

Memorandum

Title: Guide for P0 Pre-analysis
To: New Preservation Concept
From: Asbjørn Skødt

Date: 15. oktober 2020
Case no.: 19/06296

1. Introduction

This step deals with the collating of existing and new knowledge concerning the content type under investigation. A content type is a group of file formats used for the same purposes. The deliverable of P0 Pre-analysis is a report.

The collated information are applied in later steps e.g. P1 Migration Assessment and P2 Format Assessment.

2. Obligatory elements

The deliverable report of the pre-analysis must contain the following content and may contain other collated information deemed relevant.

2.1. Describe the content type

The data of the content type must be described in-depth e.g. what are the data used for, what purposes are fulfilled, what prerequisites are basis for the production and use of the data and which perspectives does the use of data have?

The description must contain a delimitation (or clarification) of the content type. This answer the questions of precisely what data are investigated and which other data are bordering.

2.2. Analyse prevalence of the content type

This section analyses the quantities of the data amongst data producers and already submitted to your archival collections. Possible questions:

- How much data are produced and/or used by data producers?
- How many and which submissions including the content type have your archive registered?
- Which file formats are typically used?
- How often does the file format extensions appear in your archived data?
- How much archival storage does the content type consume?
- How often is the content type disseminated to your users?

Information can be collated through dialogue with specialist colleagues and access to archival databases.

2.3. Analyse properties of the content type

The pre-analysis must identify file format properties of the content type. The identification can compose a preliminary draft of a table for the most known properties of the content type' file formats. This table is to be later refined in the next step P1 Migration Assessment through interviews with stakeholders and assessment of property significance.

Properties may be found through searches on the internet or through use of characterization or application software.

2.4. Analyse relevant stakeholders

The pre-analysis must identify relevant stakeholders producing or using the content type. In the next step P1 Migration Assessment, the stakeholders are contacted for an interview. The identified stakeholders must be listed in a table in the pre-analysis report with the following minimum information: Name of stakeholder organization and potentially contact person and type of stakeholder e.g. data producer, archival user, consultant etc.

Stakeholders can both be external as well as internal. However, the dialogue with colleagues are typically preparation and improvement of the questionnaires for the external interviews.

2.5. Analyse relevant file formats

The pre-analysis must identify all file formats of the content type and select relevant file formats for the step P2 Format Assessment. The file formats must be listed in a table in the pre-analysis report with the minimum information: Name of file format, extension and whether the file format are included in P2 Format Assessment.

The selection of file formats must be justified. It may be meaningful to group the rejected file formats based on certain conditions e.g. if they are technologically obsolete or obscure (rarely used or used in highly specialized cases), so that the justification applies for a group of file formats rather than individual. However, relevant file formats for P2 Format Assessment must be individually justified. Relevant file formats can be judged based on whether other comparative archives have approved the file format for submission (This information must be collated. See spreadsheet "Collation of File Formats in Archival Use"), the file formats are independent and contains as many significant properties as possible.

The purpose of the selection process is to make P2 Format Assessment as effective as possible by only spending resources on assessing the most suitable candidates.

3. Conclusion

The pre-analysis produces a report with the following obligatory elements:

1. Description of the content type
 - a. Analysis of the prevalence of the content type
2. Analysis of properties of the content type
3. Analysis of relevant stakeholders
4. Analysis of relevant file formats

The pre-analysis may contain other relevant descriptions and analysis, which the analyst encounters during the process.