

Memorandum

Title: Guide to P1 Migration Assessment
To: New Preservation Concept
From: Asbjørn Skødt

Date: 18. november 2020
Case no.: 19/06296

1. Introduction

This step is the most important step in the Concept Model for Development of Preservation Plans. The step is important because it is the results of this assessment, which determines if the investigated content type can be migrated to one or more of the existing preservation formats with an acceptable loss of quality measured in loss of significant properties.

The step P1 Migration Assessment leads to one of the following recommendations:

1. **Continue to P2 Format Assessment** if the content type cannot migrate with an acceptable loss of data quality to one or more of your archive's existing preservation formats.
2. **Continue to P5 Preservation Plan** if the content type can migrate with an acceptable loss of data quality to one or more of your archive's preservation formats.

Migration is in this context defined as the conversion of data between file formats.

2. Method

The migration assessment applies the InSPECT (Investigating Significant Properties of Electronic Content) framework for identification of significant properties. Gareth Knight developed the framework in 2009 through JISC funds and it was hosted as a cooperation between The Centre for e-Research at King's College, London, and the British National Archives.

In summary, the InSPECT framework is a method to assessing the significant properties of file formats' structure and usage. The framework enables archives to make documented preservation plans and actions supporting the technical mapping and validation of archival content and thus reinforcing authenticity of archived data. In the context of InSPECT, authenticity is considered a relative term, which defines the level of preservation through an archive's ability to preserve significant properties when migrating data.

Knight uses the following definition of significant properties:

”The characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects and their capacity to be accepted as evidence of what they purport to record.”

The framework utilizes two other methods from the field of digital preservation and engineering. The first one is the recognized and highly adopted *OAIS reference model*, which defines a group of stakeholders as a “designated community” for your archive. Thus, stakeholders become established entities which require a suitable plan for managing. The other method is *Function-Behaviour-Structure (FBS)*, which is a process-supporting design method for engineers delivering a method for cooperation with users (a particular stakeholder) concerning their needs.

Reference to memorandum: ”InSPECT framework til vurdering af signifikante egenskaber v1.0” case no. 19/06296-11.

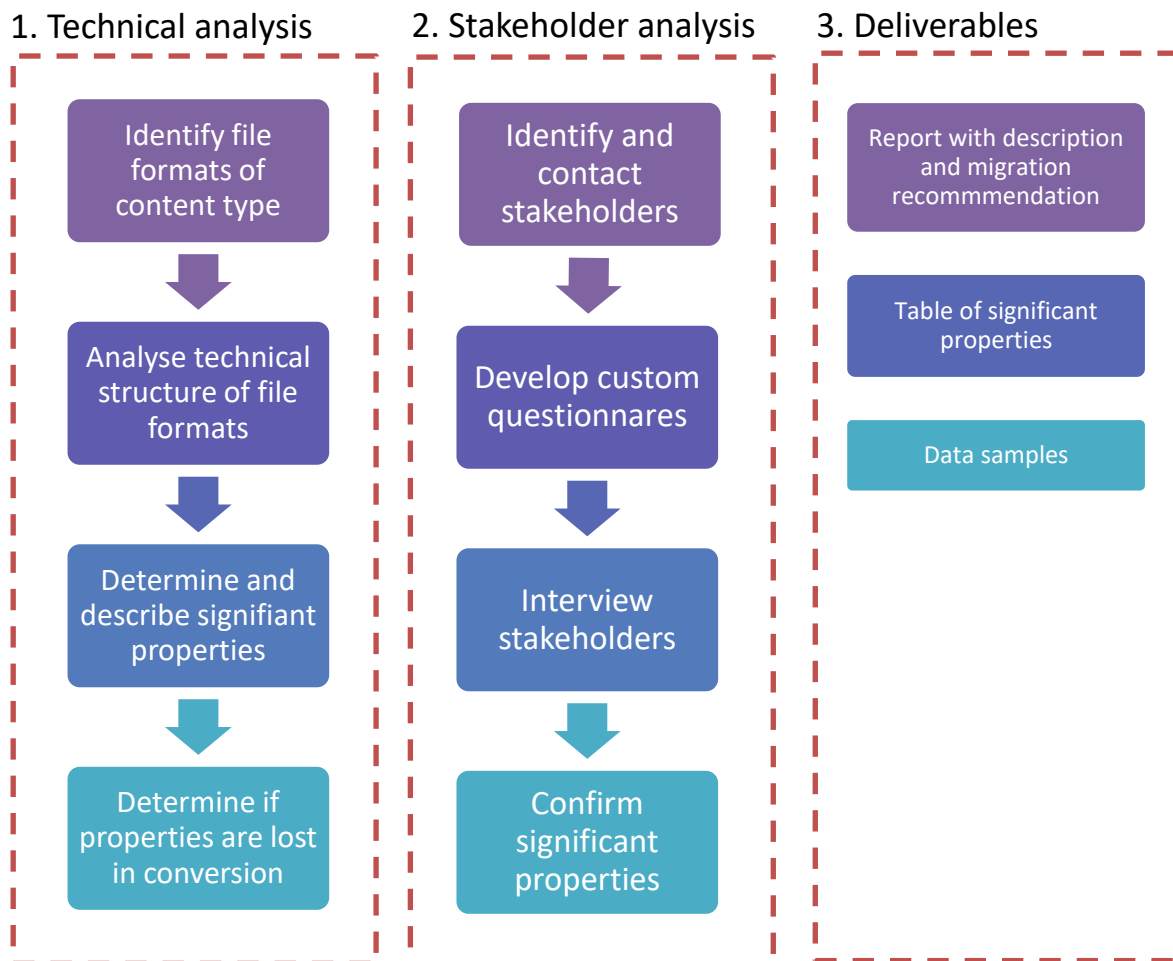
2.1. Methodological deviations

The step P1 Migration Assessment does not apply the InSPECT framework slavishly because we experienced certain aspects, which are not unequivocally expedient. This applies to the following:

- We do not require all file format properties are identified and described, because it can be too comprehensive with complex content types. Instead, we propose to concentrate the assessment on the properties, which we have a presumption are significant for the content type. We seek stakeholder confirmation of these properties. However, it is still vital that all significant properties are mapped, thus we do not require a mapping of all properties but recommend the mapping.
- The stakeholder analysis includes a step concerning the determination of stakeholders’ acceptable value boundaries e.g. min-max values for image resolution or bit depth. This determination can be used in validation requirements. However, it can be difficult for stakeholders to define these values, thus we do not require but recommend that stakeholders define acceptable value boundaries.

3. Execution

The following sections describe how to apply the Migration Assessment and which deliverables are expected. See illustrative overview below.



3.1. Technical analysis of significant properties

The identified file formats from the step P0 Pre-analysis are the objects of the technical analysis. The technical analysis of objects may adopt and further develop the preliminary table of significant properties identified in the Pre-analysis.

The technical analysis may involve the collating of specifications, test with characterization tools (may also be applied in the step P3 Test), practical use of data samples, internet research and sparring with colleagues. It might be necessary to procure or produce data samples, possibly in cooperation with stakeholders.

The identified significant properties must be listed in a table in the spreadsheet "Table_Significant properties". The table have the following columns, which must be filled:

- **Category:** Select a predetermined category suitable for the significant property. Options are "Content", "Context", "Rendering", "Structure" and "Behaviour"¹.
- **Subcategory:** Free choice of sub category for the significant property e.g. "Filtering and sorting"
- **ID:** Give an ID e.g. 1, 2, 3, etc.
- **Name:** Give a descriptive name for the significant property
- **Description:** Describe the significant property
- **Lost in conversion:** Is the significant property lost in conversion to one or more of your archive's existing preservation formats? Either yes or no
- **Stakeholder confirmed:** Have one or more stakeholders confirmed the property as significant? Either yes or no
- **Stakeholder comments:** Possibly, any comments which the stakeholder have concerning the significant property e.g. determination of acceptable value boundaries (validation requirements)

The spreadsheet also contains the following sheets:

- **Data analysis** applies quantification and overview of how many significant properties are lost in conversion and how many are confirmed by stakeholders. It uses a pivot chart. Right click on a pivot chart and click "Refresh" to put the values into the pivot charts.
- **Data formats** is a table for mapping of significant properties against file formats identified in the step P0 Pre-analysis and selected for assessment in the step P2 Format Assessment. You register "Yes/No" for match between file format and significant property. This table creates an overview of which file formats preserve which significant properties and this information is applied in P2 Format Assessment.

3.2. Stakeholder analysis

The purpose of the stakeholder analysis is to gather information concerning stakeholders' production and use of the content type. It is not a purely academic desk exercise for your archive's employees to imagine how a future user want to reuse data in e.g. 50 years from now. Therefore, a good introduction

¹ **Content:** Information content within the Information Object. For example, text, still and moving images, audio, and other intellectual productions. Examples: duration, character count.

Context: Any information that describes the environment in which the Content was created or that affect its intended meaning. Examples: Creator name, date of creation.

Rendering: Any information that contributes to the re-creation of the performance. For example, font type, colour and size, bit depth.

Structure: Information that describes the extrinsic or intrinsic relationship between two or more types of content, as required to reconstruct the performance. E.g. e-mail attachments.

Behaviour: Properties that indicate the method in which content interacts with other stimuli. For example, hyperlinks.

in a stakeholder interview is to tell the person, that the interview conveys an understanding for how the person would use data, if they imagined using the data in 50 years from now.

3.2.1. Identification of stakeholders

The stakeholder analysis typically involves 3-5 stakeholders, but no lower or upper limit are given. It is recommended to interview at least one municipality and one state agency. Stakeholders are typically data producers, other archives, users and consultants.

Identification of possible interview persons are based off the identification in the previous step P0 Pre-analysis, but more persons can be identified and contacted for an interview. You conduct an interview with the persons, that the contacted stakeholder deems relevant, but it is our experience, that the fewer participants in an interview the better.

The Migration Assessment report must include a table of which stakeholders (and persons) you conducted an interview with, which were contacted without reply and which denied interview as well as their contact information

3.2.2. Developing questionnaires

One or more questionnaires must be developed for the interviews. It is recommended to send the questionnaire to the persons before the interview so they have a chance to prepare for the interview and deliver better and more precise answers. The questionnaire must be adjusted to the individual situation.

The questions are to be formulated pedagogically yet they must be formulated in a way so that your archive receive information of which file formats are used (and for what purposes: Data creation, data exchange and/or (re)use of data) and which properties are significant and which value boundaries are acceptable if the data are reused in the future.

In conclusion, the person can be asked to supply representative data samples. This is especially relevant if the stakeholder is a data producer.

3.2.3. Conducting interviews

Interviews can be conducted through personal attendance at the location of the interviewed person or a video meeting. Interviews are recommended to last 1-2 hours.

Every meeting must have minutes written and it may ease the subsequent mapping of significant properties and usage, if this is done seamlessly during the interview when a property is confirmed. It is recommended that at least two archival employees participate in the interview to distribute the workload of writing minutes, mapping properties asking questions and maintaining the conversation.

3.3. Mapping of significant properties and stakeholder usage

The mapping of significant properties and stakeholder usage with each other either confirms or disconfirms our presumption whether the property is significant. This is done through connecting the information from the previous two analyses and hereby we achieve a confirmation in reality of what is significant rather than having to guess.

The answers from the interview must be analysed and identified significant properties must be mapped to the table of significant properties through stating “Yes/No” in the column “Stakeholder confirmed”. In addition, potentially newly identified significant properties from the interviews can be listed in the table.

4. Conclusion

The Migration Assessment must provide a recommendation to management on whether one or more of your archive’s existing preservation formats preserve the significant properties of the content type to an acceptable degree. If not, the step P2 Format Assessment must be initiated to assess the approval of one or more new preservation formats. If valid, the step P5 Preservation Plan is instead initiated and the relevant existing preservation formats are specified.

The Migration Assessment has the following deliverables:

- A report describing the process and results of the assessment. The report must contain a table of identified stakeholders. The report must also contain a table of identified file formats for application in the step P2 Format Assessment or a recommendation of which existing preservation formats are applicable for specification in the step P5 Preservation Plan. The table with file formats must contain information of whether the file format is used for data creation, data exchange and/or the (re)use of data.
- A spreadsheet with a table listing the significant properties, whether the property is lost in conversion to one or more of your archive’s preservation formats and whether the property is confirmed by stakeholders. In addition, another table listing which identified file formats contain the individual significant properties.
- Data samples provided by stakeholders. Data samples must preferably represent the stakeholder’s range of data of the content type.