Evaluation of a Large Migration Project

Alex Thirifays
Danish National Archives
Rigsdagsgården 9
1218 København K
+45 33 92 23 69

alt@ra.sa.dk

Anders Bo Nielsen
Danish National Archives
Rigsdagsgården 9
1218 København K
+45 33 92 83 26

abn@ra.sa.dk

Barbara Dokkedal
Danish National Archives
Rigsdagsgården 9
1218 København K
+45 33 95 46 89

bd@ra.sa.dk

ABSTRACT

The Danish National Archives (DNA) has ingested structurally heterogeneous public digital records since 1973. The year 2004 saw the creation of a new preservation standard into which it was decided to migrate the above mentioned archival holdings. The main objectives of this operation were to save data from technological obsolescence and to reduce the cost of both access and future migrations by streamlining the collection.

The project costs approximately 30 FSCs (one 'FSC'—Format and Structure Conversion—is the way the project's project management measured 1 person-year, and equals 1,291 person-hours). The total sum of purchasing software, hardware and external services amounted to around 135.000 Euros.

The project migrated data from both relational and hierarchical databases (for instance ERDMS and registries), and included the digitisation of audio, video as well as paper documentation. The registries counted for example the first Civil Registration System from 1968 and the State Tax Administration's final equation from 1970. Data and documentation made up a total of about 1.7 TB, consisted of 11,187 files scattered in almost 200 different structures, and constituted more than 2,000 information packages (IPs).

The overall technical objective of the migration was defined by the aforementioned preservation standard, which required:

- Common format for data files
- Common structure of documentation, metadata and documents
- Common format for documents (TIFF).

The project's main objectives were achieved, since all records were migrated, except the film collection. The goal of making access and future migrations easier was also reached, but a fully automatic migration of the collection is not yet entirely possible. The overall conclusion is that the migration project, which, to our knowledge, is the first of its kind, was of very high quality, both in terms of planning, execution and product.

A few main conclusions are:

- Standardisation of data is a prerequisite for an economically sound digital preservation: It took about 70 times longer to migrate an older, non-standardised IP, than a newer, standardised one.
- Inadequate feasibility studies remind us that the timing in digital preservation saves money: The condition of the magnetic tapes was examined through spot checking 10 years prior to the migration project and found satisfactory, which turned out to be an erroneous, expensive conclusion.
- The technical infrastructure suffered from a number of shortcomings and late decisions, resulting in precious loss of time.

Categories and Subject Descriptors

E.1.5 =E. Data, E.1 DATA STRUCTURES - Records

General Terms

Design, Documentation, Management, Measurement, Standardization, Verification.

1. INTRODUCTION

In the Performance Contract 2009-12 between The Danish National Archives and The Ministry of Culture it was agreed that the former write a self-evaluation of the migration[1] project, the Format and Structure Conversion Project (FSC), which was conducted by The Danish National Archives between 2005 and 2008.

The Format and Structure Conversion 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 IT-systems[2].

This Executive Order is now outdated and has been replaced by a new order[3].

A new "format and structure conversion" is hence expected. Therefore the experiences gained by the Danish National Archives from the FSC Project will serve as a valuable data basis for the planning of future migrations.

---

[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.*

[2] In Denmark, (technical) Executive Orders constitute the submission standards.

[3] Annex 2 in the evaluation report – *Executive Oder no. 1007 of 20 August 2010 on information packages.*

This article aims to sum up the conclusions of the evaluation report[4], which is the first phase of a project[5] designed to develop strategy proposals for logical preservation and implementing concrete planning of future migrations within the framework of the Performance Contract 2009-12.

2. FSC PROJECT DESCRIPTION

2.1 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[6]. At the end of the 1990s, it also became possible to receive electronic records and document management systems (ERDMS).

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 format- 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 was decided that all digital records should be migrated to the *preservation standard*, which coincided with the *submission standard*[7] applicable at that time. The standard describes which data structure and formats the records should be preserved in. It also describes what documentation and metadata should accompany each IP.

2.2 Objective and Introduction

The FSC project was a transformation[8] 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 calendar years to complete – from 2005-2008 – and included migration of all formats and structures that did not meet the requirements set forth in the aforementioned preservation standard. Briefly, 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' person-years and the aggregate expenses for the purchase of software, hardware and external services amounted to around 135.000 Euros. 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.

2.3 The Records

The Danish National Archives' holdings[9] of non-standardised records, which were to be migrated, originated from three periods:

- **Period A:** submitted before the 1998 preservation standard[10] 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 preservation standard. 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 packages demanded many resources as well.
- **Period C:** submitted in compliance with Circular No. 4 from 2000 and, after 2004, with Executive Order no. 342[11]. These records encompassed 942 information packages, but the number was constantly increasing as new submissions were received. 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, thus enabling a high degree of automated migration.

---

[4] http://www.sa.dk/media(3649,1030)/Final_report%2C_ Evaluation_of_the_Format_and_Structure_Conversion_Project. doc

[5] Preservation Planning Project (PPP)

[6] "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.

[7] Annex 5 in the evaluation report – Executive Order no. *342 of 11 March 2004 on information packages worthy of preservation data from IT-systems.*

[8] 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.*

[9] Note that the project was conducted by The Danish National Archives, and that the records were solely records from public administration and courts of law.

[10] Cf. *Electronic Archiving – The Danish National Archives system demands,* Danish National Archives, 1998, ISBN8774971778.

[11] These technical law texts represented quite similar submission standards, and, thus, preservation standards.

Table 1 - Overview of records in the FSC project

Record type	Submitted during the period	Data size (GB)	Number of files	Number of IPs
Period A[12]	1973-1998	171	3,109	650
Period B[13]	1998-2000	419	8,078	641
Period C[14]	2000-	1,187	-	942
Total	-	**1,777**	**11,187**	**2,233**

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 ERDMS for a 5-year period.

It has not been possible to accurately allocate the time used on each period separately, since some of the tasks performed were 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 the time spent. However, Table 2 shows a relatively accurate picture of the time spent.

Table 2 - Allocation of time used on each record type (A, B and C)

Record type	Hours	FSC-years	Time allocation (%)	Fraction of total collection (%)[15]
A	17,889	13.79	53	9.6
B	12,853	9.91	38	23.6
C	3,184	2.45	9	66.8
Total	**33,926**	**26.16**	**100**	**100**

The time spent here is not equivalent to all the time spent 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 spent on migrating e.g. audio/visual records is not included in the breakdown.

2.4 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[16]

---

[12] Data files on tape and paper documentation.

[13] Data files on CDs and disks and paper documentation.

[14] ERDMS and registries.

[15] This column's numbers are based on each collection's data size in GB, cf. Table 1.

[16] According to the Executive Order no. 342, it is optional to choose between fixed and variable record lengths. However the FSC project systematically chose fixed record lengths.

- 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 each submission, including context information, reference information and descriptive information)
  - o Documents
  - o Tables
  - o Metadata, description of the information package's tables with field description and the mark-up of relations between them

- **Common format for documents** (TIFF)
  - o Digitisation of paper documentation
  - o Migration of digital documents (e.g. from Word to TIFF)

2.5 The Project

Most of the project required specialized knowledge, which only existed in the National Archives Preservation & Disposal Department; hence most of the tasks were performed internally. Some 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[17]. All measures of the project's progression were performed in the archival database DAISY[18], while the registering of the records was performed partially in MARY[19] and partially in the 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.
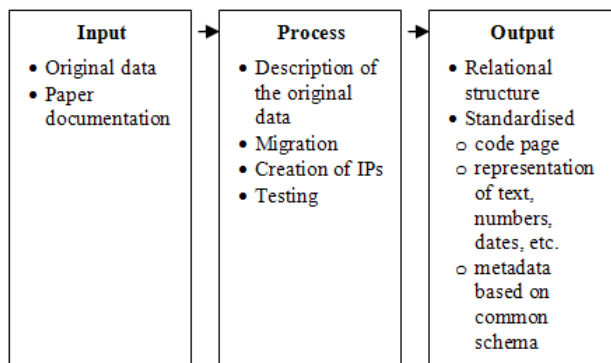
This infrastructure formed the framework for the migration process, which can be simply illustrated as follows:

---

[17] Cf. *New Implementations of the Media Strategy for The Danish National Archives, cf.* Jr. no. 2004-360-0005.

[18] DAISY is The Danish National Archives digital registry in which all user-oriented information about the creators of the IPs (authorities, companies, private individuals, etc.) as well as the IPs themselves is lodged. This database is the public's access to the holdings.

[19] MARY is The Danish National Archives preservation database used to monitor the condition of the digital collections. This database is for internal use only.

Figure 1 - FSCs migration process

Input	Process	Output

• Original data
• Paper documentation
	• Description of the original data
	• Migration
	• Creation of IPs
	• Testing
		• Relational structure
		• Standardised
		  ○ code page
		  ○ representation of text, numbers, dates, etc.
		  ○ metadata based on common schema

The project went through a series of phases as described below.

1. Pre-project

The pre-project included the surveys that were prerequisite for estimating the budget of the actual project.

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.

3. Scanning

The task included the scanning preparation and the scanning of paper documentation (stored in ca. 400 archive boxes), which belonged to the old submissions. The paper documentation was the prerequisite for analysing and migrating the older records.

4. Registration

The task dealt with a basic and comprehensive registration and continuous updating and maintenance of data in the preservation database MARY used for monitoring the holdings. Moreover, the registration data in DAISY was made accessible in order to enable users to search the records.

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 meet the preservation standard's requirements. The digital description standard was developed in XML with its own schema, and was used for 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.

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 a central migration application (KonvOld), besides a number of modules intended to handle the variations in code pages, data structures, etc.

7. Further development of test, CD-burning and registration systems

The task included the further development of a number of already existing systems. These included programmes used for testing (TEA)[20], preservation media transfer (DEA)[21] and the export of automatically generated registration data to the preservation database, MARY.

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, migrated, tested, proofread, and placed on preservation media along with their corresponding digitised paper-based documentation.

An important part of this task was to prepare 2 documents establishing respectively the principles any migration would follow, and the routines any Convertor[22] 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.

The migration was subdivided into three tasks:

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 central migration tool, KonvOld. The records also could not be handled by the digital description format. Hence there was a need for the development of record specific *ad hoc* tools and for a subsequent pre-migration, which prepared each unique record for the final migration using KonvOld. 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 the migration to proceed automatically.

b. Migration of newer records

The task included the machine migration of a number of newer information packages supplied in accordance with the then recent preservation standard, Circular No. 4, and thus almost compliant with the preservation standard in force at the time (Executive Order no. 342).

c. Migration of analogue audio/visual records

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

---

[20] Test af Elektroniske Arkivalier - Test of Electronic Records
[21] Distribution af Elektroniske Arkivalier - Distribution of Electronic Records (CD burning and transfer programme)
[22] Convertors designates the staff used for migrating and testing.

3. Conclusions

3.1 Achieving the formal goal

3.1.1 Appropriation, budget and expenditure

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

Table 3 - Appropriation versus budget versus expenditure

	Appropriation	Budget	Expenditure
$	2,900,000	2,700,000	2,800,000

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).

3.1.2 Records

Number

The project's main goal was achieved since all records were migrated, with the exception of the Danish National Archive's film collection, which had not been digitised. This was however accounted for in the revised project plan.

Quality and authenticity

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 – had not been tested correctly and in consequence deviated from their original, appertaining preservation standard *before* the FSC project migrated them. This meant that some of those records also deviated from the preservation standard *after* migration. Maintaining the authenticity of the data instead of modifying them and stating how they were modified was a priority in accordance with the project management's preservation policy decisions. The partial lack of standardisation of those records is, hence, not due to poor migration, but to one of the Danish National Archive's preservation principles.

3.2 User evaluation of the migrated records

Concurrently with the migration project, an accessibility project developed an Access tool (SOFIA), which was launched in 2008. As of July 2010[23], this tool had been used to render 227 Information Packages 790 times.

No user feedback has been gathered in a systematic manner, neither regarding the access tool nor the accessed IPs. The feedback that has been received shows however that a number IPs are faulty, but according to the Access Department, the errors are easily corrected and modest in number.

The types of errors vary, but the most common ones are:

- Incorrectly converted data types, e.g. TIMESTAMP to STRING
- Inaccessible .TIFF context documentation documents

---

[23] No newer numbers are available.

- Missing end date of the IP
- Occurrences of duplicates of document ID
- Missing validation of the IP after migration

3.3 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. The conclusions are substantiated in more detail in the evaluation report.

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 following conclusions, which question a number of issues, do not dispute the overall conclusion, which is that the project was of very high standard regarding planning, execution and outcome.

3.3.1 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 30 man-years) and the project's crucial prerequisites: the rescuing of magnetic tapes (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.

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 analysis of records and the development of prototype programmes.
- **Digitisation of film, budgeted at DKK 700,000, was scrapped.** This decision was considered prudent by the project, as it was estimated that the analogue media the films were stored on were 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 purchase of 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 time of the migration itself had been satisfactory.
- **Completing the handling of period C records was postponed** until after the end of the project.

- **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' organizational entity, Access, has been very satisfied with the migrations. A log has been created including inappropriateness and errors that have been detected.
- **The Danish National Archives 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 migrated information packages.

Preservation policy principles can be 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. 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:
- **In the recovery project**, which migrated damaged magnetic tapes, it cost on average 6.874 Euros to migrate just one magnetic tape, while the price for migrating a magnetic tape of good quality was 46 Euros (factor 149).
- **Unique records**. It took 70 times longer time to migrate older, non-standardised records (cf. period A) than newer, standardised ones (cf. 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. Measured per file, it took ca. 158 times longer to migrate the older records compared to the newer period C records (3.16 hours versus 0.02 hours). Per GB, it took 253.5 times longer to migrate one older record than a newer one (63.36 hours versus 0.25 hours).

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. There are, however, still examples of records that cannot be processed in an automated fashion by migration tools.
- **System development of tools for non-standardised records is expensive.**
  - Development of tools needed to handle the older records (period A and B), accounted for over 80 % of the total system development costs.

- 175 programmes were developed to handle the 167 records with unique structures and formats.
- **The flexibility of the central migration application (KonvOld) allowed for easier addition of modules and thereby ease of handling many formats.** Only 167 out of 1,291 records (just under 13 %) were to be pre-migrated by *ad hoc* tools before they were suitable as data input to KonvOld. The remainder of the data could be used directly as standard input. KonvOld handled thus the automated migration of ca. 87 % of records. This was only possible due to the ongoing, iterative optimisation and expansion of KonvOld.
- **A standardised, system-independent format was chosen for preparing the digital descriptions:** An XML schema. This made it possible to use alternatives to InfoPath tools used for reading and maintaining the digital descriptions.
- **Two preservation databases are one too many.** Having to do (double) registering in MARY and DAISY required much manual resources - it took ca. 1 person-year. It is, however, important to note that it might not have been economically feasible to merge the two systems within the scope of the FSC project.

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 bullet below, '**Feasibility studies'**, pertaining to an insufficient random check of the magnetic tapes.
- **Test tools must test exactly what is described in the preservation standard,** otherwise, the records are not tested in compliance with the standard, which makes 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 describe to what degree it lives up to the preservation standard.

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.
- **Tools performed poorly in certain contexts.**
  - When using InfoPath for the digital descriptions, the speed of loading 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.

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 testing when receiving records.

- **1.6 % of the 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.
- **The quality of the input data was 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, otherwise automatic processing is impossible.
- **The quality of the documentation was not satisfactory** and did not always give a complete description of the records. This meant that many resources were used to examine the formats, structures and content of the records in order to optimise the documentation in its digital form.

3.3.2 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.

- **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 working procedures. The documents were discussed and updated at the weekly project meetings.
- **Quick access to information.** Several factors were of great value to the daily work within the project, namely: online access to the majority of the project's information, the 2 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* the standardization of the records, that knowledge is no longer needed.
- **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:
  - **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.
  - **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 migration principles and work-flows were invaluable in the daily work. They ensured a uniform, high quality and efficient processing of the records.

- **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.
- **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. These series were semantically similar, enabling recognition and identical treatment, and the sharing of responsibility partially explains the diligence with which the work was performed.
- **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.

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. One example illustrates this:
- **Feasibility studies. Several instances of insufficient feasibility studies which, if done, could have provided potential savings and efficiencies.**
  - **The condition of the magnetic tapes had already been examined in 1995 through a random check. The tapes' condition was found to be suitable, which it was not**. Therefore, a migration of the media's data was not initiated at that time, which in turn resulted in the huge expenses incurred to salvage the tapes 10 years later (cf. on average 6.874 Euros to migrate just one damaged magnetic tape).
  - **The pre-project did not have enough time to examine a suitable sample of the data to be migrated**, which meant that it was not possible to produce a detailed requirements specification for KonvOld, which had to be developed simultaneously with the migration process (iteration). The fact that this turned out to be a very advantageous solution could not have been predicted.
  - **Preparation for the system development was perhaps inadequate.** If there had been set aside more time for surveying the market and trying out various commercial or open source tools, it might have been possible for the FSC project to save resources on system development.

Lengthy decision-making processes

- **In certain instances, the chain of command was bureaucratic and impeded the quick implementation of decisions.**
  - **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, the FSC project did not succeed in getting more resources for backup and hence the 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 is both acceptable and necessary.**

Outsourcing

- **Outsourcing of tasks does not entail freeing up of resources to internal activities pertaining to the task.** The major share of the expenses for scanning of documentation, which was outsourced, was taken from sub-tasks performed in-house (55 %).
- **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.

Insufficient communication among the various projects

Despite a common steering group for the FSC project and the concurrent Accessibility project (TGP), 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. It was 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 tool, SOFIA, by around 1 year.
- During the process of buying hardware, the IT department had a fall out with the FSC project staff. A better cooperation would have ensured a more powerful hardware performance than what was used earlier in the project, which would have ensured faster migration.

- There was inadequate integration between the components for preservation (FSC), registration (Access department with DAISY) and accessibility (TGP). It was e.g. not possible to automatically import data from KonvOld to DAISY.

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.

4. REFERENCES

[1] ANSI/ARMA, 2007, *The Digital Records Conversion Process: Program Planning, requirements, Procedures*, ARMA International, 13725 West 109th Street, Suite 101, Lenexa, KS 66215, 913.341.3808.

[2] Consultative Committee for Space Data Systems, 2002, *Reference Model for an Open Archival Information System (OAIS),* CCSDS Secretariat, Space Communications and Navigation Office, 7L70, Space Operations Mission Directorate, NASA Headquarters, Washington, DC 20546-0001, USA Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington. DOI = http://public.ccsds.org/publications/archive/650x0b1.PDF

[3] Consultative Committee for Space Data Systems, 2004, *Producer-Archive Interface Methodology Abstract Standard*, CCSDS Secretariat, Space Communications and Navigation Office, 7L70, Space Operations Mission Directorate, NASA Headquarters, Washington, DC 20546-0001, USA Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington. DOI = http://public.ccsds.org/publications/archive/651x0m1.pdf

[4] CRL, The Center for Research Libraries, OCLC Online Computer Library Center, Inc., 2007. *Trustworthy Repositories Audit & Certification: Criteria and Checlist*. 6050 South Kenwood Avenue, Chicago, Illinois, 60637-2804 USA

[5] DNA, The Danish National Archives, 1998, *Electronic Archiving – The Danish National Archives system demands,* ISBN8774971778.

[6] Dollar, Charles M., 2000. *Autrhentic Electronic Records: Strategies for Long-Term Access.* Cohasset Associates, Inc. Chicago, Illinois, USA

[7] ISO 15489-1:2001, *Information and documentation -- Records management -- Part 1: General*

[8] ISO 23081-1:2006, *Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles*

[9] ISO/DIS 13008, 2011, *Information and documentation -- Digital records conversion and migration process*

[10] ISO/IEC 29121:2009, *Information technology -- Digitally recorded media for information interchange and storage -- Data migration method for DVD-R, DVD-RW, DVD-RAM, +R, and +RW disks*