Fewer than anticipated document registrations
  - Cf. Chapter 6.10.1.3.1, Fejl! Henvisningskilde ikke fundet, page 102, it is possible that some of the manual registering were performed a bit (too) quickly at the expense of quality. One of the consequences is that the naming policy is not complied with 100%, which makes machine handling completely impossible.

- **Pre-migration** ended up 2 man-years cheaper because:
  - Changed work procedures transferred a greater part of the hours to tasks
  - Development of the digital description format KonvOld and ancillary programme modules.
  - Migration including creating information packages and test
  - It was possible to re-use pre-migration tools for all records because of the occurrence of the same problems several times in various parts (fewer structures than anticipated).
  - It was easier than anticipated to render KonvOld more flexible and thus handle more structures.

- **Project management, planning etc.,** was DKK 3,252 or 81.34 % cheaper because some to the tasks were transferred later on in the process and because the project manager went on maternity leave and therefore no hours were registered on that account for the remaining 8 month of the project.

The largest negative discrepancies from the budget were:

- **Development of the digital description format KonvOld and ancillary programme modules** was just over one man-year more expensive than anticipated because:
  - Part of task 6 “Pre-migration” was done here.

There were more data types and code pages along with many faulty originally assumed; cf. Chapter 6.10,
- Migration and Preparation of Metadata, page 84
- More time was used than anticipated on improving user-friendliness and presentation of errors (via formatting of logs and showing on screen)

**Migration including creation of information packages and test** exceeded the budget by just under 72% or 6.7 man-years for three reasons:
- A great proportion of task 6, Pre-Migration, was performed within the scope of this task.
- The collected yield output of the machines and the programmes was insufficient compared to the budgeted estimates.
- There was – perhaps – a larger discrepancy between the data description in the documentation and the actual data. This could be because the spot check sample was too small, or perhaps because it took longer time than anticipated to correct errors.

**10.a Learning and Training** ended up at 483 hours or 30.51% more expensive than expected since the Management of the States Archives decided, towards the end of the project, to send several of the project’s temporary staff members on an XML training course.

### 6.13.2 Actual time used

The table and diagram below show the final internal time used in the project as well as the percentage of time used on each task. Note that sub-divided tasks that are of similar type are tied together alphanumerically and by colour coding.

<table>
<thead>
<tr>
<th>Task</th>
<th>Hours used</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Pre-project</td>
<td>3,192</td>
<td>7.45%</td>
</tr>
<tr>
<td>1. Restructuring records to new preservation structures</td>
<td>1,782</td>
<td>4.16%</td>
</tr>
<tr>
<td>2a. Developing scanning programmes</td>
<td>301</td>
<td>0.70%</td>
</tr>
<tr>
<td>2b. Preparing for scanning and handling</td>
<td>995</td>
<td>2.32%</td>
</tr>
<tr>
<td>3. Registering in DAISY and MARY</td>
<td>315</td>
<td>0.74%</td>
</tr>
<tr>
<td>4. Developing a digital descriptive format, KonvOld and ancillary programme modules</td>
<td>3,724</td>
<td>8.69%</td>
</tr>
<tr>
<td>5. Further development of test, CD burning and registering systems</td>
<td>995</td>
<td>2.32%</td>
</tr>
<tr>
<td>6. Migration method</td>
<td>2,279</td>
<td>5.32%</td>
</tr>
<tr>
<td>7a. Pre-migration (periods A and B)</td>
<td>1,583</td>
<td>3.69%</td>
</tr>
<tr>
<td>7b. Migration of filing systems (period A)</td>
<td>2,049</td>
<td>4.78%</td>
</tr>
<tr>
<td>7c. Migration of registry (period A)</td>
<td>7,205</td>
<td>16.81%</td>
</tr>
<tr>
<td>7d. Migration of filing systems (period B)</td>
<td>2,493</td>
<td>5.82%</td>
</tr>
<tr>
<td>7e. Migration of registry (period B)</td>
<td>4,714</td>
<td>11.00%</td>
</tr>
<tr>
<td>7f. Quality control (periods A and B)</td>
<td>2,071</td>
<td>4.83%</td>
</tr>
<tr>
<td>8. Migration of Circular no. 4 records</td>
<td>224</td>
<td>0.52%</td>
</tr>
<tr>
<td>9. Migration of analogue and audio/visual records</td>
<td>3,345</td>
<td>7.81%</td>
</tr>
<tr>
<td>10. Preservation media transfer</td>
<td>6</td>
<td>0.01%</td>
</tr>
<tr>
<td>11a. Project management and reporting</td>
<td>2,513</td>
<td>5.86%</td>
</tr>
<tr>
<td>11b. Project group meeting</td>
<td>588</td>
<td>1.37%</td>
</tr>
<tr>
<td>11c. Documentation and journaling</td>
<td>897</td>
<td>2.09%</td>
</tr>
<tr>
<td>12. Knowledge building and training</td>
<td>1,583</td>
<td>3.69%</td>
</tr>
<tr>
<td><strong>In Total pre-project and actual project</strong></td>
<td><strong>42,853</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
The biggest expense stems from tasks related to the migration of periods A and B (dark blue in diagram). This amounted to just below 50% of the total time consumed internally in the project.

The migration was a production process that was partly done manually requiring individual processing of each single information package.

Going forward, much of the expenses related to migration will be possible to eliminate or reduce significantly since it will be possible to perform simple technical migrations of data and metadata.

However, there still remain a number of open issues that will have to be resolved before enabling full automatic migration, cf. chapter 5.1.3, Draft to list of outstanding tasks related to the FSC project’s aims and the condition of the archival holdings., page 29.
The following table provides a glossary and explanation of abbreviations.

Please note that the terms 'digital' / 'electronic'; and 'migration' / 'conversion' are used synonymously in this report. We prefer the former of the terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive end date</td>
<td>End date for submitted data. Format: ccymmd. See also ARKVER.TAB.</td>
</tr>
<tr>
<td>Archive start date</td>
<td>Start date for submitted data. Format: ccymmd. See also ARKVER.TAB.</td>
</tr>
<tr>
<td>Archive View</td>
<td>Archive View is a function in the test tool TEA, which presents the content of an information package divided into media. The function makes it possible to:</td>
</tr>
<tr>
<td></td>
<td>• Test file structures</td>
</tr>
<tr>
<td></td>
<td>• Test and show general information (GENINFO.TAB)</td>
</tr>
<tr>
<td></td>
<td>• Show information package information (ARKVER.TAB)</td>
</tr>
<tr>
<td></td>
<td>• Show creator of information (SKABER.TAB)</td>
</tr>
<tr>
<td>ARKVER.TAB</td>
<td>ARKVER.TAB is a help table in an information package that described the version through 7 fields:</td>
</tr>
<tr>
<td></td>
<td>1. Information package number</td>
</tr>
<tr>
<td></td>
<td>2. Code denoting submission type</td>
</tr>
<tr>
<td></td>
<td>3. Unique media name including information packages metadata and help tables</td>
</tr>
<tr>
<td></td>
<td>4. Information package number of earlier submissions</td>
</tr>
<tr>
<td></td>
<td>5. The official name of the system, in which all abbreviations are removed</td>
</tr>
<tr>
<td></td>
<td>6. Start date for submitted data</td>
</tr>
<tr>
<td></td>
<td>7. End date for submitted data</td>
</tr>
<tr>
<td>AVID (information package ID)</td>
<td>A unique information package number defined as an 8 digit number.</td>
</tr>
<tr>
<td>Create end date</td>
<td>The time when the archive creator stops entering data into the system. Format: ccymmd. See also SKABER.TAB.</td>
</tr>
<tr>
<td>Create start date</td>
<td>The time when the archive creator starts to enter data into the system. Format: ccymmd. See also SKABER.TAB.</td>
</tr>
<tr>
<td>Creator</td>
<td>Name for every individual authority that has entered data into the system. If a public authority’s organisation unit have not all used the system or if they have not started using the system at the same time, the organisational unit is stated as archive creator. An organization unit can be a section, an office, a department etc. See also SKABER.TAB.</td>
</tr>
<tr>
<td>DAISY</td>
<td><em>Dansk Arkivalie Information system (DAISY)</em> is The Danish</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>National Archives’ digital registry including information on the archive creator and the records that are part of The Danish National Archives’ collections.</td>
<td></td>
</tr>
<tr>
<td>Database View</td>
<td>DatabaseView is a function in the test tool TEA that shows the archive’s registry (tables, ARK files) in the form of an E/R-diagram. Each registry shows the appertaining fields with mark-up of primary keys and relations to other tables.</td>
</tr>
<tr>
<td>DEA Distribution of Electronic Records – burning programme</td>
<td></td>
</tr>
<tr>
<td>&quot;Digital” is used instead of &quot;electronic”, except when used for formal references to the FSC project, which used the latter term.</td>
<td></td>
</tr>
<tr>
<td>Document View</td>
<td>DocumentView is a function in the test tool TEA that can show each document in TIFF view.</td>
</tr>
<tr>
<td>DOKMAP.TAB (help table in an information package)</td>
<td>DOKMAP.TAB is a help table in an information package showing the relation between each document and its position in the storage media submitted to the public archive. The table includes a row of the following three columns:</td>
</tr>
<tr>
<td></td>
<td>1. Name or ID denoting a unique document</td>
</tr>
<tr>
<td></td>
<td>2. Name of storage media in which the documents are stored</td>
</tr>
<tr>
<td></td>
<td>3. Sub-directory in which the documents can be found</td>
</tr>
<tr>
<td>E-number Reference number for the identification of a digital submission (see also heuristic unit)</td>
<td></td>
</tr>
<tr>
<td>Filing system number Number that uniquely identifies a subject in an archive series</td>
<td></td>
</tr>
<tr>
<td>FILMAP.TAB</td>
<td>FILMAP.TAB is a help table in an information package that lists all tables that are part of the information package, including the help tables GENINFO.TAB, SKABER.TAB and DOKMAP.TAB as well as the metadata file. The tables include these two columns:</td>
</tr>
<tr>
<td></td>
<td>1. File or library name</td>
</tr>
<tr>
<td></td>
<td>2. Unique name of media in which the file or library is placed.</td>
</tr>
<tr>
<td>Geninfo</td>
<td>Geninfo is a document attached to an information package. The document includes information on the identifications (media and library identification), a description of which documentation is included in the submission (e.g. user manual, legal information, screen dumps, user interfaces, E/R diagrams, etc.), and where the documentation can be found.</td>
</tr>
<tr>
<td>GENINFO.TAB</td>
<td>GENINFO.TAB  is a help table in an information package indicating location (library name) including documents with the information package’s general information. The table includes rows with the following three columns:</td>
</tr>
<tr>
<td></td>
<td>1. Unique name of the media where the library is placed.</td>
</tr>
<tr>
<td></td>
<td>2. Name of the library including the specific document.</td>
</tr>
<tr>
<td></td>
<td>3. Description of each individual document, including general information.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Heuristic unit</strong></td>
<td>The heuristic unit denotes records which can be collectively identified by:</td>
</tr>
<tr>
<td></td>
<td>1. The heuristic unit’s name</td>
</tr>
<tr>
<td></td>
<td>2. The name of the attached participant</td>
</tr>
<tr>
<td></td>
<td>3. The heuristic unit type (Component, E-system, etc.)</td>
</tr>
<tr>
<td></td>
<td>4. Reference (e.g. E-number)</td>
</tr>
<tr>
<td>IP</td>
<td>Information package. Equivalent to Information Package in OAIS</td>
</tr>
<tr>
<td>ITM</td>
<td>IT &amp; Method department was during the FSC project The Danish National Archives’ section for the support of digital preservation.</td>
</tr>
<tr>
<td>KonvOld</td>
<td>KonvOld was the central migration application for the migration of periods A and B records. Besides the basic functionalities, the programme included a number of modules intended for handling variations in character set code pages, data structures, etc.</td>
</tr>
<tr>
<td>KORAID</td>
<td>Converting raid used as an interim storage during the migration process.</td>
</tr>
<tr>
<td>MARY</td>
<td>MARY (Methods Archival Registration sYstem) is The Danish National Archives’ preservation database used for surveillance of the digital collection.</td>
</tr>
<tr>
<td>MAVDA</td>
<td>Mass production of Information packages of Digital Records. This programme was developed to create information packages of the digitized audio/visual materials.</td>
</tr>
<tr>
<td>MedieID</td>
<td>A unique number assigned to each media (CD-R) an information package is made up of.</td>
</tr>
<tr>
<td>MetadataView</td>
<td>MetadataView is a function in the test tool TEA showing the content of the meta data file in text or table format as well as validating (parsing) the XML structure.</td>
</tr>
<tr>
<td>Migration (converting)</td>
<td>The transformation of data performed when older digital records formats and structures are migrated so as to render them compliant with the newest demands for preservation and accessibility. The term migration is used instead of “conversion” except when specific reference is made to the FSC project, which has used the term conversion.</td>
</tr>
<tr>
<td>Participant</td>
<td>Participant includes – or is synonymous with – authority, institution, archive creator and archive responsible. In The Danish National Archives’ preservation database, DAISY, Participant denotes archive creator.</td>
</tr>
<tr>
<td>Provenance information</td>
<td>Provenance information (Provenance Information) – equivalent to the OAIS terminology for the information documentation the preservation worthy data’s origin and history, besides any changes that may have been performed and who has owned them over time. Provenance information is found in the information package both as administrative and systematic documentation. The provenance information is also the documentation prepared after submission to an archive in connection with test and migration. An example of</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>Provenance number</td>
<td>A number allocated by The Danish National Archives to all archive creators. See also provenance information.</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>It was possible to reconstruct data after the migration of an archival information package from periods A or B, for which a programme, TGP02, was developed in the TGP project based on uploading of data in the MS SQL database. The tool was not used often, since the project had several test and viewing tools available.</td>
</tr>
<tr>
<td>SB</td>
<td>The State Library (Statsbiblioteket) stores one of the The Danish National Archives’ three editions.</td>
</tr>
<tr>
<td>SKABER.TAB</td>
<td>SKABER.TAB is a help table in an information package that describes the archive creator(s) by providing the following fields:</td>
</tr>
<tr>
<td></td>
<td>1. Name of all individual authorities that have entered data into the system. The time when the archive creator started to enter data into the system.</td>
</tr>
<tr>
<td></td>
<td>2. The time the data creator stops entering data into the system.</td>
</tr>
<tr>
<td>System name</td>
<td>The system’s official name without any abbreviations</td>
</tr>
<tr>
<td></td>
<td>One field in ARKVER.TAB.</td>
</tr>
<tr>
<td>TableView (function in test tool TEA)</td>
<td>TableView is a function in the test tool TEA showing each individual registry at record level. The function provides for the following:</td>
</tr>
<tr>
<td></td>
<td>• Test of data types</td>
</tr>
<tr>
<td></td>
<td>• Test of keys</td>
</tr>
<tr>
<td></td>
<td>• Test of documents</td>
</tr>
<tr>
<td></td>
<td>• Viewing of codes</td>
</tr>
<tr>
<td></td>
<td>• Browsing and searching of records</td>
</tr>
<tr>
<td>TEA</td>
<td>Test of Electronic Archives (records) – tool developed by The Danish National Archives for testing submissions.</td>
</tr>
<tr>
<td>TGP</td>
<td>The aim of the accessibility project was to develop an accessibility application (SOFIA) based on the migrated information package from the FSC project. The projects were simultaneous.</td>
</tr>
</tbody>
</table>