

Memorandum

Title: Summary of Concept Model
To: New Preservation Concept
From: Asbjørn Skødt

Date: 16. december 2020
Case no.: 19/06296

Table of contents

1. Introduction.....	1
2. Concept Model.....	1
3. Screening and Priority List	3
4. P0 Pre-analysis.....	3
5. P1 Migration Assessment.....	3
6. P2 Format Assessment.....	4
7. P3 Testing	4
8. P4 Consequence Assessment	4
9. P5 Preservation Plan	5
10. Recommendation of Decision.....	5

1. Introduction

This document summarizes the content of the *Concept Model for Development of Preservation Plans* created under the project *New Preservation Concept* by the Danish National Archives. The purpose is to enable readers to understand and apply the methods, guides and documents of the Concept Model in actual assessments of data. Before this step, your archive has identified and assessed that these data cannot immediately be migrated to one or more of your archive's accepted preservation formats.

2. Concept Model

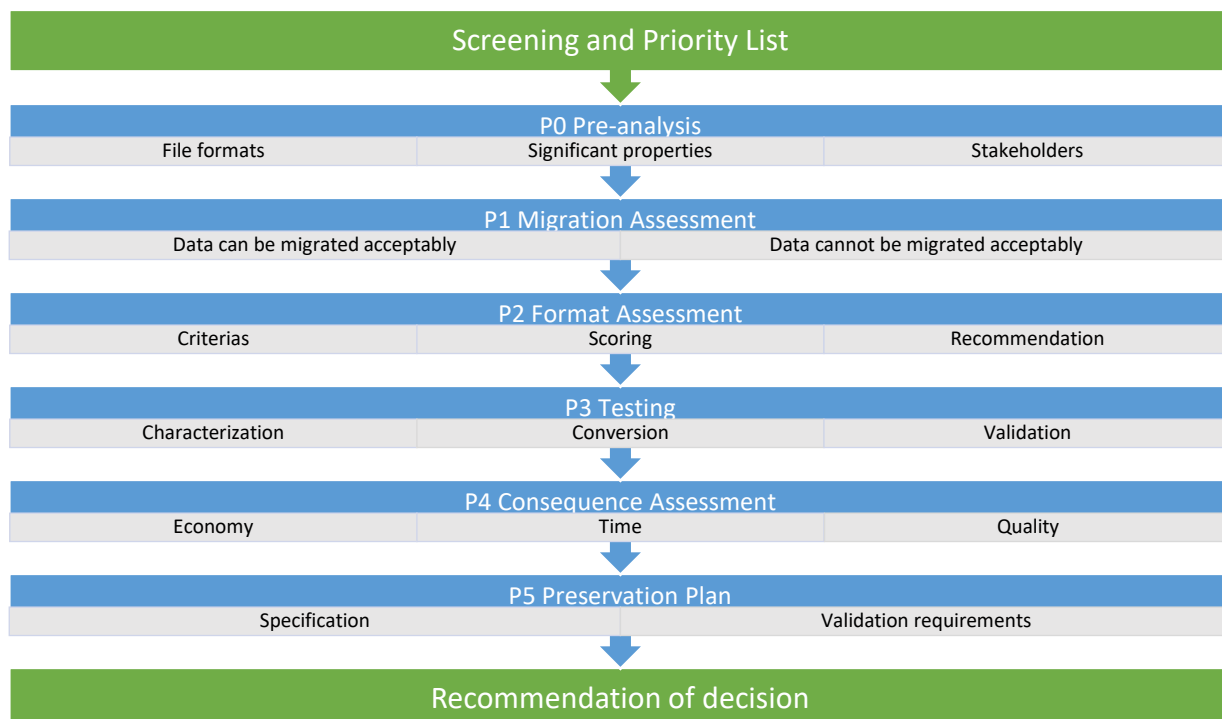
The Concept Model has gone through several iterations before its current version and it is under continuous discussion, evaluation and adaptation. The important thing to consider is that the model must be useful and generate new, relevant and insightful knowledge concerning a wide range of aspects related to problematic and unknown data, rather than you abide by the form.

Data selected for investigation by applying the Concept Model should be prioritized on a yearly meeting with specialists from relevant departments of your archive participating. The meeting considers both new data and the continuous monitoring of data in existing collections, as potential candidates for investigation. A Concept Model investigation is a time consuming task and it prerequisite that the relevant units of your archive can allocate time and resources for the investigation. Ideally, the prioritization of the list happen as an integral and coordinated part of your archive's annual planning.

The investigation begins with the carrying out of the Pre-analysis (P0), which then leads to a Migration Assessment (P1). This step is crucial, because it reexamines methodically whether the data can be converted to one of your archive's already accepted preservation formats. The Migration Assessment results in either an existing preservation format are selected for the Preservation Plan (P5) or the continuation of the investigation with Format Assessment (P2), Testing (P3), Consequence Assessment (P4) and ending with a recommendation of a new preservation format or level for Preservation Plan (P5).

Then a new process begins, which results in adopted legislation, new procedures for submission and revised validation software. However, this process lies outside the scope of the Concept Model. The investigation is formally concluded by the delivery of the reports, templates and a Decision Recommendation for your archive's management.

Figure 1. The steps of the Concept Model.



3. Screening and Priority List

The purpose of the Priority List is to collate the reporting of problematic data, which your archive become familiar with, typically through public notification of IT systems, submissions or supervision.

Before a new content type are added to the list, a screening must be performed. The purpose is to assess if any existing preservation plans (migration strategies) are suitable or a full Concept Model investigation is necessary to determine the appropriate preservation plan. If the latter, then the content type are registered on the Priority List.

The Priority List should be prioritized yearly in a meeting with specialists from relevant organizational units. The meeting results in the selection of one or more content types for Concept Model investigation. The list is an evolving document, which collates any preservation challenges your archive is facing and enables discussion. The list may also be the document where you collate information on file formats in danger of technological obsolescence and hence support the general monitoring of file formats in your archive's collections.

4. P0 Pre-analysis

Pre-analysis is fundamentally a step of collating knowledge concerning the content type's data e.g. describing how data are typically used, for what purposes, in which formats, which experience from previous and current submissions does your archive have, which stakeholders may be relevant to interview and what actualize the investigation.

Pre-analysis is delivered as a report which may contain a wide range of information concerning the content type and it may be relevant to later assess if entire sections should be moved to reports from later steps in the Concept Model. However, Pre-analysis should not speculate regarding the content in later steps and it should not provide any opinion concerning any potential results in later steps. Instead, Pre-analysis should enable later steps e.g. the identification of relevant stakeholders should be applied in P1 Migration Assessment and the identification of file formats should be applied in P2 Format Assessment.

5. P1 Migration Assessment

The second step applies the information obtained in Pre-analysis. The Migration Assessment must fundamentally answer the question, whether the investigated content type can be migrated to one of your archive's approved preservation formats without a significant loss of data quality and the potential for future reuse i.e. an unacceptable loss of significant properties.

Before any assessment may be completed, an object analysis of the content type or file format's technical properties and an analysis of the use cases of stakeholders are necessary. Finally, the technical properties and use cases are aligned for a confirmation of which properties are significant and must be preserved. The list of technical properties must be sufficient but not necessarily exhaustive.

The use cases of stakeholders are mapped by conducting interviews with an array of data producers, users and/or consultants. The interviews applies targeted questionnaires. The technical properties for the object analysis may be mapped by collating technical properties tables from the internet, testing the data of the content type or reading standards. The Migration Assessment method applies an adapted version of the InSPECT framework developed by Gareth Knight and the British National Archives in 2009.

6. P2 Format Assessment

The Format Assessment step compares a relevant selection of file formats. Relevancy are described in either the Pre-analysis or the Format Assessment report. Relevancy exists because file formats may be unsuitable for comparison despite belonging to the same content type. This may be due to technological obsolescence or obscurity. Instead focus should be given to those file formats, which are assessed to be immediately most suitable and from this comparison draw conclusions on which file formats to recommend as new preservation formats.

Format Assessment are conducted by applying a matrix, which consists of a large number of criteria related to suitability (or robustness), which may be checked off. The criteria are weighted differently depending on the criteria's importance in suitability. Some of the criteria are of crucial importance and if a criterion is difficult to apply to a file format, the criterion may be flagged. The criteria are automatically added for a sum score, which enables your archive to draw conclusions from the comparison. The Format Assessment are then collated in a report, which must describe each criteria for each file format, any potential red flags and a recommendation of which file formats are suitable for long term preservation.

7. P3 Testing

The step has the purpose of testing software for identification, characterization, conversion and validation of the file formats (typically 1-2 file formats) deemed most suitable as new preservation formats from the Format Assessment. The practical experience with testing file formats provides knowledge concerning the immediate applicability of any suitable file formats by means of tools whether they are open-source, commercial or developed in-house.

This knowledge must be collated in a report, where the test setup and results are described. The results may lead to a reassessment of relevant criteria in the Format Assessment, and the results are applied in the next step, Consequence Assessment, and the final step, Preservation Plan, where any tools and validation criteria are specified.

8. P4 Consequence Assessment

The Consequence Assessment measures the economic, temporal and qualitative consequences of accepting new preservation formats.

The economic consequences are e.g. costs for storage, monitoring, adopting legislation and developing new validation software. The temporal consequences may be experienced in manual workflows versus automatic workflows. The qualitative consequences are the potential of improving the data quality for users of archival data.

The three types of consequences are measured in a five-value scale ranging from “greatly negative consequence” to “greatly positive consequence”. The assessment applies presupposed economic values e.g. hourly rates at your archive or consultants, price of media types and applies estimates on expected growth of data and yearly submissions.

9. P5 Preservation Plan

Determining the preservation plan is the last methodical step in the Concept Model. The preservation plan must provide the recommended guidelines for preservation of the content type based on the assessments performed in the previous steps. The plan must ideally go into detail and specify all the necessary aspects of the submission and validation of data, so that your archive receives and preserves data in a controlled and documented manner.

The preservation plan may later be reformulated according to necessity. For instance, this could stem from further work on drafting legislation for adopting the preservation plan, development of tools for validation and later the general monitoring of the content type.

10. Recommendation of Decision

When the investigation is done, a recommendation of decision must be delivered to management. The recommendation summarizes how the content type should be preserved, which implications it has and if any communication is necessary. The communication should describe what should be communicated, for whom and through which channels. The recommendation should include appendices summarizing the conclusions of the different steps and links to the documentation. Use any relevant management template in your organization to write and deliver your recommendation.