Costs of Digital Preservation

Project Report for Phase 2

Ulla Bøgvad Kejser (The Danish Royal Library),
Anders Bo Nielsen and Alex Thirifays (The Danish National Archives)

February 2011

Table of Contents

1 Resumé (Executive summary)

1.1 Background

The project was created as a consequence of a proposal to the Danish Ministry of Culture's "Committee on Digital Preservation" from The Danish Royal Library (KB), The State and University Library (SB) and The Danish National Archives (SA) to prepare a cost model for the preservation of digital records.

The project so far consists of two phases:

In the first phase of the project (2009) we developed a cost model which is designated "Cost Model for Digital Preservation" (CMDP). The underlying method is based on two elements: the archive standard Open Archival Information System (OAIS) Reference Model[1] and the activity-based cost model, ABC[2].

The OAIS model describes the various functions in a long-term digital preservation system from the ingestion of data into the preservation institution's archives, until the data are preserved and made available to the institution's users. We used the OAIS model to identify functions and divide them into delineated cost-critical activities. The calculations solely include activities that take a minimum of 1 person-week to complete. We then identified the cost parameters (variables) related to the individual activities, and expressed them as formulæ in a spreadsheet.

Costs are stated as the time it takes to perform an activity multiplied by the hourly price. In addition, there are the costs of any acquisitions. In principle, CMDP must include all direct costs: development, establishment and operating expenses, as well as indirect administration costs (overhead). So far, however, overhead has only been implemented in the model indirectly as the salary level. Other outstanding issues with regard to the distribution of costs are the incorporation of depreciation in the model, as well as determining whether the costs of general system support, for example the costs of infrastructure establishment, should be categorised under the OAIS archive functions, or under the support function of Common Services. In time, the model will also be able to take financial adjustments, for example for inflation, into account.

Due to limited resources it has been necessary to operationalise the cost model in stages.

In the first phase of the project we compiled the costs of logical preservation and focused on the costs of the preservation strategy that are based on migration. The first phase of the project was reported on in "The Cost of Digital Preservation, November 2009"[3]. The report presents an introduction to the subject of the costs of digital preservation and reviews existing cost models. The report also describes the method we have used to develop the cost model. Finally, the report describes how we have operationalised the model for logical preservation by migration, and presents the results of our test of the model on existing cost data from migration projects conducted by The Danish National Archives.

---

[1] http://public.ccsds.org/publications/archive/650x0b1.pdf

[2] http://en.wikipedia.org/wiki/Activity-based_costing

[3] The CMDP spreadsheet, project report and other documentation from the first phase of the project can be downloaded from the project website: www.costmodelfordigitalpreservation.dk.

In the second phase of the project, which took place in 2010, we focused on the costs related to the ingest of digital collections into the archives.

This report describes the second phase of the project. The report is targeted at the project's steering group and others with an interest in determining the costs of long-term digital preservation and of mapping preservation activities.

Where relevant, the OAIS model's function names are stated in upper case. A glossary is presented in Chapter 11.

1.2 Objective

The Project's overall objective is to create a generic model for the compilation of actually defrayed costs, and to estimate the expected costs of preserving digital records at ABM institutions.
The objective of this phase was first and foremost to chart the costs of ingest of records in the archives, and to improve the CMDP model's ability to estimate the costs of logical preservation by the migration strategy. The objective was also to obtain cost data for what it actually costs to ingest digital records, and to test and adjust the model in relation to this input. The model was also to be expanded with the area of bit preservation (Archival Storage). Finally, the intention was to create a user-friendly interface to the CMDP spreadsheet and a guide to use of the model.

1.3 Key activities

The project comprises a number of key activities, as presented below. The activity numbers refer to the project initiation document.

5.3.1 Expansion of the model with "submission" and "reception" (ingest)
5.3.2 Collection of data on ingest costs
5.3.3 Further development of the model in the area of "migration" (Preservation Planning)
5.3.4 Application of the adjusted model to the data obtained
5.3.5 Expansion of the model with "bit preservation" (Archival Storage)
5.3.6 Overall report and spreadsheet for the model
5.3.7 User-friendly-interface to the spreadsheet
5.3.8 Translation of the final report into English

At the present time we have expanded the model with Ingest and obtained information on the actual costs defrayed for Ingest.

The results we have obtained from our questionnaires to public authorities (Producers) concerning their submission to public archives are varied. The results do, however, show, that approximately two thirds of the costs for the authority relate to the actual creation of the Information Package for the archive, i.e. for migration (normalisation) of documents.

We have also gained more knowledge of the impacts on the costs of logical preservation, and also improved the model's ability to estimate these costs. We must note, however, that the model is still based on a number of estimates, and that the cost estimates show wide variation, since by their nature these costs are dependent on the quality of preservation that is selected: there is, for example, a big difference between giving priority to preservation formats that are estimated to last for 8 or 20 years; whether priority is given to simple or complex preservation formats; and how many errors in migration are acceptable.

The estimates are based primarily on experience from the archive sector, since so far the libraries do not have much experience with the execution of migrations and documentation of the costs. For the same reason it must be expected that the estimates are most relevant for archives.

1.4 Resources

The project was allocated 12 man-months (mm), in total DKK 600,000, from the Ministry of Culture. The resources were distributed as 3 mm to one project participant from The Danish Royal Library, and 8 mm to three project participants from the Danish National Archives and one project participant from the Archives of the City of Copenhagen (Københavns Stadsarkiv). In addition, the Danish State and University Library was allocated 1 mm for review of the project report and the CMDP model (spreadsheet). The allocated resources have been used.

1.5 Products

The products from the project are the following:

- Cost model (spreadsheet) operationalised for the areas Ingest, Preservation Planning and selected functions under Administration.
- Project report in Danish

1.6 International orientation

The project has also stayed abreast of the international development of cost models for digital preservation.

We have been in regular contact with the British LIFE project (http://www.life.ac.uk/). In August, the CMDP project hosted a meeting at The Royal Library with the project manager and a project team member from the LIFE project. The purpose of the meeting was to evaluate the LIFE3 beta model (spreadsheet). Considerable changes have been made from the previous version of the LIFE project's spreadsheet, and the model's focus is now on the costs of digitisation and storage. All calculations for preservation planning and migration, called the Generic Preservation Model (GPM), have been removed from the model because it was too complex an area to operationalise. At the meeting, the LIFE project manager described how the objective was to create a web-based spreadsheet solution. The LIFE3 project closed at the end of August 2010 with the hope that the activities could be continued in a LIFE4 project, or under Open Planets Foundation[4]. The status today is that neither the spreadsheet nor the web solution is accessible.

Another cost project we have followed is Keeping Research Data Safe (KRDS). This project is designed by the private firm of consultants Charles Beagrie (http://www.beagrie.com/). The KRDS projects have published various reports and guidelines, but the model outlined has not yet been operationalised.

CMDP has also contributed a presentation at an international meeting of experts on the costs of digital preservation held by the Netherlands Coalition for Digital Preservation at the National Library of the Netherlands in The Hague[5].

[4] http://www.openplanetsfoundation.org/

[5] http://www.ncdd.nl/en/documents/20100916PriceTagsConferenceReportfinal.pdf

1.7 Future activities

The steering group intends to hold a workshop in March 2011 on the costs of digital preservation at The Royal Library, attended by a rather broader group from the institutions. The workshop will focus on the actual CMDP model, which is to be tested on a number of cases.

Expansion of the model to include the costs of bit preservation is expected to take place under the Archival Storage project. This task will be relatively simpler to undertake, as SB (The State and University Library), SA (The Danish National Archives) and KB (The Royal Library) have longstanding experience from bit preservation, and cost data for the area are available.

Furthermore, the opportunities to promote the CMDP model and the project's other results internationally will be investigated, including translation of the project report into English and its publication on the project's website. We have also discussed the possibility of a brochure (A4 page printed on both sides) with facts about the CMDP model.

We could draw even more attention from presenting the project results in an international journal or at an international conference. The presentation of the first phase of the project at the iPRES 2009 conference thus led to an invitation to publish an article in the International Journal of Digital Curation. In addition, as stated, we were invited to present our work at a meeting of experts at the National Library of the Netherlands in The Hague.

Website exposure and attending conferences and meetings of experts are not just a way of promoting the institutions' work, but also make it possible to gain new knowledge for the benefit of the institutions, besides creating opportunities to cooperate with other projects on the continued work on the cost model.

1.8 Conclusions

Overall, the method can be used to identify costs. However, the overall method is not yet sufficiently detailed to give accurate results for all types of records. Since all empirical data are from public archive materials, the model is currently most suitable for estimating the costs for this particular type of materials.

Generally, our work on the costs of digital preservation shows that to a great extent preservation institutions depend on being able to use standardised solutions, as it is very expensive for them to develop all solutions themselves.

A key conclusion from the work on Ingest, as confirmed by the questionnaire survey, is that normalisation of formats is the largest Ingest cost item, at around two thirds of the total costs.

Another key conclusion is that the choice of the digital object (the format), with its complexity and volume, as the calculation basis makes the model potentially generic and thereby capable of calculating the costs for various digital collections.

If CMDP is to be completed in terms of both method and precision, and expanded to include not only all OAIS functions, but also the opportunity for calculation of different preservation strategies and different digital collections, for example, additional resources are needed.

2 Introduction

This report describes the second phase of the "Costs of Digital Preservation" project.

The costs of digital preservation comprise all costs related to the establishment and operation of an OAIS-based long-term digital preservation system, i.e. from the ingest of digital records at the archive, to their preservation, and availability to the users of the archive.

According to the OAIS standard, an Archive is an organisation comprising systems and people who are responsible for receiving and storing the records and making them available to the archive's user group (Designated Community). The term archive relates to all types of preservation institutions, and not especially to the archive sector. The OAIS standard contains a functional model that describes an archive as seven main functions (Ingest, Archival Storage, Data Management, Administration, Preservation Planning, Access and Common Services), and three roles (Producer, Consumer, Management). The OAIS standard also defines three types of information packages: Submission Information Package (SIP), Archival Information Package (AIP) and Dissemination Information Package (DIP).

In the first phase of the project we created formulæ for activities under the OAIS main function of Preservation Planning, and for the functions under Administration that are part of, and underpin, the migration processes. In the second phase, we created formulæ for activities under Ingest and selected functions under Administration, which is also part of the ingest process. We also used the PAIMAS[6] standard to identify activities in connection with Ingest. PAIMAS, which is based on the OAIS standard, describes the activities related to the design of a Submission Agreement between Producer and Archive, as well as the actual transfer of data (Transfer) and validation of transfer (Validation).

2.1 Guide for the reader

In Chapter 3 we describe how we have used the OAIS and PAIMAS standards to expand the CMDP model with Ingest. Then, in Chapter 4, we describe what the Ingest costs consist of, and what they are dependent on. In Chapter 5 we assess the two preservation strategies: emulation and migration, and explain why the migration strategy is the most prevalent. We then in Chapter 6 describe the costs of the migration strategy distributed on the ongoing monitoring and migration of data, including a discussion of migration quality, format lifetime and migration frequency. This discussion was also the basis for our further development of the formulæ for logical preservation. In Chapter 7 we list the cost elements that are part of bit preservation. The analysis of these costs in relation to OAIS has not yet taken place. Chapter 8 presents the results of a questionnaire survey to obtain cost data from actual submissions. Chapter 9 presents a brief guide to use of the cost model, which does not yet have a user-friendly interface. Chapter 10 presents the project conclusions and the usability of the model. Finally, Chapter 11 provides a bibliography, and Chapter 12 a glossary of the terms used in the report and in the spreadsheet.

[6] http://public.ccsds.org/publications/archive/651x0m1.pdf

3 Ingest of digital data to the archive

Ingest is generally described in OAIS, but the creators of the OAIS standard have also found it necessary to make a more detailed description of the interaction that takes place between Producer and Archive when data is submitted. This is done in the Producer-Archive Interface Methodology Abstract Standard (PAIMAS). The focus of PAIMAS is on what is required in order to prepare a Submission Agreement, and describes a provisional and a final phase for the establishment of an agreement, and a transfer and a validation phase.

As the first step in identifying activities related to the ingest of digital records, on the bass of the OAIS standard we prepared a flow diagram of the process. Below we first describe this flow, in order to show which activities the process involves. It should be noted that the activity Generate SIP is not found in OAIS. We have added it in CMDP, in order to have a place to allocate any costs relating to the normalisation of the information packages before they are submitted to the archive.

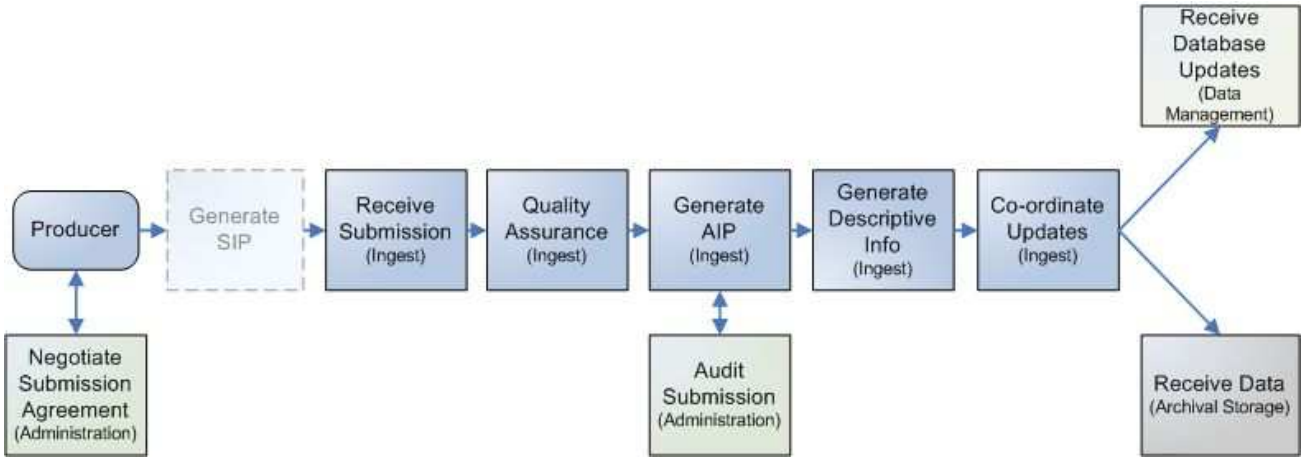

Diagram 1 shows the flow between OAIS functions from Producer to Archive on submission of digital records.

Negotiate Submission Agreement

Before data are transferred from Producer to Archive, a Submission Agreement can be concluded between the parties. In the OAIS model, the Submission Agreement is negotiated between the Producer and the function under Administration called Negotiate Submission Agreement.

Generate SIP

If the Archive can make requirements of the Producer concerning how the information packages are to be submitted it may be necessary for the Producer to normalise data to validated submission formats and/or enrich the packages with metadata before the submission information packages (SIP) are transferred to the Archive. This preparation of the packages is cost intensive, and as the activity is not included specifically in the OAIS model we have added this element to CMDP and called it Generate SIP.

Receive Submission

When the Submission Agreement has been concluded the Producer sends the SIP (submission information packages) to the Archive (Transfer), where they are received by Receive Submission and placed in temporary storage. Transfer may be by movable storage media such as DVD or hard disk, or via a network.

Quality Assurance

The submission is then checked for errors by the Quality Assurance function. This may, for example, be a check-sum control. If the packages are in order a confirmation of receipt is sent to the Producer. If there are errors or defects the Producer is informed, so that the packages can be transferred once again.

Generate AIP
The SIP are then migrated to AIP in the function Generate AIP under Ingest. This transformation from SIP to AIP may entail that formats are normalised to validated preservation formats, that the Representation Information is changed, and/or that the information in the packages is re-organised. The function may also request the Data Management function to send the information necessary for a full description of the package.

Audit Submission
The Audit Submission function then checks whether the packages fulfil the provisions of the Submission Agreement and sends an audit report back to Generate AIP. If there are errors or defects the Producer is furthermore informed, and the Producer then corrects the errors and transfers the packages once again.

Generate Descriptive Information
The Generate Descriptive Information function then pulls the Descriptive Information, i.e. metadata used to search and retrieve the packages, from the AIP package, and sends the information to Co-ordinate Updates. There is no function in OAIS to generate metadata and other documentation. Metadata are only pulled from existing sources. As generating metadata can be a significant cost, in CMDP we have included this activity in Generate SIP.

Co-ordinate Updates
Co-ordinate Updates sends on the AIP package(s) to Archival Storage, which confirms receipt with a confirmation and allocation of an ID for the AIP package when storage has been performed and verified. Co-ordinate Updates includes this ID in the descriptive information and sends it on to Data Management.

3.1 Preparation of submission agreement

The PAIMAS standard describes the preparation of a Submission Agreement between Producer and Archive, the performance of data transfer (Transfer), and validation that the transfer has proceeded as planned (Validation). The standard divides a submission project into four phases:

- The clarification phase must show whether a submission project is feasible and financially viable. The phase comprises the first contact between Producer and Archive, the provisional definition of the project's objective and context, a draft description of the digital data and its structure, and a draft of the actual Submission Agreement.
- The definition phase must result in a formal Submission Agreement between Producer and Archive. The phase must describe the SIP design, as well as exactly which digital data are to be submitted. It must also determine legal and contractual terms as well as security, and describe how transfer and validation of transfer are to take place. Finally, it must set up a timeframe for the project.
- The transfer phase must ensure that the digital records (SIP) are transferred from Producer to Archive, and that the Archive's initial processing of data takes place according to the guidelines laid down in the Submission Agreement.
- The validation phase must ensure that the transfer of the digital data is validated according to the guidelines set out in the Submission Agreement.

PAIMAS is a check list that describes the individual activities in each phase in approximate chronological order. The first two phases are undertaken in OAIS by the function Negotiate Submission Agreement under Administration. The transfer phase corresponds to Receive Submission under Ingest, while the validation phase corresponds to the functions Quality Assurance under Ingest and Audit Submission under Administration. The strength of the PAIMAS standard lies in the detailed description of what a Submission Agreement must include, how data must be transferred, and how the transfer must be validated.

We began by analysing PAIMAS to identify cost-critical activities and their interdependencies. This analysis showed that the degree of detail in PAIMAS exceeds the limit for cost-critical activities we operate with in CMDP (1 person week). We therefore amalgamated a number of the activities.

Diagram 2 presents the cost-critical activities on the preparation of a Submission Agreement. The largest boxes represent the main activities, and the smaller boxes, which are connected to the main activities, show which sub-activities they contain.

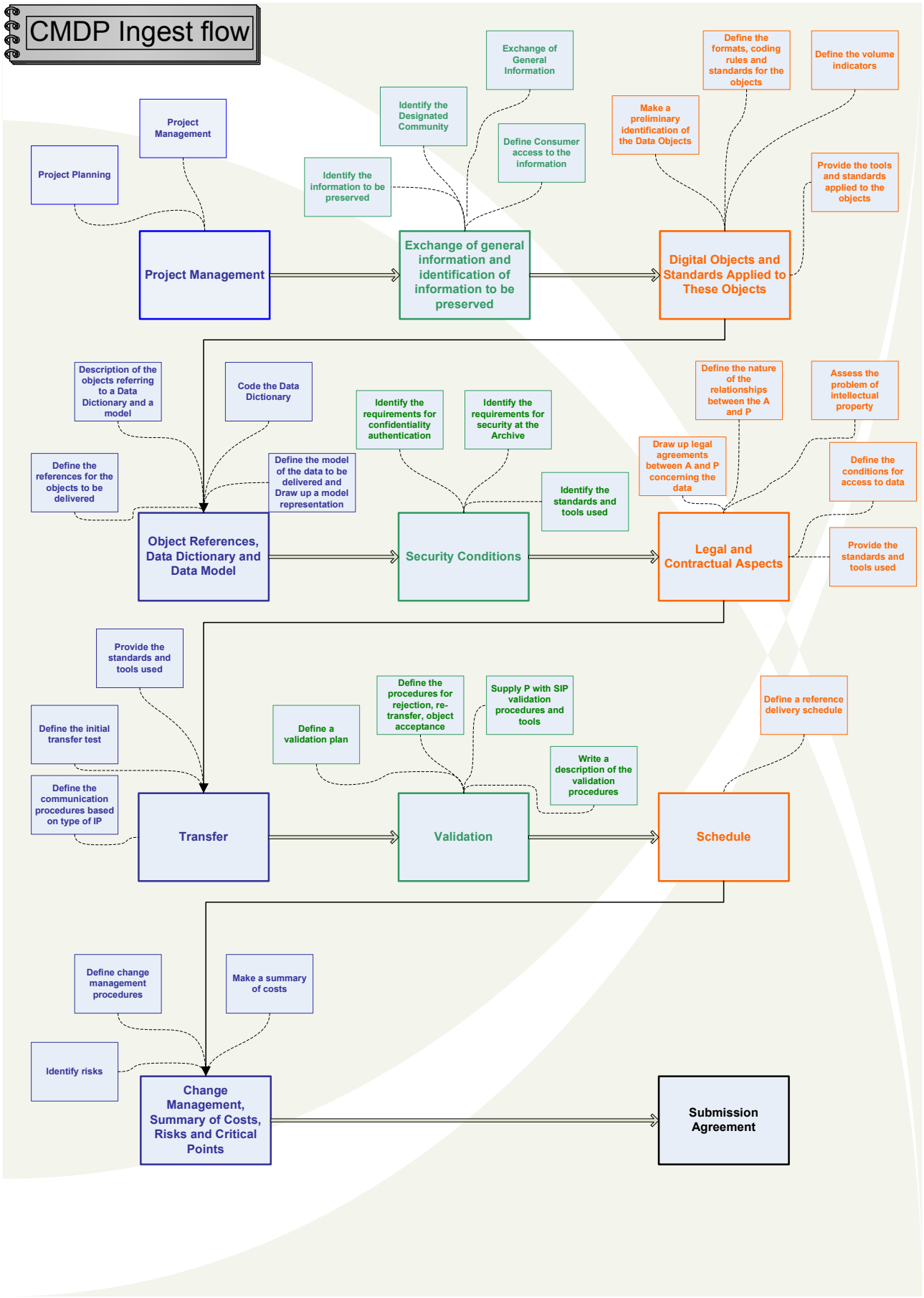

Diagram 2. Cost-critical activities in connection with the preparation of a Submission Agreement between Producer and Archive.

4 Costs of Ingest

Ingest of data worthy of preservation from the Producer can be a significant digital preservation cost. Below we review the cost-critical activities relating to Ingest and their interdependencies.

The costs of preparation of a Submission Agreement are first and foremost dependent on the distribution of strength between Producer and Archive, the diversity and complexity of the data, how well the data are documented, and the size of the submission project.

If Ingest includes the normalisation of data from the formats and structures used by the Producer to those used in the Archive this will entail considerable costs, irrespective of whether these costs are carried by the Producer or the Archive.

In OAIS, normalisation is named under the description of the Ingest function Generate AIP, where it is stated that migration from SIP to AIP may entail normalisation of formats (p. 4-6). PAIMAS also mentions briefly under the description of the Submission Agreement (p. 3-21) that if the existing format dos not match the format specified the Producer and Archive must find a solution, for example normalisation. It is thus the distribution of strength between Archive and Producer that determines where the costs are defrayed. The costs of normalisation are described in Chapter 6 ”Costs of migration”.

The costs related to the transfer of data from Producer to Archive comprise the establishment of an infrastructure and of temporary storage. We have not yet clarified whether these and similar system support expenses must be categorised as part of the Ingest costs, or placed under the OAIS main function Common Services.

After transfer, data must be checked for bit errors, but this process is automatic and only entails initial costs for the establishment of the quality assurance system.

The subsequent generation of the AIP package(s) may entail normalisation. This cost is described in Chapter 6. The addition of descriptive metadata for search and retrieval is normally extracted automatically from the package, and is therefore not cost-critical. On the other hand, the validation of the AIP by the Audit Submission function under Administration can be cost intensive. This first and foremost depends on any requirements made in the Submission Agreement, which again depend on the distribution of strength between the preservation institution (Archive) and the creator of the archive (Producer). Validation typically comprising ensuring that the information packages are complete, that integrity is maintained, and that they fit the data model, including that the agreed data formats have been used, and their syntax has been maintained.

The subsequent coordination with Data Management and transfer to Archival Storage normally also take place automatically and are therefore not cost-critical either.

5 Logical preservation (Preservation Planning)

In the first project report we described the identification of cost-critical activities and interdependencies for logical preservation by the migration strategy.

Costs of logical preservation comprise the costs to reproduce the data content in a logical form. The most realistic preservation strategies for logical preservation based on both a technical and a financial approach are estimated to be:

- Emulation of the original systems that can use the data in unchanged form
- Migration of data so that it can be used by contemporary systems

Both strategies entail the risk that the data content is not preserved completely, either because it is not possible to completely emulate the original systems, or because the data cannot be fully migrated so that it is reproduced correctly in a contemporary system.

5.1 Brief status of emulation

Over the last ten years emulation has steadily improved. Within the preservation institutions' area the National Library of the Netherlands has used a UVC (Universal Virtual Computer) system to make images available, and among other things via the EU project Planets developed the Dioscuri emulation system that can emulate x86 machine architecture.

The project Preserving Virtual Worlds, which focuses on the preservation of games and interactive fiction, and the EU project KEEP (Keeping Emulation Environments Portable), also work with emulation. The status of this research is that there are promising prototypes, but products that can function in a preservation system are still far away. The task is too demanding for the financial resources available.

The computer games industry has also contributed to the development of emulation. Games console manufacturers have thus developed emulation systems for their games consoles, which games manufacturers use on PCs. There are also user groups within computer games that develop emulators so that old games, for example for Commodore 64, can run on contemporary machines.

Emulation of the original systems requires that besides the original data the equivalent original systems and underlying operative systems are also saved, as well as documentation of their configuration and utilisation.

On reproducing data content may years after it is created it must be anticipated that this will naturally be according to the premises of the original system. The users must thus be able to operate older systems and operative systems in order to access data. The development in system operation in current decades points to relatively uniform user interfaces. It is remarkable, however, that systems dating only from the mid-1990s which do not have a graphic user interface, but are based on function keys and commands, appear to be a major barrier for young users that have never used such systems themselves.

The current status for emulation as a preservation strategy is thus that emulation must be considered to be troublesome to introduce, and relatively uncertain to maintain. It can furthermore be difficult to foresee whether the reproduction of content and functionality will be feasible in the long term. The preservation institutions that have an active preservation strategy often avoid the emulation strategy if their data can in any way be preserved via the migration strategy.

There are, however, types of records, for example interactive records, whose functionality cannot be preserved by the migration strategy. In this case some preservation institutions focus on bit preservation of

all the components required should it become possible at a later time to give access to the records via emulation.

5.2 Brief status of migration

The opportunity for migration of data as the migration of data from one data format to another has always existed, and in the last ten years, especially most recent years, this has become a less cost-intensive task. This is because today fewer formats than before are used; they are used for relatively longer; and they are relatively better documented (and for this reason among others can be used by several systems).

Even though migration has become relatively less demanding, it is in absolute terms still a very cost-intensive task. It consists among other things of the acquisition, operation and further development of systems for ongoing monitoring of which formats have become obsolete, so that their content must be migrated to other formats, and in particular the actual migration of data from one format to another, and controlling that the data is not changed unintentionally, i.e. in addition to that entailed by the chosen format.

Data in itself cannot normally be used, but requires a system that can interpret the data format in order to reproduce the data content. The more complex the data content and data format become, the greater the dependence on systems to interpret the data format. The migration strategy separates data from its system, eliminating the risk of not being able to fully reproduce the data content in another system than the system in which the data were created. This especially applies to data formats where the Producer has no interest in them being interpreted by other then the original systems.

The migration strategy is thus by no means perfect, but currently in technical and financial terms the most simple and secure preservation strategy, and probably for the same reason the strategy used in practice by most preservation institutions.

Below the costs of logical preservation will be considered solely in respect of the migration strategy.

6 Costs of migration

There are numerous costs of migration, of which the most important are the acquisition, operation and further development of systems for:

- Ongoing monitoring of which formats are obsolete, and of which the content must be migrated to other formats.
- Actual migration of data from one format to another, including control that the data are not changed unintentionally.

The cost estimates are based on own experience and a review of the source material on this subject. With regard to the further development of the migration cost formulæ, we have been inspired in particular by the guide: Software Development Cost Estimating Guidebook[7] (see also the Bibliography).

6.1 Selection of migration quality

As for many other costs, the selection of migration quality is decisive to the level of costs. Migration quality is determined primarily by the choice of the format to which data are migrated (hereafter the output format), and by the error tolerance on migration of date from the input format (the format from which data are migrated) to the output format.

6.1.1 Selection of output format

High quality in terms of an advanced output format (functional preservation) rather than a simple output format will entail significantly higher costs. This is because on migrating data from input format to output format programs must be determined and acquired to handle how all data in the input format is migrated to an equivalent site in the output format. In addition, on the actual migration it must be controlled that this has taken place, cf. below concerning error tolerance. For example, migration from one word processing format to another word processing format will entail significantly higher costs than migration to a simple format in the form of a graphic bitmap format, as the word processing format contains far more information than a graphic format. This is a general observation, since in practice the situation may be that the system that migrates data from the input format to an advanced output format is far superior to the system that migrates data to a simple output format. The choice of output format is also vital to determining how often migration should be performed (see below).

6.1.2 Selection of error tolerance

With regard to error tolerance on migration of data, high quality in the form of a low error tolerance will entail significantly higher costs than a high error tolerance. This is because a low error tolerance will typically require the use of extra funds for the acquisition, operation and further development of the system for the migration of the formats in question. In addition, on actual migration it will be necessary to use extra resources for error control, and especially error correction. As stated under selection of output format, irrespective of the choice of error tolerance there will normally be higher costs for the error handling of an advanced output format than of a simple output format. This is because there is more chance of something going wrong, and it is more expensive to correct the individual errors.

[7] Software Technology Support Center (STSC) Cost Analysis Group, U.S. Air Force, Software Development Cost Estimating Guidebook, 2010, http://www.stsc.hill.af.mil/consulting/sw_estimation/SoftwareGuidebook2010.pdf

Selection of output format and error tolerance can furthermore be combined. Depending on the purpose of preservation and the data content, there may be a wish to select an advanced output format and a high error tolerance. On the other hand, a simple output format and a low error tolerance may be preferred. The selection of an advanced output format thus does not necessarily entail the selection of a low error tolerance, just as the selection of a simple output format does not necessarily entail the selection of a high error tolerance.

6.2 When should migration take place?

An important factor with regard to the costs of migration is when, and especially how often, migration should be performed. The immediate answer is as seldom as possible, while bearing in mind the risk of obsolete data. This is because each migration entails a risk of losing information when data are migrated from one format to another, and because each migration entails costs.

There are different tactics for when it is best and least expensive to migrate, including to which output format.

One tactic is called migration to contemporary standardised format, and consists of migrating data to a contemporary standardised format as seldom as possible. The argument behind this tactic is that by migrating to a contemporary standardised format the number of migrations is reduced, and thereby the risk of unintended changes. The reason is that the lifetime of a standardised format is expected to be significantly longer than for other formats, as several systems will be able to read data in the format and write it in another. In addition, the standardised format should make it less expensive to acquire, operate and maintain systems for actual migration, due to the larger range available.

On the other hand, the number, dissemination and system support of contemporary standardised formats is estimated to be modest. It is therefore necessary to either select output formats of low quality (simple formats), or to perform migration almost as frequently as if the next generation of the input format had been chosen as output format.

Another tactic is called migration to the latest format, and consists of continuously migrating data to the currently latest output format. The argument behind this tactic is that it adopts the situation of other IT users with a need to migrate data from the previous generation of the format to the latest as correctly and inexpensively as possible. This makes it possible to benefit from the systems (programs) for the latest generation of the format, which must be assumed to be the best for reading the immediately preceding generation of the format.

On the other hand, the frequent migrations are cost intensive and increase the risk of unintended changes. Moreover, the programs for the newest generation of the format are not always the best to interpret the previous generation. Sometimes it is necessary to wait for the following generation to achieve better reproduction. In addition, suppliers and users seem to be so much more interested in being able to create new data in new formats, rather than reading older data in older formats in the new generations of the systems, that nothing particular is done to facilitate migration. It is thus difficult to find systems that handle mass migration of the previous to the current generation of the format.

A third tactic is called migration on demand and entails that if the data are in a relatively common and documented format the data are retained in this original format and not migrated to another format until the data are requested. The argument behind this tactic is that it is estimated that the number and variation in the use of data formats is continuously narrowing, and that dissemination, openness and documentation are widening. The probability that in a few years it will be possible to read a previously relatively common and

documented format is therefore so high that there is no reason to perform migration before then. This saves a large number of intervening cost intensive and hazardous migrations.

On the other hand, the risk is considered by some to be too high, i.e. the probability that after a number of year there will, after all, not be any system that could interpret the format. In addition, depending on the output format, it is often an advantage to migrate shortly after the data are created, as many formats are not isolated, but depend on external data, for example fonts in the system, or references to images or other data outside the format that may have been altered after a number of years. These are external dependencies of which the encapsulation requires systems that have to be acquired, operated and further developed. Some standard programs, such as MS Word 2010, now support partly embedded fonts.

6.2.1 When is a format obsolete?

When a format is obsolete is related to the previous discussion of when migration should take place, and is primarily a question of costs.

Virtually irrespective of its format data in themselves are of little value, as it requires a system, a program, that can interpret the data format in order to reproduce the data content in an understandable form. The obsolescence of formats is therefore dependent on the obsolescence of the software that is to interpret the format.

With regard to data preservation, it is not sufficient that a program can interpret data in their current form (input format), as it must also be able to write data in a suitable contemporary format (output format); otherwise the result is a dead end. Especially in earlier times it has been emphasised that there was often a projector – a viewer – for a format, which was therefore not obsolete. This is not a tenable argument; however, as the result is dependence on new generations of the viewer, and in addition the data cannot be further processed in other systems/programs.

The answer thus merely leads to the new question of when the systems in question that can read data in its present format (input format), and write the data in a contemporary output format, are obsolete?

For as long as new generations of systems are developed that can read data in its present format (input format), and write the data in a contemporary output format, there is no real problem of obsolescence. This does require, however, that the system is tested, and that the reproduction is acceptable. When a generation of the system is developed in which the data format in question can no longer be interpreted, or just cannot be written in a suitable contemporary format, it is necessary to use the *previous* generation of the system to do this, thereby becoming dependent on its lifetime.

More and more formats can be interpreted by systems that are several generations younger, although naturally there are limits. Appendix 11.1 presents an example of MS Excel 2007 from which the possibility of reading a number of older formats has been removed. It is therefore necessary to use the previous generation of the program to read the formats and save them in a contemporary output format. As for most new generations of programs and data formats there are a number of functionalities and derived data that are not supported in the newest generation.

The lifetime of systems/programs does not end on the same day that a new generation of the system is born, or a competing system takes over the market. The lifetime of systems/programs is dependent on the costs of their use and maintenance. For as long as a system/program can run on contemporary hardware and be integrated with contemporary systems the costs of its use and maintenance are manageable. Thus, neither the system nor the data formats it can interpret and write to, are obsolete.

A known example of obsolescence is the BBC Domesday Project: In 1986 the BBC published an extensive modern multimedia edition of the famous Domesday Book that describes England in the 11th century. The BBC Domesday edition consisted of letters, maps, images, statistical data, videos, etc. stored on two interactive laser discs, LaserVision Read Only Memory (LV-ROM). In 2002, it was feared that the discs would become unreadable due to the technological obsolescence of the data storage medium and it was necessary to use migration, emulation and re-digitisation in order to preserve the data. This was technically possible with great difficulty, and the high costs were a clear indication that the formats had become obsolete.

"The lesson of this digital preservation project is that if you have enough time, individual skill, dedication and imagination then almost anything is possible, provided that you don't leave it too late. If you start counting the cost this may seem an expensive project, but then the value of the record is high too - and that applies equally to the original Domesday Project. There is of course a great need to preserve other electronic records in a routine and predictable manner, and this rescue project is not a suitable model to be followed in such cases. The National Archives is working on ways to make this possible in future"[8].

This is despite the fact that from the outset the project's creators were aware of the preservation risk and had in due time submitted data and documentation to an archive that did not handle the matter satisfactorily.

"The deputy editor of the Domesday Project, Mike Tibbets, has criticized the UK's National Data Archive to which the archive material was originally entrusted, arguing that the creators knew that the technology would be short lived but that the archivists had failed to preserve the records effectively."[9]

Do we always have to rely on existing systems to be able to read data in a given format? In practice yes, since even with exhaustive documentation of the format it is normally a very demanding task to develop a system, or have one developed, to read data in one format and write it in another.

The exception is the very simple formats in which, at a modest cost, it is possible to develop systems, or have them developed, that can read data in one format and write it in another. Examples include TIFF, UTF-8 or XHTML.

6.3 How often should migration take place?

On the basis of the current situation our tentative estimate of when a format is obsolete is eight to 20 years after its introduction on the market.

Twenty years is based on the furthest horizon we dare estimate within digital preservation. Eight years is based on the time within which we estimate that it will still be possible to run a program that can read data in its input format and write it in a suitable, contemporary output format.

Both extremes are tentative estimates, and in a few years' time it may be possible to estimate an even longer lifetime, if only because we will then know the time that is now the future, and will once again make an estimate 20 years into the future.

---

[8] Darlington, J., Finney, A. & Pearce, A., Domesday Redux: The rescue of the BBC Domesday Project videodiscs, Ariadne Issue 36, 2003, http://www.ariadne.ac.uk/issue36/tna/intro.html

[9]Mike Tibbetts, ACM Committee on Computers and Public Policy, Forum on Risks to the Public in Computers and Related Systems, Volume 25: Issue 44, Saturday 8 November 2008, http://catless.ncl.ac.uk/Risks/25.44.html#subj7

6.3.1 Important lifetime parameters

It is extremely difficult to estimate the lifetime for a given format between the extremes of eight and 20 years, but we assess the vital parameters to be the dissemination, complexity and documentation of the format. Lifetime is increasing due to wide dissemination, low complexity and good documentation. The three parameters are mutually dependent, which does not make the estimate easier. Simple, well-documented formats are often very widely disseminated, and simple formats are often well-documented.

In this context dissemination concerns the number of users, but especially the number of different systems that use the format. IT is a market with considerable network effects, and the aim is to develop programs that can fully read a competitor's format, but only write in their own formats; otherwise it is necessary to compete on the competitor's home turf, or on an equal footing.

Complexity is dependent on the number of types of information in the format, including the functionality in the system that is reflected in the format. Highly complex formats are often replaced more quickly (than formats of low complexity) by new generations of the format, as producers or users require even more functionalities. As stated, formats of very low complexity can be independent of existing systems because on the basis of the documentation, if it is good, it will be possible, without prohibitive costs, to develop a system, or have a system developed, to interpret the format.

Documentation concerns the description of the structure and use of the format. A characteristic of good documentation is that it gives others besides the original creator of the format a feasible opportunity to develop systems that can interpret the format. It will at times also be necessary to have partial documentation of the system in order to understand how to interpret the format. For documentation to be good it must first of all be accessible, and secondly include the entire structure and use of the format, and finally be explanatory, i.e. intended to ensure that others besides the original developers can understand the format.

6.4 Costs of systems for monitoring and migration

- Ongoing monitoring of which formats are obsolete, and of which the content must be migrated to other formats.

- Actual migration of data from one format to another, including control that the data are not changed unintentionally beyond that entailed by the chosen format.

6.4.1 Costs of ongoing monitoring

Costs must be defrayed for the acquisition, operation and further development of a system for identification and registration of all formats for all data, stating the precise version of each data unit.

In practice this entails that on ingest of data in the preservation system all data are analysed, so that its formats can be identified and registered, and so that all data in a given format can be retrieved when it is transferred to another format.

This task can be handled by the submitting party if the receiving party can get the submitter to undertake the task, and trust the result, but in practice most preservation institutions will handle this themselves.

Identification should in practice be followed by validation and partial characterisation. This is because far too much data do not comply with their format, and that many formats are so rich in content that it can be necessary to have information on their characteristics, i.e. which parts of the format contain data.

6.4.1.1 Acquisition

Acquisition of such a system currently entails that it has to be developed, or required to be developed, although there are partial solutions in the form of PRONOM, DROID and JHOVE.

We estimate that the costs of acquisition of the core of a modular system that via specific modules for the individual formats can perform reasonable identification, partial validation and a small degree of characterisation will be 12-24 person months.

The costs of the development of the individual modules depend on the formats' complexity and documentation, and are estimated to be respectively exponentially increasing and diminishing. We estimate that the cost per format will be from a few person weeks for simple formats to several person weeks for advanced formats.

Going beyond what we unclearly call reasonable identification, partial validation and a small degree of characterisation, we estimate that there will be a highly exponential increase in the costs. It has, for example, still not been possible to achieve a complete validation of PDF/A. It is currently necessary to use validators from several suppliers to cover as many areas as possible.

It will not be possible to avoid incorrect identification or incorrectly formatted data. In practice, it must be hoped that the programs to migrate data to other formats are relatively error tolerant. It will not be possible to avoid a few errors without very high costs.

In a situation such as ingest to the preservation system, where it must be possible to ensure migration of an almost unlimited number of formats, the cost of the identification of all these many formats will entail that it will be necessary to accept a general rough identification, with best identification of the most used formats.

6.4.1.2 Operation

Actual monitoring takes place by manual review of the list of formats used and comparison of them with the development in the market for the formats in question, in order to assess whether some formats are becoming obsolete. Work is also taking place on the establishment of a joint international format register, the Unified Digital Format Registry (UDFR)[10], which will be able to streamline monitoring. Monitoring of the market entails in practice that for each format there is one or several system(s)/program(s) that must be registered and stated as necessary to interpret the format. These systems/programs' lifetimes must also be assessed, including whether the format is supported in the newest generation of the system/program.

The task is highly manual, and we estimate that the cost is proportional to the complexity of the format. On this basis it is estimated that monitoring will take from individual person days to a few person weeks, and that it will most frequently have to take place every second year for a given format.

6.4.1.3 Maintenance

Besides general maintenance, the maintenance of the system, for example in connection with a new operating system, also includes the development of new profiles for identification, validation and characterisation of any new formats that might be used in the preservation system.

[10] Unified Digital Format Registry (UDFR), http://www.udfr.org/

6.4.2 Costs of data migration

The migration of data from one format to another lies at the core of migration, and in terms of costs can be divided into a number of sub-tasks.

6.4.2.1 Acquisition

Acquisition of a migration system entails the development of a system/that a system is developed with modules to handle the following tasks:

- General module that on the basis of central registration of data and their format can retrieve the data in an information package (an SIP or AIP) of which the format is estimated to be obsolete, and unpackage this data.

- General module to manage all information packages and data retrieved in the obsolete formats, as well as their status, via the migration process. For each body of data in a format the module must request the specific module created for each format, register the result, and if successful send the migrated data in its new format for "repackaging" with the unaltered data from the package, so as to create new packages. To ensure efficiency the module must be able to parallelise its requests.

- Specific modules for each format that ensure that the data in the format is migrated with the system/program considered to be the most suitable to migrate date from the input format to the output format in the required quality. The programs in question will normally be the same as were registered in conjunction with the monitoring of the format's obsolescence To be able to automate migration there will be a need for the module to be able to control parts of the program's behaviour, for example so that it is not stopped by enquiries from the program. If an advanced output format is selected there may furthermore be a need for further management of the program in order to migrate all the required information to the output format.

The costs of developing the above system, or having it developed, are considerable, and reuse of others' solutions is an obvious alternative. We do not know any turnkey solutions, but a number of sub-solutions, such as Apache Hadoop or Berkeley Boinc, might be used.

We estimate that the general modules do not require development over more than 12-24 man-months.

The costs of the specific modules are not necessarily proportional to the number of formats, if a series of formats use the same program for migration. The test of the correct functioning of the module with a given format is, however, proportional, and the cost can therefore be almost proportional. Reuse of others' solutions is an obvious path to take, but we do not know of any such solutions. For each format, primarily the advanced formats, where there is a need, there are often full or partial solutions, such as Apache POI or Microsoft Open XML Format SDK, that can manipulate the running of a program or directly access the format. We assess, however, that the cost of directly accessing the format in the case of advanced formats, such as ODF or OOXML, exceed what is feasible for an individual preservation institution. The institutions must therefore await development in a wider community if the quality is to exceed that offered by turnkey programs.

We estimate the cost per format to be exponential to the format's complexity, and vice versa in terms of error tolerance.

We estimate that a module for a simple format with a low error tolerance will take a few person weeks, while an advanced format with a low error tolerance will take several person weeks.

6.4.2.2 Operation

The costs of operating the system are primarily related to error handling. As for so many other systems, the extent of error handling depends on how reliably the system has been developed to operate. In this respect the costs of development and subsequent error handling are often inversely proportional, and it is not easy to calculate the optimum distribution.

Error handling comprises actual operational interruptions in the areas for which the system has not been developed to operate reliably enough. It also includes the identification of errors in the individual modules, when a format cannot be migrated as expected. Finally, error handling concerns errors that the system does not know that it makes, and which can only be detected via subsequent random sampling. In other words, handling errors that it is known will arise; errors that are assumed to arise; and errors that are not expected to arise. When the errors have been identified it is necessary to decide whether, and if so, how, they are to be corrected.

Depending on the migration quality selected, primarily the complexity of the output format and the migration's error tolerance, we estimate that per format per TB (Terra Byte) from one person day to a few person weeks must be devoted to monitoring. We estimate that error correction takes up to ten times longer than monitoring.

Even though the costs can be compiled per format, there are still economies from migrating several formats simultaneously, for example on packaging and unpackaging, storage, and error handling. As formats do not die on the same day that they are declared to be obsolete several obsolete or virtually obsolete formats should be gathered for simultaneous migration.

6.4.2.3 Maintenance

Maintenance of the system comprises general maintenance, for example in connection with a new operating system and the development of new modules for new formats.

7 Costs of bit preservation (Archival Storage)

The costs of physical preservation comprise the costs of data storage, in particular the costs of acquisition, operation and further development of systems that can handle:

- Acquisition of a sufficient number of media (storage capacity).
- In a sufficient number of copies.
- Distributed on several geographical locations in order to reduce the risk of fire, theft, vandalism, terror and natural disasters.
- Distributed on several types and makes of media in order to reduce the risk of general production errors.
- Adequate frequent monitoring of the content of media to ensure, or at least show to be probable, that they are unaltered at the time of control.
- Documented transfer of data from old to new media, including new types of media, when existing media are becoming technically or technologically obsolete.

8 Collection of cost data for Ingest

A questionnaire was sent to a number of public authorities in order to collect information on their actual consumption of time and resources to produce information packages of data from IT systems in connection with submission to The Danish National Archives. The data collected will be used to test and adjust the cost model in relation to the newly-developed Ingest part of the model. If the authorities used an external supplier to produce the information package they were also requested to submit a copy of the contract/invoice for the assignment in order to obtain a full overview of the costs.

The questionnaire (see below) was prepared on the basis of the PAIMAS standard.

	Questions regarding submission of IT system to archive	Time	Amount	Remarks
A	How many hours were spent on project planning? For example: attending meetings, ongoing contact with receiving archive and supplier, preparation of contract, etc.			
B	How many hours were spent on identification and description of submission data and documentation? Examples: data retrieval, identification of which data to submit and how it would be submitted (possible submission model), retrieval of technical and administrative documentation, etc.			
C	How many hours/how much money were/was spent on production of the information package? Establishment and set-up of hardware/software Extraction of data from the authority's system Conversion of data for archiving (tables, types, structure) Document conversion Machine processing time Error correction (manual, technical and administrative), such as handling of documents with errors Preparation of metadata Packaging of the information package			Please attach any supplier contract
D	How many hours were spent on the actual physical submission of the information package to the receiving archive?			
E	How many hours were spent on testing the information package up to approval?			

The questionnaire was sent to 34 authorities, of which approximately half replied. Table 1 below presents cost data for selected authorities on submission to a public archive. The responses received point in many directions and show that the authorities found it difficult to understand the questionnaire and compile the consumption of resources. Based on the responses received a tentative conclusion for large submission projects (>160 hours) is that project management (A) costs approximately 13% of the total submission project. Identification and description of the digital objects and their references (B) accounts for approximately 16%. Production of the information package (normalisation) (C) accounts for approximately 66%. Test of the information package (E) accounts for approximately 5%. The responses concerning the time spent on the actual physical submission (D) are not included as the responses showed that the question was not understood correctly. It should furthermore be noted that the actual expenses defrayed are stated, since the total costs are not known. Furthermore, in the case of the authorities that have used consultants, a high

price is not necessarily equivalent to high revenue for the consultant, as the price/revenue ratio is not equal for all consultants.

Table 1. Costs for selected authorities on submission to a public archive.

Institution	Total time consumption	Total amount (DKK)	System	MB	Document formats	Contract year	Project Planning (hours / DKK)		Data-ID and description (hours/DKK)		Production of IP (hours / DKK)		Physical transfer (hours / DKK)		Test	Producer	Submitted to?	Price per MB	Hours per MB	Hourly price
A	49	17.100	E-J	10,0	TIF	2010	9		6		30	17.100	2		2	1	KSA	1.710,00	4,900	349
B	40	12.540	E-J	100,0	TIF	2010	9		6		22	12.540	1		2	1	KSA	125,40	0,400	314
C	681	487.918	ESDH	180.000,0	.xlsx, .xls, .	2008	181	61.400	500			426.518				2	SA	2,71	0,004	716
D	20	36.000	Register	700,0	SQL							36.000				6	SA	51,43	0,029	1.800
E			ESDH	2.100,0	mails og dc	2010										2	SA			
F	160	207.628	ESDH	27.284,0	AcroExch.D	2008	40		40			207.628	70		10	2	SA	7,61	0,006	1.298
G	772	571.638	ESDH		ACE, AVL,	2007-08	60					571.638				3	SA			740
H	235	1.621.885	ESDH	86.603,0		2007	175		60		547	656.884				4	SA	18,73	0,003	6.902
I																5				
J	112	100.000	ESDH	44.000,0	alle	2006	31		74		50	100.000			7	1	KSA	2,27	0,003	893
K	182		ESDH	328,5	wordperfec	2006-10										1	KSA		0,554	
L	17,5	754.812	ESDH	189	se aftale m	2007	5		12,5			194.812		560.000		2			0,093	43.132
M		662.660	E-J og ESDH	134,0		2006						662.660				2	SA	4.945,22	0,000	
N	577	451.350	Register/ESDH	22.000,0	TIFF	2010	20		5		531	451.350	1		20	1	KSA	20,52	0,026	782
O	218	158.950	Register	6,5	TIFF	2009	10		5		187	158.950	1		15	1	KSA	24.453,85	33,538	729
P		2.159.999	ESDH	1.499.000,0				280.000		60.200		1.819.799						1,44		

9 Guide to the CMDP model (spreadsheet)

As an element of this report a spreadsheet called CMDP_2.xls has been created. The purpose of the spreadsheet is to facilitate use of the model, by which is meant select formats and quantities, and adjust constant values and thereby formulæ.

As the spreadsheet is intended to be as open and accessible as possible, the creation of own macros and functions has been avoided. However, there has proved to be such a great need for functionalities such as look-up and calculations that many of the formulæ become quite complex, even with the help of sub-totals directly in the spreadsheet to ensure transparency.

The spreadsheet is thus open, but not as accessible or transparent as required. The problem is that the spreadsheet format as such cannot fulfil both the high functionality requirements and the need for openness and transparency. In any next version of the spreadsheet there will thus be a need to implement most of the functionalities in own macros and formulæ in order to reduce complexity, which on the other hand will make the spreadsheet is less open and the forumulæ less transparent.

9.1 Structure

The spreadsheet contains the most important cost elements from the functional entities (Functional Entity) in OAIS considered in this project, i.e. Ingest (IN), Administration (ADM) and Preservation Planning (PP). The spreadsheet is structured as several sheets (tabs). A sheet for each functional entity is thus used.

The first sheet, ”Overview”, is an overview of OAIS functional entities, including Ingest, Administration and Preservation Planning.

The next sheet, ”Cost”, compiles costs from the individual sheets for each functional entity. ”Cost” states the costs over a 20-year period based on stated annual and periodical costs as found in the individual sheets for Ingest (IN), Administration (ADM) and Preservation Planning (PP).

The sheets IN, ADM and PP are based on the same template, stating element and sub-element. Each cost has its value stated in person time (typically person weeks) and distributed on types of manpower and frequency.

The sheets Ingest (IN), Administration (ADM) and Preservation Planning (PP) have other sub-totals, which are sheets called PAIMAS and Migrate (the spreadsheet format does not allow sheets to be ranked, and this order is therefore not immediately visible.) The IN sheet gets some of its data from the PAIMAS sheet and from the Migrate sheet. The ADM and PP sheets get some of their data from the Migrate sheet. The spreadsheet contains a sheet with data from cases, and the sheet’s name begins with ”Case”.

9.2 The Migrate sheet

This sheet calculates the costs of migration from one format to another. The idea is to select the formats to be migrated from and to, the quantity for each format, and how the migration software is obtained (based on existing programs or development).

Based on the many pre-determined parameters, which can be changed, a result is shown (compiled in person weeks) for both the first normalisation (Ingest) and the following migrations (from preservation format to preservation format).

The sheet comprises a series of blocks:

Format identification
Here the format's type, name and group are stated. A number of known types and formats are stated in advance. When a more comprehensive joint format register has been developed it will be possible to apply its definitions. Until then, the user must add new formats (by inserting rows).

User choices for size and format
The data volume in Gigabytes (GB) and the output format are stated here.
For each input format the choice of output format is limited to a small number of formats (they are shown in the same column block as the input formats). The delineation is made according to which output formats have been found to be most optimal from a preservation perspective. The criteria have been expected lifetime and low complexity. All output formats are also shown as input formats.

Here it is also shown how the software for migration is achieved, i.e. whether existing software is used, or software is developed.

The starting point is existing software, as for virtually all input formats there are programs than can read these formats and write them in the selected output format. These should be selected, as the costs involved are far lower than for development. As an example, it will be far less expensive to acquire, for example, Word Perfect X5 to read the format Wordperfect 7 and write in the format ODT 1.0 than to develop software for this.

Format characteristics
This block shows a series of characteristics for each format, which are specifically its complexity, its documentation and how it is implemented. All values are stated in relation to the format type of unformatted text. The characteristics in question have been selected as we consider them to be of significance to the costs of preservation.

Complexity is closely related to functionality. The contention is that complex formats more quickly become obsolete, as they are replaced by new versions, simply because additional or other functionalities are required. In addition, complex formats are inherently more expensive to handle than simple formats in many areas, quite simply due to the wider outcome range.

The relative level of complexity is based on the assessment of formats that took place in the first part of the project. The assessment was to a high degree based on the volume of documentation that had to be studied in order to understand the format. Among the relatively few types of format stated, spreadsheets and word processing have the greatest complexity.

Documentation states the degree to which a format has documentation that is accessible, adequate and usable. In practice this is only fulfilled 100% for the simple text formats, which with UTF-8 may be an approximation.

Implementation shows the extent to which the format has been implemented in open source code, the extent to which the format is widely used, and the extent to which the format is currently used. The statement of open source code is rarely 100%, as it is rare for all elements of the format to be supported in open source code. Widely used means that the format is used in many different sectors and countries. This improves opportunities to find software that can read the format.

Present use states the extent to which the format is used within its area, and its rough share of use of its format type. For example, Word Perfect 7 is very widely used as it is widely available, but its current use is

minimal, whereas various contemporary versions of MS Word have more than 90% of the current market for word processing.

Format Life Time

The calculated lifetimes are not used directly in the model, but merely serve as a guide. Lifetime is calculated on the basis of the format's characteristics, with high weight (40%) given to the complexity of the format. The reason for the low weighting of documentation and implementation in open source code is that in cost terms it is most advantageous to base migration on existing programs, so that dissemination and current used are given higher weighting. Review of the results will cause many to remark that one format or another is expected to have a far shorter or longer lifetime. This only serves to show that other variables, which we have not yet identified, also play a role.

Software provision using existing software

This compiles the cost of obtaining software for the individual format. It is assumed that the optimum migration procedure is to develop a general migration system for which implementations for each format are created.

The cost for each format is sub-divided into procuring and testing the software and implementing it in the migration system. Exactly which concrete software to acquire for a given format is not stated. Other tools, such as Plato from the EU project Planets, can be a help in selecting tools, but it is generally assumed that the programs created with the format are used, if they can write in the required output format.

The cost is to a wide extent based on the format's complexity, and in particular on the complexity of the output format. The model in this version does not take into account that a large number of formats are so similar that the costs are diminishing rather than directly proportional, for example MS Word 2000 and MS Word 2002. The argument is that on migration from something complex to something simple (for example from ODT to TIFF) it is the output format that determines how much complexity to transfer, and thereby the costs such migration entails.

The output format is weighted at 75% in the model, and complexity entails a cost increase to the power of 1.2 (exponent of 1.2). The cost is thus not directly proportional to complexity, but rising. An example to support this assumption is that experience from the processing of the open document formats ODT and OOXML showed that the costs of determining whether data had been migrated correctly from closed to the open formats, or between the open formats, rose strongly on application of the complex functions of which the results could be stored by the output formats.

The cost of implementing the migration software in the migration system is set at 50% of the costs of finding and testing the software. The assumption is that it must be possible to automate the software so that the migration system can, without risk of interruptions, request the software in question when one of its formats is to be migrated. An example is that many of the programs that are best at reading the input formats in question and write in the required output formats are not immediately suitable for automatic processing (for example as batch run without GUI). Own experience, as well as from ESDH deliveries, shows that a lot of effort is needed to "encapsulate" programs in order to achieve automatic processing.

The lifetime of the migration software is stated in this block, and it is an important assumption that, like all other software, the software has a limited lifetime. The format itself often lives far longer than the concrete software selected for migration. The stated lifetime of eight years may appear short, but our limited experience does not show many examples of software that is still in full operation after eight years. Dependencies on operating systems, etc. have so far entailed a relatively short lifetime, although it appears to be lengthening.

Software provision developing software
Here the cost of procuring software for the individual format is stated. The assumption here is that it has not been possible to find software that could read the input format and write in the output format. The basic cost is several times higher, and the exponent is also higher, as complexity is considered to be even more important when own development of software for the format has to be undertaken.

Documentation carries far greater weight that on use of existing software, just as an existing implementation in open source code reduces the cost.

The cost of implementation in the migration system is lower than for existing software.

Software provision for migration system
This block states the cost of the migration system. The central migration system is assumed to request other software for each format. The migration system must thus manage the process from input to output format so that migration can take place as automatically and quickly as possible, including with the fewest possible interruptions. The system must thus be able to parallelise migration, handle interruptions for individual tasks, and restart them. The cost is estimated at 80 person weeks, and the system is expected to exist for eight years.

Processing
This calculates the time it takes to migrate each format, both for machines and for the individuals who are to monitor the process, make sample checks and correct errors. The calculations are based on the complexity of the format, with unformatted text as the basis. The calculation is adjusted for each sub-process, but with this starting point it cannot be avoided that formats that deviate considerably from text are given a misleading calculation (video, for example, can be a complex format to migrate, but sample checks can be made relatively quickly). A more accurate calculation will require more parameters for each format, which was not possible with this project.

Migration is based on the format's complexity, and input format and output format carry equal weighting. The exponent is below one, as it is assumed that the processing time is not directly proportional to the complexity, but falling. It is assumed, for example, that even if a spreadsheet format is 14 times more complex than an unformatted text format it only takes eight times longer to process the spreadsheet format.

Monitoring is set at 5% per machine hour. Sample checks are of 1 per thousand of the data volume. It is assumed that this control costs 16 hours per GB of unformatted text. A data volume of 1,000 GB will thus entail a sample check of 1 GB, which will take 16 person hours, equivalent to 1 minute/MB subject to sampling.

Error correction is based on 1 per thousand errors per GB for unformatted text, and that this increases more than proportionally to the format's complexity, and that it takes five minutes to correct an error for unformatted text. It should be noted that the co-variation between sample size and error rate is not included in this version.

Format Interpretation Factor
Here each format's "format interpretation factor" is stated, based directly on the format's characteristics. The vital parameter is the basic value of one person week per 100% of the format's complexity. There are also small additions for documentation and implementation.

Format interpretation factor is among other things used to determine the manpower required by the archive to monitor whether the format has become technologically obsolete.

Migration Factor
The equivalent output format data are stated in the rows after all input formats. As all output formats are also input formats they have the same characteristics, and they are mainly stated for the sake of clarity.

Output formats must also be migrated to new output formats. In other words, at some point a format such as JPEG2000 must also be migrated. In this respect the model is very simple and very tentative, as all formats are migrated simultaneously after eight years. This must be understood to mean that every eighth year there is migration to formats of equivalent complexity (for example, JPEG2000 is migrated to an output format of equivalent complexity after eight years, and this new output format is in turn migrated after another eight years to a new output format of equivalent complexity, and so on).

For each format, the cost of developing the software to migrate the format in question is stated, and the sum for all formats is called the ”Total Migration Factor”.

9.3 The PAIMAS sheet

The PAIMAS sheet is divided into series of blocks for each main group of tasks in PAIMAS, while the individual cost parameters are shown in columns.

The main groups are Project Management, Information to be preserved and users (Content), Digital Objects and Standards Applied to These Objects, Security Conditions, Legal and Contractual Aspects, Transfer and Validation.

The aforementioned tasks are the basis for stating the cost of preparing a submission agreement - a cost that in OAIS is placed in Administration under Negotiate Submission Agreement.

For each main group a share of the total costs is determined. These shares are determined on the basis of the typical pattern in the questionnaire issued under the project, as well as the project members’ experience. Each main group is sub-divided according to the tasks that are included in the main group.

For example, the main group 3.12.5. Security Conditions is allocated a share of 10%. This 10% is again divided into the three groups P23 & P25, P24 and P26 with respectively 45%, 10% and 45%.

The cost parameters are shown in columns, for example Producer/Archive Strength. For each parameter the value Low, Medium or High, applying to all conditions, is stated. For each task the extent to which the task in question is affected by the cost parameter is stated, i.e. irrespective of whether its value is Low, Medium or High.

For example, Producer/Archive Strength can generally be “Low”, which means that Producer does not have much strength in relation to Archive. For some tasks this parameter is of great significance to cost, while for others it has either low, or no, influence.

The value of the impact is determined for each column. For example, for Producer/Archive Strength for Low it is respectively 4%, 6% and 9% (L, M, H); for Medium respectively 20%, 30% and 45%, and for High respectively 100%, 150% and 225%.

Depending on which parameters are of significance to which tasks, this results in a cost supplement which is the basis for compiling the cost of preparing a submission agreement.

In addition to the above tasks, there are Actions in the Transfer Phase (T) and Actions in the Validation Phase (V). These are not included in the basis for compilation of the costs of preparing a submission agreement, but are the implementation of parts of this agreement.

10 Conclusions

In overall terms, we believe that the method of identifying costs is viable, although the cost model is not yet sufficiently detailed to give accurate results for all types of records. All empirical data are from public archive materials, which entails that the model is currently best suited to estimate the costs for this particular type of material.

Generally, our work on the costs of digital preservation shows that preservation institutions depend to a great degree on being able to use standardised solutions, as it would be very expensive for them to develop all solutions themselves.

In terms of the first phase of the project, which was to expand the model with the OAIS function Ingest, the project spent a relatively large amount of time on analysing the PAIMAS standard. The analysis work resulted in a detailed list of activities and a review of the variables that affect the costs of the individual activities in a submission. The variables included the size of the project and the digital collections, the homogeneity and complexity of the digital collections in terms of content, structure and format, and the impact of such issues as legislation and security.

An important conclusion from the work on Ingest, as confirmed by the questionnaire survey, is that normalisation of formats is the highest Ingest cost, namely around two thirds of the total costs. The fact that normalisation also entails cost-sensitive choices such as migration frequency (when should it take place?), migration level (to which format?), player (who is to perform migration?), and error tolerance (which and how many errors are tolerated?), emphasises that this particular cost requires very particular focus in consideration of the precision of the cost model, but also in relation to the choice of migration strategy, cf. Chapter 6.2, When should migration take place?, page 3.

Producers can contribute to reducing the costs of normalisation if they include the cost of future normalisation in their everyday use, in particular how they will migrate from use format to preservation format.

The project's analyses also indicate that the division of strength between producer and preservation institution has great influence on the costs, and their distribution, so that this is an essential parameter in the model.

The cost model's method has been developed further in the second phase of the project.

A key conclusion is that the choice of the digital object (the format), its complexity and volume as the calculation basis makes the model potentially generic and thereby capable of calculating the costs for various digital collections.

In order to achieve accurate results for all types of digital records there is, however, a need to expand with several parameters for each object type, for example number of objects.

Another conclusion concerning the method is that it is dependent on a number of estimates such as format lifetime and thereby migration frequency. In order to address this problem the model makes it possible to state other values than those proposed by CMDP.

From a sound methodological starting point we believe that the most important remaining aspect of the model is its lack of precision. We believe that precision can be enhanced by adding additional parameters and by testing the model with the help of collected cost data.

Implementation of the model in the spreadsheet has proved to be problematic. The requirements of transparency and precision cannot be fulfilled simultaneously. In a new version of the model, with greater precision, it will therefore be necessary to sacrifice transparency and state the formulæ in code.

An actual user guide to the spreadsheet, or the establishment of a user interface, is another deficiency, as the model in its current form is very difficult for external parties to use.

For the purpose of further development of the model the project has stayed abreast of the international development of cost models for digital preservation. The possibility exists that this focus will lead to formal or informal cooperation in the future, as it is assessed that both the interest in and the necessity of greater certainty in this field are generally considered to be substantial.

All in all, the CMDP project (projects #1 and #2) has spent 24 man-months (DKK 1.2 million). For comparison, the LIFE project has spent 42 man-months (DKK 4.6 million), and the KRDS project 14 man-months (DKK 1.0 million), and NASA CET has since 2002 spent 216 man-months (DKK 13.5 million).

If CMDP is to be completed in terms of both method and precision, and expanded to include not only all OAIS functions, but also the opportunity for calculation of different preservation strategies and different digital collections, for example, additional resources are needed.

11 Litteraturliste

Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., & Wheatley, P. 2008. The LIFE2 Final Project Report. http://eprints.ucl.ac.uk/11758/

Beagrie, N., Chruszcz, J., & Lavoie, B. 2008. Keeping Research Data Safe. A Cost Model and Guidance for UK Universities, Copyright HEFCE 2008.
http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf

Beagrie, N., Lavoie, B., Woollard, M., Keeping Research Data Safe 2, Final Report, Charles Beagrie Limited, 2010,
http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

Björk, B.C., Economic evaluation of LIFE methodology, Research report. LIFE project, 2007, p. 5-6, 18-19. http://eprints.ucl.ac.uk/7684/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation, Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008. http://brtf.sdsc.edu/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Sustainable Economics for at Digital Plant: Ensuring Long-Term Access to Digital Information, Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010,
http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Boehm, B., Abts, C., Winsor Brown, A., Chulani, S., Clark, B. K., Horowitz, E., Madachy, R., Reifer, D.J., Steece, B. 2000. Software cost estimation with COCOMO II. Englewood Cliffs, NJ:Prentice-Hall, ISBN 0-13-026692-2

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval), Review of the State of the Art, 2007, http://www.casparpreserves.eu/Members/cclrc/Deliverables/review-of-state-of-the-art-1/at_download/file.pdf

Center for Research Libraries (CRL) & RLG OCLC Programs, Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist, Version 1.0., 2007 http://www.dcc.ac.uk/tools/trustworthy-repositories/

Charles Beagrie Limited and JISC, Keeping Research Data Safe Factsheet,
http://www.beagrie.com/KRDS_Factsheet_0910.pdf

Construx Estimate (værktøj til estimering af omkostninger forbundet med udvikling af programmer),
http://www.construx.com/Page.aspx?nid=68

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, 2002, (ISO14721:2003).
http://public.ccsds.org/publications/archive/650x0b1.pdf

Digital Curation Centre (DCC) & DigitalPreservationEurope (DPE), Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), Version 1.0., 2007 http://www.repositoryaudit.eu/

Egger, A., Shortcomings of the Reference Model for an Open Archival Information System (OAIS). TCDL Bulletin Vol. 2(2). 2006. http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html

Foxite.com, Software Development Costs and Choice of Tools, http://www.foxite.com/articles/read.aspx?id=37&document=software-development-costs-and-choice-of-tools

Grace, S., Knight, G. Montague, L., Investigating the Significant Properties of Electronic Content over Time (InSPECT), Final Report, 2009, http://www.significantproperties.org.uk/

Higgins, S. & Boyle, F., Response to CCSDS's comments on the OAIS Five-year review: recommendations for update, The Digital Curation Centre (DCC) and The Digital Preservation Coalition (DPC), 2006, http://www.dpconline.org/events/previous-events/427-oais-5-year-review-follow-up

Karjalainen, M., Large-scale migration to an open source office suite: An innovation adoption study in Finland, Academic Dissertation, Faculty of Information Sciences of the University of Tampere, 2010, http://acta.uta.fi/pdf/978-951-44-8216-8.pdf

Hole, B., Lin, L., McCann, P. & Wheatley, P., LIFE3: A Predictive Costing Tool for Digital Collections, In: Proceedings of iPRES 2010, 7th International Conference on Preservation of Digital Objects, Austria http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/hole-64.pdf

Marshall, C., How People Manage Information over a Lifetime, 2008, http://www.csdl.tamu.edu/~marshall/PIM%20Chapter-Marshall.pdf

McLeod, R., Wheatley, P., & Ayris, P. 2006. Lifecycle information for e-literature: full report from the LIFE project. Research report. http://eprints.ucl.ac.uk/1854/

NASA Cost Estimation Toolkit (CET). http://opensource.gsfc.nasa.gov/projects/CET/CET.php

Nationaal Archief, Digital Preservation Testbed, Cost of Digital Preservation, 2005. http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf. Testbed computational model: http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=185&categorie=6
Both links are broken.

OECD, International Standard Cost Model Manual, International SCM Network to reduce administrative burdens,2004, http://www.oecd.org/dataoecd/32/54/34227698.pdf
Erhvervs- og Selskabsstyrelsen (www.amvab.dk) vedligeholder forskellig dokumentation og en dansk udgave: Manual til Aktivitetsbaseret Måling af Virksomhedernes Administrative Byrder (AMVAB): http://www.amvab.dk/graphics/DK/Metode%20info/AMVAB-manual-okt2010.pdf

Officeshots.org, Creates transparency in the Office market, http://www.officeshots.org/pages/about

Planets. 2007. Report on tool and service approach, pp. 1, 32-33. http://www.planets-project.eu/docs/reports/Planets_PA4-D1_ReportOnToolAndServiceApproach-Final_Public.pdf

Rosenthal, D.S.H., Robertson T., Lipkis, T., Reich V., Morabito, S., Requirements for Digital Preservation Systems. A Bottom-Up Approach. D-Lib Magazine, Vol. 11 No. 11, ISSN 1082-9873, 2005, http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

Sanett, S. 2002. Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity. College & Research Libraries 63, 5, pp. 388-404.

Shenton, H., Life Cycle Collection Management, LIBER QUARTERLEY, 13, pp 294-272, ISSN 1435-5205, 2003. http://liber.library.uu.nl/publish/articles/000033/article.pdf

Software Technology Support Center (STSC) Cost Analysis Group, Software Development Cost Estimating Guidebook, http://www.stsc.hill.af.mil/consulting/sw_estimation/SoftwareGuidebook2010.pdf

Stephens, A., The application of life cycle costing in libraries. British Journal of Academic Librarianship 3, pp. 82-88, 1988.

Stephens, A., The application of life cycle costing in libraries: a case study based on acquisition and retention of library materials in the British Library. IFLA journal, 1994, 20(2) pp. 130-140

Stocholm, J. L., SaaS og dokumentformater, Version2, 2010, http://www.version2.dk/artikel/16409-saas-og-dokumentformater

Watson, J., *The LIFE project research review. Mapping the landscape, riding a life cycle, 2005.* http://eprints.ucl.ac.uk/1856/

Wheatley, P.: Costing the Digital Preservation Lifecycle More Effectively, iPRES2008 Conference, 2008 http://www.bl.uk/ipres2008/presentations_day1/19_Wheatley.pdf

11 Ordliste

English	Danish	Definition, English	Definition, dansk
Access	**Tilgængeliggørelses-funktions-enheden**	This entity provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.	I OAIS den funktionsenhed, som omfatter de services og funktioner, som gør arkivbeholdningen synlig og tilgængelig for brugerne. Funktionen kan således oplyse brugerne om et arkivalies eksistens, beskrivelse, placering og tilgængelighed. Tilgængeliggørelsesfunktionen gør det ydermere muligt for brugerne at anmode om og konsultere arkivalier. Funktionen skal tillige sørge for en sikkerhedsopmærkning af beskyttede arkivalier.
Administration	**Administrationsfunktionsenheden**	This entity provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve archive operations, and to inventory, report on, and migrate/update the contents of the archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests.	I OAIS den funktionsenhed, som indeholder de services og funktioner, der er behov for i forbindelse med styringen af det OAIS kompatible arkivs andre funktionsenheder og som servicerer dets overordnede operationelle niveau. Administrationsfunktionerne udarbejder retningslinjer for afleveringsbestemmelser, kontrollere afleveringer for at sikre sig, at de overholder afleveringskravene og vedligeholder konfigurationsstyringen af hardware og software i test- og bevaringssammenhænge. Administrationsfunktionen tilvejebringer ligeledes systemudviklerfunktioner med henblik på at overvåge og forbedre de arkivtekniske forhold samt at styre, migrere, ajourføre og indberette om arkivets indhold. Funktionen er også ansvarlig for at etablere og vedligeholde arkivstandarder og -politikker samt at yde brugersupport.
Archival Information Package (AIP)	**Arkiverings-pakke**	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.	I OAIS de informationspakker, bestående af indholdsinformation og bevaringsbeskrivelsesinformation, som arkiveres i arkivet.
Archival Information Update	**Arkivinformationsopdateringsfunktionen**	The **Archival Information Update** function provides a mechanism for updating the contents of the archive. It receives *change requests*, *procedures* and *tools* from Manage System Configuration. It provides updates by sending a *dissemination request* to Access, updating the contents of the resulting *DIPs* and resubmitting them as *SIPs* to Ingest.	I OAIS den funktion under Administration, der tilvejebringer en mekanisme til opdatering af indholdet af arkivet. Den modtager ændringsanmodninger, procedurer og værktøjer. Den opdaterer ved at sende anmodning til Access, ajourfører indholdet af de deraf følgende DIPs og genindsender dem som SIPs.
Archival Storage	**Bitbevaringsfunktions-enheden**	This entity provides the services and functions for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfil orders.	I OAIS den funktionsenhed, som sørger for bevaring samt vedligeholdelse og fremfinding af materialer. Enhedens funktioner inkluderer modtagelse af AIPer fra indtagningsfunktionen, tilføjelse af disse til permanent bitlagring, styring af hierarchical storage management (HSM), udskiftning af lagringsmedier, udførelse af fejltjek, tilstedeværelse af katastrofeberedskab og servicering af Access.
Audit	**Afleverings-**	The **Audit Submission** function will verify	I OAIS den funktion under Administration,

English	Danish	Definition, English	Definition, dansk
Submission	**revisions-funktionen**	that submissions (*SIP or AIP*) meet the specifications of the Submission Agreement. This function receives *AIP/SIP reviews* from Preservation Planning and may also involve an outside committee (e.g., science and technical review). The audit process must verify that the quality of the data meets the requirements of the archive and the review committee. It must verify that there is adequate Representation Information and PDI to ensure that the Content Information is understandable and independently usable to the Designated Community. The formality of the review will vary depending on internal archive policies. The Audit process may determine that some portions of the SIP are not appropriate for inclusion in the archive and must be resubmitted or excluded. An *audit report* is provided to Ingest. After the audit process is completed, any *liens* are reported to the Producer, who will then resubmit the SIP to Ingest or *appeal* the decision to Administration. After the audit is completed, a *final ingest report* is prepared and provided to the Producer and to Negotiate Submission Agreement. Audit methods potentially include sampling, periodic review, and peer review.	som kontrollerer, at SIP eller AIP opfylder specifikationerne i afleveringsaftalen. Denne funktion modtager AIP/SIP reviews fra Bevaringsplanlægningsfunktionen. Revisionsprocessen skal kontrollere, at data opfylder kravene i arkivet. Den skal kontrollere, at der er tilstrækkelig repræsentationsinformation og bevaringsbeskrivende information for at sikre, at oplysninger om indholdet er forståelige og uafhængigt anvendelige til den relevante brugergruppe. Graden af formalitet af denne gennemgang vil variere afhængigt af de interne arkiv politikker. Høringsprocessen kan bestemme, at visse dele af SIP ikke er egnede til optagelse i arkivet og skal forelægges på ny. Høringsmetoder omfatter potentielt stikprøver, periodisk gennemgang og peer review.
Base cost software development	**Basis omkostning software udvikling**	In CMDP one of the parameters used in the calculation of how much time it takes to develop migration software. Default is 4 person weeks (pw) for unformated text	I CMDP et af de parametre, der anvendes til beregning af hvor meget tid det tager at udvikle migreringssoftware. Sat til 4 personuger for uformateret text
Bit preservation	**Bitbevaring**	Bit preservation ensures that the bits remain intact and readable.	Aktiviteter, der sikrer, at bits forbliver intakte og læsbare over tid. Disse aktiviteter varetages i OAIS af funktionsenheden Archival Storage.
Calendar week	**Kalenderuge**	In CMDP the default is 168 hours.	I CMDP sat til 168 timer.
Common Services	**Understøttende funktionsenhed**	**Operating system services** provide the core services needed to operate and administer the application platform, and provide an interface between application software and the platform. **Network services** provide the capabilities and mechanisms to support distributed applications requiring data access and applications interoperability in heterogeneous, networked environments. **Security services** capabilities and mechanisms to protect sensitive information and treatments in the information system. The appropriate level of protection is determined based upon the value of the information to the application end-users and the perception of threats to it.	I OAIS den funktionsenhed som inkluderer:Operativsystemtjenester – er de kerneydelser som er nødvendige for at drive og administrere applikationsplatformen. Netværkstjenester som tilvejebringer mekanismer til understøtning af distribuerede applikationer, der kræver adgang til data og applikationsinteroperabilitet i heterogene, netværksbaserede miljøer. Sikkerhedstjenester - mekanismer til at beskytte følsomme oplysninger i informationssystemet.
Consumer	**Bruger**	The role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISs, as well as internal OAIS	I OAIS de brugere eller systemer, som interagerer med arkivet.

English	Danish	Definition, English	Definition, dansk
		persons or systems.	
Content complexity	**Indholdskompleksitet**	I CMDP the semantic complexity of the data, how hard the data are to understand. E.g.: Understanding biomolecular data is of a higher complexity than understanding an inventory of furniture	I CMDP datas semantiske kompleksitet - hvor svære data er at forstå. Fx: Biomolekylær data har en større kompleksitet end en opgørelse af møbler.
Content Information	**Indholds-information**	The set of information that is the original target of preservation. It is an Information Object comprised of its Content Data Object and its Representation Information. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc.	I OAIS den information, der skal bevares. Indholdsinformationen består af dataobjekt og repræsentationsinformation. Fx en tabel med tal der viser og er forståelig som temperaturer. Ikke dokumentation, der beskriver tabellens historie og oprindelse eller hvordan tallene relaterer til andre observationer mm.
Content variation	**Indholdsvariation**	I CMDP the semantic variation of the data, how diverse the areas are. E.g.: Understanding a submission consisting of biomolecular data, geographical data and Russian poetry has a higher complexity than just understanding Russian poetry.	I CMDP datas semantiske variation - hvor forskellige områderne er.Fx: Forståelse af en aflevering, der består af biomolekylære data, geografiske data og russisk poesi har en større kompleksitet end bare at forstå russisk poesi.
Context Information	**Kontekst-information**	The information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects.	I OAIS den information, der dokumenterer indholdsinformationens relationer til dets miljø. Fx hvorfor indholdsinformationen blev skabt og hvordan den relaterer til anden indholdsinformation.
Coordinate Updates	**Koordinér opdateringer funktionen**	The **Coordinate Updates** function is responsible for transferring the *AIPs* to Archival Storage and the *Descriptive Information* to Data Management. Transfer of the AI*P* includes a *storage request* and may represent an electronic, physical, or a virtual (i.e., data stays in place) transfer. After the transfer is completed and verified, Archival Storage returns a *storage confirmation* indicating (or verifying) the storage identification information for the AIP. The Coordinate Updates function also incorporates the storage identification information into the Descriptive Information for the AIP and transfers it to the Data Management entity along with a *database update request*. In return, Data Management provides a *database update response* indicating the status of the update. Data Management updates may take place without a corresponding Archival Storage transfer when the SIP contains Descriptive Information for an AIP already in Archival Storage.	I OAIS den funktion under Ingest, der koordinerer opdateringer af informationspakkernes metadata og dokumentation og som overfører informationspakkerne til Archival Storage og sender metadata og anden dokumentation til Data Management.
Cost critical activity	**Omkostningskritisk aktivitet**	I CMDP a cost critical activity is a task which is estimated to take more than one person week (pw) per year to complete.	I CMDP er en omkostningskritisk aktivitet en opgave, som anslås at tage mere end én personuge om året til at udføre.
Cost Model for Digital Preservation (CMDP)	**CMDP**	Generic cost model for calculating and estimating cost of long-term digital preservation.	Omkostningsmodel til beregning og estimering af omkostninger for bevaring af digitale materialer over tid.
Cost	**Omkostningsp**	In CMDP the parameters that influence the	I CMDP et parameter som har indflydelse på

English	Danish	Definition, English	Definition, dansk
parameter	**arameter**	cost critical activities and that constitute part of the variables in the spreadsheet formula.	de omkostningskritiske aktiviteter og som udgør en del af variablerne i regnearket.
Data	**Data**	A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.	I OAIS information, der er udtrykt på en formaliseret måde og som kan fortolkes. Eksempler på data er bitsekvenser, en tabel med numre, tegn på en side, en lydoptagelse af en person der taler, en prøve fra en månesten.
Data Management	**Datastyringsfunktions-enheden**	This entity provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. Data Management functions include administering the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading new descriptive information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets.	I OAIS den funktionsenhed, der stiller services til rådighed med henblik på at udfylde, vedligeholde og tilgå både fremfindingsinformationen, som identificerer og dokumenterer arkivalierne, og de administrative data, der bruges til at styre arkivet. Datastyringsfunktionerne inkluderer administration af arkivdatabasefunktionerne (vedligeholdelse af *schemaer* og *view* definitioner samt den referentielle integritet), udførelse af databaseopdateringer (af fremfindingsinformationen og administrative data), udførelse af databaseforespørgsler med henblik på generering af resultatsæt og udarbejdelse af rapporter på baggrund af disse.
Data object	**Dataobjekt**	Either a Physical Object or a Digital Object.	I OAIS enten et fysisk eller et digitalt objekt.
Degree of documentation	**Dokumentationsgrad**	In CMDP this expresses how well the data formats are documented, what formats do data belong to, and how detailed are these formats documented. E.g.: A high level of documentation requires that the producer has documentation for all data and all formats, such as documenting that a certain file is in a Word Perfect 5.1 format and having documentation for that specific format.	I CMDP afspejler det hvor godt dataformater er dokumenteret, og detaljeringsgrad af dokumentationen. Fx: En høj detaljeringsgrad forudsætter at producenten har dokumentation for alle data og alle formater, som f.eks. dokumenterer, at en bestemt fil er i et Word Perfect 5.1 format.
Descriptive Information	**Beskrivende information**	The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers.	I OAIS den information som lagres i Data Management og som bruges til at finde, bestille og hente informationspakkerne. Kan bestå af både en unik identifikator samt titel og andre referencer.
Designated Community	**Brugergruppe**	An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.	I OAIS de brugergrupper, som arkivet har specificeret, som de der skal kunne bruge arkivets materialer.
Destination format	**Udformat**	In CMDP the destination format of a migration.	I CMDP det format, der migreres til.
Develop Packaging Designs and Migration Plans	**Udvikling af informationspakke design og migrerings-plan funktionen**	The Develop Packaging Designs and Migration Plans function develops new IP designs and detailed migration plans and prototypes, to implement Administration policies and directives.	I OAIS den funktion under Preservation Plannning, som udvikler nye designs for informationspakker, migreringsplaner og migreringprogrammer med henblik på at implementere politikker og retningsliner fra Administration.
Develop Preservation Strategies and	**Udvikling af bevaringsstrategier og**	The Develop Preservation Strategies and Standards function is responsible for developing and recommending strategies and	I OAIS den funktion under Preservation Planning, der på baggrund af rapporter fra Monitor Designated Community og Monitor

English	Danish	Definition, English	Definition, dansk
Standards	**standarder funktionen**	standards to enable the archive to better anticipate future changes in the Designated Community service requirements or technology trends that would require migration of some current archive holdings or new submissions.	Technologi, løbende udvikler og anbefaler strategier og standarder for arkivet med henblik på at undgå forældelse.
Digital Preservation	**Digital bevaring**	Digital preservation designates the methods and systems which are needed to assure access to digital materials over time. Digital preservation may be divided into bit preservation, which must ensure that the bits remain intact and readable, and logical or functional preservation, which must ensure that the bits remain comprehensible. The concept digital preservation is also used in a broader context for the whole life cycle process from the production of the digital materials, acquisition by the OAIS archive, and long term preservation, until they are made available to the users.	Bevaring af digitalt materiale. Materialet kan være skabt digitalt eller være resultat af en digitalisering af papirdokumenter, lyd eller levende billeder i analog form m.v.
Error Handling (EH)	**Fejlhåndtering**	In CMDP the time it takes to manually correct flawed files in a migration processing action.	I CMDP den tid det tager manuelt at rette fejlbehæftede filer som følge af en en migreringshandling.
Establish Standards and Policies	**Etabler standarder og politikkerfunktionen**	The Establish Standards and Policies function is responsible for establishing and maintaining the archive system standards and policies. It receives budget information and policies such as the OAIS charter, scope, resource utilization guidelines, and pricing policies from Management. It provides Management with periodic reports. It receives recommendations for archive system enhancement, and proposals for new archive data standards from Preservation Planning. It also receives performance information and archive holding inventories from Manage System Configuration. Based on these inputs, archive standards and policies are established and sent to other Administration functions and the other Functional Entities for implementation. The standards include format standards, documentation standards and the procedures to be followed during the Ingest process. It provides approved standards and migration goals to Preservation Planning. This function will also develop storage management policies (for the Archival Storage hierarchy), Including migration policies to assure that archive storage formats do not become obsolete, and database administration policies. It will develop disaster recovery policies. It will also determine security policies for the contents of the archive, including those affecting Physical Access Control and the application of error control techniques throughout the archive.	I OAIS den funktion under Administration, som er ansvarlig for at etablere og vedligeholde standarder og politikker for hele arkivet, på baggrund af de overordnede retningslinjer (formål og budget) udstukket af Management. I relation til Ingest sender funktionen format og dokumentationsstandarder til Ingest samt de procedurer som skal følges i forbindelse med Ingest.
Fixity information	**Integritets-information**	The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content	I OAIS garantien for, at indholdet i en informationspakke ikke er blevet ændret på udokumenteret vis. Et eksempel på

English	Danish	Definition, English	Definition, dansk
		Information object has not been altered in an undocumented manner. An example is a Cyclical Redundancy Check (CRC) code for a file.	integritetsinformation er resultatet af Cyclical Redundancy Check (CRC) på en fil.
Format charateristics	**Formategenskaber**	The characteristics of each format is given by its complexity, documentation and implementation measured relatively to unformatet text	For hvert format er angivet dets kompleksitet, dokumentation og implementering målt i forhold til formatet og uformateret tekst.
Format Interpretation factor	**Format forståelsesfaktor**	In CMDP it denotes how difficult a format is to understand. It depends on the complexity, the documentation and the implementation. The complexity is based on the time it takes, measured in person-weeks, to identify and read the format specifications and any other relevant documentation. A addition is made if the documentation is not 100% available, covering and implementable, likewise additions are made if the format is not 100% implemented is open source, widespread and in current usage. The complexity level is by default the most important part of the format interpretation factor.	I CMDP angiver faktoren hvor svært et format er at forstå, hvilket afhænger af kompleksiteten, dokumentationen og implementeringen. Kompleksiteten er baseret på den tid som det tager målt i personuger at identificere og læse formatspecifikationen og anden relevant dokumentation. Der er et tillæg, hvis dokumentationen ikke er 100 % tilgængelig, dækkende og implementerbar, ligeledes er der tillæg hvis implementering ikke findes 100% i åben kildekode, er udbredt og i aktuel brug. Som udgangspunkt er det kompleksiteten, der udgør det vigtigste aspekt af forståelsesfaktoren.
Format Life Time (Expectancy)	**Formaters forventede levetid**	In CMDP the average estimated lifetime of a format has been calculated, based on complexity, documentation and implementation. In this version of CMDP due to uncertainty the default max life time is set to 8 years.	CMDP estimerer et formats forventede levetid ud fra dets kompleksitet, dokumentation og udbredelse. I denne udgave af CMDP er der grundet usikkerhed sat en længste levetid for alle formater på 8 år.
Functional preservation	**Funktionel bevaring**	Functional or logical preservation ensures that the bits remain comprehensible in the long term.	I digital bevaring er funktionel – eller logisk – bevaring en sikring af, at den bevarede fil kan fortolkes over tid.
Generate AIP	**Generér AIP**	The **Generate AIP** function transforms one or more *SIPs* into one or more *AIPs* that conform to the archive's *data formatting and documentation standards*. This may involve file format conversions, data representation conversions or reorganization of the content information in the SIPs. The Generate AIP function may issue *report requests* to Data Management to obtain *reports* of information needed by the Generate AIP function to produce the Descriptive Information that completes the AIP. This function sends *SIPs or AIPs for audit* to the Audit Submission function in Administration, and receives back an *audit report*.	I OAIS den funktion under Ingest, der genererer AIPer, herunder omformer migrerer SIPer til AIPer. Omformningen Migreringen af pakker kan indebære, at formater og/eller strukturer migreres (normaliseres) til godkendte bevaringsformater og/eller strukturer, at repræsentationsdata ændres og/eller at informationen i pakkerne reorganiseres. Funktionen kan også bede Data Management om yderligere information, der er nødvendig for beskrive pakken fuldstændigt.
Generate Descriptive Information	**Generér beskrivende information**	The **Generate Descriptive Information** function extracts Descriptive Information from the *AIPs* and collects *Descriptive Information* from other sources to provide to Coordinate Updates, and ultimately Data Management. This includes metadata to support searching and retrieving AIPs (e.g., who, what, when, where, why), and could also include special browse products (thumbnails, images) to be used by Finding Aids.	I OAIS den funktion under Ingest, der genererer beskrivende information, dvs. metadata og dokumentation om informationspakkerne. Den beskrivende information trækkes fra pakken selv eller tilføres pakken fra Data Management.

English	Danish	Definition, English	Definition, dansk
Generate Report	**Generer rapport funktionen**	The Generate Report function receives a report request from Ingest, Access or Administration and executes any queries or other processes necessary to generate the report that it supplies to the requester. Typical reports might include summaries of archive holdings by category, or usage statistics for accesses to archive holdings. It may also receive a report request from Access and provides descriptive information for a specific AIP.	I OAIS den funktion under Data Management, der modtager bestillinger på rapporter fra Ingest, Access eller Administration, og som udfører forespørgsler og andre processer, som er nødvendige for at levere rapporten.
Influence on Consumption	**Indflydelse på brugere**	In CMDP the influence the OAIS archive has on what formats the users use. We assume, for example, that the entity Monitor Designated Community depends on how much influence the archive has on the production and use of formats: The more influence, the fewer costs.	I CMDP den indflydelse OAIS arkivet har på hvilke formater brugerne bruger.
Influence on Production	**Indflydelse på produktion**	In CMDP the influence the OAIS archive has on what formats the producers use. We assume, for example, that the entity Monitor Designated Community depends on how much influence the archive has on the production and use of formats: The more influence, the fewer costs.	I CMDP den indflydelse OAIS arkivet har på hvilke formater producenterne bruger.
Information	**Information**	Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret a string of bits as numbers representing temperature observations measured in degrees Celsius (the representation information).	I OAIS alle typer viden, der kan udveksles. I en udveksling er information repræsenteret af data. Et eksempel på information er en bitsekvens (data) og den tilhørende beskrivelse af hvordan bitsekvensen skal fortolkes som numre, der udtrykker temperatur observationer målt i celsius grader (repræsentationsinformationen). Svarer til OAIS begrebet indholdsinformation.
Information package	**Informations-pakke**	The Content Information and associated Preservation Description Information which is needed to aid in the preservation of the Content Information. The Information Package has associated Packaging Information used to delimit and identify the Content Information and Preservation Description Information.	I OAIS består en informationspakke af indholdsinformation og bevaringsbeskrivelsesinformation, dvs. den information, der er nødvendig for at bevare indholdsinformationen. Informationspakkerne selv beskrives af pakkeinformation, der identificerer og afgrænser indholdsinformation og bevaringsbeskrivelsesinformationen.
Ingest	**Indtagning**	This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management.	I OAIS den funktionsenhed, som yder de services, som gør det muligt for et arkiv, at modtage informationspakker fra producenter og forberede dem til videre bevaring. Funktionerne inkluderer: Modtagelse af informationspakker (SIP), kvalitetssikring af informationspakkerne (SIP), generering af arkiveringsegnede informationspakker (AIP), udtrækning af beskrivende information til brug i arkivdatabaserne, koordinering af ajourføringer i forhold til Archival Storage og Data Management.
Machine processing	**Maskintid**	I CMDP the machine time is based on the complexity of the source and destination formats relative to unformated text. Default	I CMDP er maskintiden baseret på kompleksiteten af indformat og udformat i forhold til uformateret tekst. Standardværdien

English	Danish	Definition, English	Definition, dansk
time		for unformated text is 1 min/GB.	for uformateret tekst er 1 min/GB.
Management	**Ledelse**	The role played by those who set overall OAIS policy as one component in a broader policy domain.	I OAIS den rolle, som fastsætter de overordnede OAIS politikker som en del af en større strategi.
MB/s	**MB/s**	Megabytes per second.	Megabytes pr. sekund.
Migration factor	**Migreringsfaktor**	In CMDP the migration factor is the cost in person weeks of providing software for migration of a format.	I CMDP angiver migreringsfaktoren omkostningen i personuger til at tilvejebringe programmel til migrering af et format
Migration frequency	**Migreringsfrekvens**	In CMDP the frequency of migration is based on average estimated lifetime of formats, which we for simplicity and to be conservative have set to be 8 years, based on the lifetime of the migrations software.	I CMDP er migreringsfrekvensen baseret på formatets levetid, men af hensyn til enkelhed i modellen og forsigtighed sat til hvert 8. år for alle formater baseret på levetiden af migreringssoftwaren.
Monitor Designated Community	**Brugerover-vågning**	The **Monitor Designated Community** function interacts with archive Consumers and Producers to track changes in their *service requirements* and available *product technologies*. Such requirements might include data formats, media choices, preferences for software packages, new computing platforms, and mechanisms for communicating with the archive.	I OAIS den funktion under Preservation Planning, der følger udviklingen i producenters og brugeres krav til arkivets services og teknologi, og rapportere ændringer til Develop Preservation Strategies and Standards.
Monitor Technology	**Teknologiover vågning**	The Monitor Technology function is responsible for tracking emerging digital technologies, information standards and computing platforms (i.e., hardware and software) to identify technologies which could cause obsolescence in the archive's computing environment and prevent access to some of the archives current holdings.	I OAIS den funktion under Preservation Planning, der løbende overvåger de formater, strukturer og systemer som arkivet benytter, og rapportere ændringer til Develop Preservation Strategies and Standards.
Negotiate Submission Agreement	**Forhandling af afleveringsaftale**	The **Negotiate Submission Agreement** function solicits desirable archival information for the OAIS and negotiates *Submission Agreements* with Producers. This function also negotiates a *data submission schedule* with the Producer. It maintains a calendar of expected Data Submission Sessions that will be needed to transfer one or more SIPs to the OAIS and the resource requirements to support their ingestion. This function receives *AIP/SIP templates* and *customization advise* from Preservation Planning and sends *SIP designs* and *SIPs* to the Audit Submission function as part of the submission approval process. The data submission formats and procedures must be clearly documented in the archive's data submission policies, and the deliverables must be identified by the Producer in the Submission Agreement.	I OAIS den funktion under Administration, der forhandler afleveringsaftalen mellem producent og arkiv på plads.
Number of Computers (NoC)	**Antal computere**	In CMDP NoC refers to the number of computers used in a migration processing.	I CMDP er NoC det antal computere, der benyttes til migreringen.
Number of Pages (NoP)	**Antal sider**	In CMDP NoP refers to the number of pages it is necessary for a computer scientist to read and comprehend in order to be able to develop migration software for the format.	I CMDP henviser NoP til det antal sider det er nødvendigt for en datalog til at læse og forstå for at kunne udvikle migreringssoftware.
OAIS	**OAIS**	An archive, consisting of an organization of people and systems that has accepted the	Et Open Archival Information System (OAIS) er et arkiv, der har accepteret ansvaret for at

English	Danish	Definition, English	Definition, dansk
		responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.	bevare information og gøre denne tilgængelig for en specificeret brugergruppe ved at leve op til ISO standarden OAIS (14721:2003).
PAIMAS	**PAIMAS**	The purpose of the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) is to identify, define and provide structure to the relationships and interactions between a Producer and an Archive. The standard defines the methodology to follow to manage the activities and the interfaces from the initial contact between the Producer and the Archive until the objects of information are received and validated by the Archive. These activities cover the first stage of the Ingest Process as defined in the Reference Model for an Open Archival Information System (OAIS).	PAIMAS er en ISO standard, hvis formål er at identificere, definere og strukturere relationer og interaktioner mellem en producent og et arkiv. Standarden definerer en metode til styring af aktiviteter mellem producenten og arkivet. Disse aktiviteter omfatter første fase af indtagningsprocessen (Ingest + Administration).
Periodic	**Periodisk**	In CMDP a cost that only occurs with a certain time span	I CMDP omkostninger som kun forekommer med et givent interval
Person week (PW)	**Personuge**	In CMDP one person week is set to 32 hours.	I CMDP svarer en personuge til 32 timer.
Preservation Description Information	**Bevarings-beskrivende information**	The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.	I OAIS den der er nødvendig for tilfredsstillende bevaring af bevaringsværdige data. Den består af proveniensinformation, referenceinformation, integritetsinformation og kontekstinformation.
Preservation Planning	**Bevarings-planlægningsfunktions-enheden**	This entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs IP templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals.	I OAIS den funktionsenhed, der overvåger arkivmiljøet og rådgiver om, hvordan det sikres, at digitale arkivalier forbliver tilgængelige over tid, selvom det oprindelige computermiljø forældes. Enheden evaluerer arkivets indhold og rådgiver om, hvornår det er nødvendigt at ty til f.eks. migrering (se migreringsstrategi) samt hvordan dette bør gøres. Enheden udarbejder også anbefalinger ift. arkivstandarder og -politikker og overvåger teknologiforandringer og ændringer i brugernes adfærd. Bevaringsplanlægningsfunktionen designer også informationspakker til specifikke datamodeller. Endelig fremstiller den detaljerede bevaringsplaner, software prototyper og testplaner med henblik på at muliggøre implementeringen af bevaringsinstitutionens bevaringsstrategi.

English	Danish	Definition, English	Definition, dansk
Producer	**Producent**	The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems.	I OAIS de personer eller systemer, der leverer den information til arkivet, som skal langtidsbevares.
Producer / Archive strength	**Producent / Arkiv styrkeforhold**	In CMDP this indicates the formal (legal) relationship between Producer and Archive E.g. the influence A has on P in the choice of preservation formats.	I CMDP angiver det det formelle (juridiske) forhold mellem producent og arkiv. F.eks. den indflydelse arkivet på producentens valg af bevaringsformater
Provenance Information	**Proveniens-information**	The information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. Examples of Provenance Information are the principal investigator who recorded the data, and the information concerning its storage, handling, and migration.	I OAIS den information, der dokumenterer indholdsinformationens historie: Beskriver indholdsinformationens oprindelse eller kilde, ændringshistorik, ejerskabshistorik. Proveniensinformation kan fx være hvem, der har optaget data og hvordan indholdsinformationen har været lagret, håndteret og migreret.
Quality Assurance	**Kvalitetssikring**	The **Quality Assurance** function validates (*QA results*) the successful transfer of the *SIP* to the staging area. For digital submissions, these mechanisms might include Cyclic Redundancy Checks (CRCs) or checksums associated with each data file, or the use of system log files to record and identify any file transfer or media read/write errors.	I OAIS den funktion under Ingest, der kontrollerer at alle bits er intakte og fuldstændige efter overførsel af SIP'er til arkivet. Kontrollen sker typisk vha. checksumsalgoritmer. Kvalitetssikringen omfatter ikke kontrol af pakkens indholdsmæssige fuldstændighed.
Receive Data	**Datamodtag-elsesfunktionen**	The Receive Data function receives a storage request and an AIP from Ingest and moves the AIP to permanent storage within the archive. The transfer request may need to indicate the anticipated frequency of utilization of the data objects comprising the AIP in order to allow the appropriate storage devices or media to be selected for storing the AIP. This function will select the media type, prepare the devices or volumes, and perform the physical transfer to the Archival Storage volumes. Upon completion of the transfer, this function sends a storage confirmation message to Ingest, including the storage identification of the AIPs.	I OAIS den funktion under Archival Storage, som modtager anmodning om bitbevaring og AIPer fra Ingest og overfører pakkerne til arkivets det permanente lager.
Receive Database Updates	**Modtag databaseopdateringer funktionen**	The Receive Database Updates function adds, modifies or deletes information in the Data Management persistent storage. The main sources of updates are Ingest, which provides Descriptive Information for the new AIPs, and Administration, which provides system updates and review updates. Ingest transactions consist of Descriptive Information which identifies new AIPs stored in the archive. System updates include all system-related information (operational statistics, Consumer information, and request status). Review updates are generated by periodic reviewing and updating of information values (e.g., contact names, and addresses). The Receive Database Updates function provides regular reports to Administration summarizing the status of updates to the database, and also sends a	I OAIS den funktion under Data Management, der opdaterer metadata og anden dokumentation lagret i Data Management.

English	Danish	Definition, English	Definition, dansk
		database update response to Ingest.	
Receive Submission	**Modtag afleverings-funktionen**	The **Receive Submission** function provides the appropriate storage capability or devices to receive a *SIP* from the Producer (or from Administration). Digital SIPs may be delivered via electronic transfer (e.g., FTP), loaded from media submitted to the archive, or simply mounted (e.g., CD-ROM) on the archive file system for access. Non-digital SIPs would likely be delivered by conventional shipping procedures. The Receive Submission function may represent a legal transfer of custody for the Content Information in the SIP, and may require that special access controls be placed on the contents. This function provides a *confirmation of receipt* of a SIP to the Producer, which may include a *request to resubmit* a SIP in the case of errors resulting from the SIP submission.	I OAIS den funktion under Ingest, der modtager afleveringer og placerer dem på et midlertidigt lager, hvor overførslen kvalitetssikres. Funktionen sender også en kvittering på modtagelsen tilbage til producenten. Hvis afleveringen er mangelfuld eller fejlbehæftet, anmoder funktionen producenten om at genoverføre afleveringen.
Reference Information	**Reference-information**	The information that identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides identifiers that allow outside systems to refer, unambiguously, to a particular Content Information. An example of Reference Information is an ISBN.	I OAIS den information, der identificerer og evt. beskriver en eller flere mekanismer, der bruges til at tildele globalunikke persistente identifikatorer til indholdsinformationen. Herunder identifikatorer, der gør det muligt for eksterne systemer at referere entydigt til indholdsinformationen. Referenceinformation kan fx være et ISBN nummer.
Representation Information	**Repræsentations-information**	The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.	I OAIS den information, der forbinder et dataobjekt til mere meningsfuld information – og som er nødvendig for at brugerne kan forstå dataobjekterne. Repræsentationsinformation giver mening (semantik) og format (struktur) til dataobjektets digitale bits, således at det sikres, at et dataobjekt kan forstås af mennesker eller maskiner på sigt. Den strukturelle information om et dataobjekt oversætter objektets bits til noget meningsfuldt, f.eks. værdier som bogstaver, tal, pixels, tabeller, osv. Den semantiske information er f.eks. information om, hvilket sprog teksten er skrevet i, hvad den indeholdte videnskabelige terminologi betyder. Et eksempel er, at en bitsekvens tolkes af den strukturelle information som værende en række af tal: 2, 1, 3, 8, 14, 16, 18, 19, 13, 7, 4, 3. Dem semantiske information fortæller os, at der er tale om grader i celsius, og at hvert tal i den angivne rækkefølge er et givet lands månedlige gennemsnitstemperatur.
Salary, high	**Løn, høj**	In CMDP one of the three salary levels. The default value is 900 E/PW.	I CMDP et af de tre lønniveauer. Sat til 900 euro / personuge.
Salary, low	**Løn, lav**	In CMDP one of the three salary levels. The default value is 300 E/PW.	I CMDP et af de tre lønniveauer. Sat til 300 euro / personuge.
Salary, medium	**Løn, middel**	In CMDP one of the three salary levels. The default value is 600 E/PW.	I CMDP et af de tre lønniveauer. Sat til 600 euro / personuge.
Software	**Software**	In CMDP the Software Provision factor is the	I CMDP angiver denne faktor omkostningen i

English	Danish	Definition, English	Definition, dansk
Provision factor	**tilvejebringelsesfaktor**	cost (in person weeks) of providing migration software for a chosen combination of source and destination format. Provision is made by using existing software or developing software and following implementing it in the migration system. The cost is dependent on the complexity and documentation of the source and destination formats. The format software life time is set to 8 years.	personuger for at tilvejebringe migrationssoftware for en valgt kombination af indformat og udformat. Tilvejebringelsen sker ved at fremskaffe bestående software eller udvikle den, og dernæst indføre den i migreringssystemet. Omkostningen afhænger af kompleksiteten og dokumentationen for formaterne. Levetiden for formatsoftware er sat til 8 år.
Source format	**Indformat**	In CMDP the format to be migrated.	I CMDP det format, som skal migreres.
Stages	**Trin**	In CMDP each stage corresponds to a functional entity in OAIS, ie. Ingest, Access, etc.	I CMDP svarer hvert trin til en funktionel enhed i OAIS, f.eks. Access.
Submission Agreement	**Afleverings-aftale**	An agreement (OAIS) concluded between the producer and the archive regarding a submission. The agreement defines all matters pertaining to a submission project, including which data and documentation to be preserved, legal, contractual and security aspects, procedures for transfer and admission, budget and time schedule.	I OAIS en aftale, der indgås mellem producent og arkiv i forbindelse med en aflevering. Aftalen definerer alle relevante forhold i forbindelse med et afleveringsprojekt, herunder hvilke data og anden dokumentation skal bevares, lov-, kontrakt- og sikkerhedsmæssige forhold, overførsel og godkendelsesprocedurer, samt budget og tidsplan.
Submission Information Package (SIP)	**Afleverings-pakke**	An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.	I OAIS de informationspakker som producenten afleverer til arkivet.
Supervision	**Overvågning**	I CMDP supervision designates how much manual work it takes (calculated in person weeks) to supervise the processing of a migration. The model estimates the cost of man power monitoring this process to be 5% of the machine processing time.	I CMDP angiver overvågning hvor meget manuelt arbejde, det kræver (beregnet i personuger) at overvåge processeringen af en migrering. Modellen anslår udgifterne til 5 % af maskintiden.
System/ Format complexity	**System/format kompleksitet**	In CMDP the syntactical complexity of the data, how complex are the systems and formats. E.g.: Multidimension formats for geodata, biochemical structures are much more complex than formats for 2-dimensional graphics	I CMDP datas/systemers syntaktiske kompleksitet. Fx: Multidimensionale formater for geodata og biokemiske strukturer er langt mere komplekse end formater til 2-dimensionel grafik.
System/ Format variation	**System/format variation**	In CMDP the syntactical variation in the systems and formats to be submitted. E.g.: many different file formats or data structures, such as various formats for sound, also includes variation in complexity of formats	I CMDP den syntaktiske variation i systemer og formater, der skal indtages. Fx: mange forskellige filformater eller datastrukturer, såsom forskellige formater til lyd, inkluderer også variation i formatkompleksitet.
Time per Page (TpP)	**Tid pr. side**	In CMDP this refers to the time it takes a computer scientist to read and comprehend a page in a format's documentation in order to be able to develop migration software for the format. The parameter is set to 20 minutes per page and is regulated by the format's complexity, thus 20 minutes times 0%, 25% or 50%.	I CMDP henviser det til den tid, det tager en datalog til at læse og forstå en side i et formats dokumentation for at kunne udvikle migreringssoftware. Parameteren er sat til 20 minutter per side, og er reguleret af formatkompleksitet samt basis tid, og er dermed 20 minutter gange 0 %, 25 % eller 50 %.
Total Format Interpretation factor	**Total format forståelses-faktor**	In CMDP the Total Format Interpretation factor is the sum of an institution's format interpretation factors, each of which pertains to a specific format.	I CMDP er den totale forståelsesfaktor summen af forståelsesfaktorer for hvert format som haves i en bevaringsinstitution.

English	Danish	Definition, English	Definition, dansk
Transparency of producer	**Producentens gennemsigtighed**	In CMDP it designates that the processes and activities of an organization's record keeping program shall be documented in an understandable manner and be available to all personnel and appropriate interested parties. E.g. : Existence of a record keeping program, documentation of processes	I CMDP udtryk for at en organisations digitale forvaltning skal være dokumenteret på en forståelig måde og være tilgængelig for alle medarbejdere og relevante berørte parter.
Year	**År**	In CMDP this refers to the cost for a specific year, consisting of annual and periodic cost.	I CMDP refererer det til omkostninger for et konkret år, bestående af årlige omkostninger og periodiske.

12 Appendix

12.1 File format support in Microsoft Excel 2010

Historically Excel has supported many different data formats. We have determined that a number of these older formats are seldom, if ever used. We are removing support for some file types to allow us to devote more of our efforts towards the file formats that are being used. Theses formats are being deprecated in 2 ways. For the set of file formats with the lowest usage, we will be discontinuing support for opening and saving of these formats. For the second set that has some minimal usage, we will support loading the files in Excel 2007 to allow you to save them in a newer format.

The following formats cannot be opened or saved in Excel 2007:
WK1 (1-2-3), WK4 (1-2-3), WJ3 (1-2-3 Japanese) (.wj3), WKS (1-2-3), WK3,(1-2-3), WK1,FMT(1-2-3), WJ2 (1-2-3 Japanese) (.wj2), WJ3, FJ3 (1-2-3 Japanese) (.wj3), DBF 2 (dBASE II), WQ1 (Quattro Pro/DOS), WK3,FM3(1-2-3), Microsoft Excel Chart (.xlc), WK1,ALL(1-2-3), WJ1 (1-2-3 Japanese) (.wj1), WKS (Works Japanese) (.wks).

The following formats may be opened, but not saved to in Excel 2007:
Microsoft Excel 2.1 Worksheet, Microsoft Excel 2.1 Macro, Microsoft Excel 3.0 Worksheet, Microsoft Excel 3.0 Macro, Microsoft Excel 4.0 Worksheet, Microsoft Excel 4.0 Macro, Microsoft Excel 97- Excel 2003 & 5.0/95 Workbook, Microsoft Excel 4.0 Workbook, DBF 3 (dBASE III), DBF 4 (dBASE IV).

Microsoft Excel 2010 - The official blog of the Microsoft Excel product team. Deprecated features for Excel 2007 by David Gainer, 24 Aug 2006 2:18 PM[11]

[11] http://blogs.msdn.com/b/excel/archive/2006/08/24/718786.aspx

12.2 Cost of Ingest in the Danish National Archives

	2008			2007			2006			2005			2004			2003			2002			2001		
A. Approval and Disposal																								
	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit	Production	Hours	Per unit
Registries	16	506,5	31,7	59	862,0	14,6	73	352,0	4,8	84	359,0	4,3	65	269,4	4,1	31	205,3	6,6	105	420,0	4,0	120	466,0	3,9
Electronic filing systems	13	132,3	10,2	14	120,0	8,6	13	213,0	16,4	59	280,5	4,8	24	328,8	13,7	51	460,3	9,0	40	674,3	16,9	25	434,0	17,4
Electronic document managemen	32	665,8	20,8	24	621,0	25,9	22	576,0	26,2	22	425,5	19,3	22	641,8	29,2	19	978,8	51,5	8	261,4	32,7	9	399,0	44,3
Electronic document managemen	10	356,0	35,6	10	702,0	70,2	18	689,0	38,3	2	139,5	69,8												
B. Submission preparation																								
Registries	148	1.577,3	10,7	138	1.804,0	13,1	88	1.510,0	17,2	109	1.135,0	10,4	224	1.649,5	7,4	202	1.673,0	8,3	122	1.862,0	15,3	118	1.448,0	12,3
Electronic filing systems	81	819,5	10,1	38	983,0	25,9	59	883,5	15,0	48	533,0	11,1	44	489,0	11,1	33	636,4	19,3	34	688,0	20,2	103	692,0	6,7
Electronic document managemen	22	756,5	34,4	19	359,0	18,9	9	320,5	35,6	7	213,0	30,4	7	176,0	25,1	3	119,5	39,8	3	57,0	19,0	0	22,0	
C. Testing																								
Registries	203	2.548,5	12,6	219	2.956,0	13,5	66	2.142,5	32,5	118	2.005,0	17,0	225	3.050,0	13,6	110	1.459,5	13,3	134	1.621,0	12,1	56	1.717,0	30,7
Electronic filing systems	33	369,0	11,2	69	690,0	10,0	43	395,0	9,2	26	537,5	20,7	30	463,0	15,4	35	358,5	10,2	102	719,0		38	820,0	21,6
Electronic document managemen	4	269,0	67,3	11	653,0	59,4	2	234,5	117,3	5	358,0	71,6	0	143,5			5,0					0	0,0	

2001-2008			
A. Approval and Disposal			
Registries	553	3.440,2	6,2
Electronic filing systems	239	2.643,2	11,1
Electronic document managemen	198	6.455,7	32,6
Total	**990**	**12.539,0**	**16,6**
B. Submission preparation			
Registries	1.149	12.658,8	11,0
Electronic filing systems	440	5.724,4	13,0
Electronic document managemen	70	2.023,5	28,9
Total	**1.659**	**20.406,7**	**17,6**
C. Testing			
Registries	1.131	17.499,5	15,5
Electronic filing systems	376	4.352,0	11,6
Electronic document managemen	22	1.663,0	75,6
Total	**1.529**	**23.514,5**	**34,2**