RIGSARKIVET

Memorandum

Title: Screening of Content Type   Date: 2. marts 2021
To: New Preservation Concept   Case no.: 19/06296
From: Asbjørn Skødt

Table of contents

1. Introduction

Screening of a content type serves the purpose of determining what should happen when your archive, through various channels, come across problematic data, where it is presumed the data cannot be immediately converted to one or more of your preservation formats. Screening involves the following items 2-4.

2. Investigate content type

The scope of the investigation should be determined with relevant experts in your archive e.g. it should clarified if the investigation concerns a content type or only a specific file format.

Investigation involves:

- Conduct meeting with relevant experts concerning the problem

- Contact and conduct a meeting with relevant experts. Clarify possible issues regarding the problem and inform them of your further process for the investigation.
- Possibly, interview one or more data producers concerning their use case.
  - Contact and conduct meeting with a data producer concerning their use cases for creation, usage, submission and reuse of the content type. Apply the questionnaire in Appendix 1 or develop your own questionnaire in advance of the meeting.
- Delimit the content type
  - Identify and delimit the content type based on the data producer's use case. Determine whether the content type is already in the Priority List.
- Briefly describe the significant properties and file formats of the content type
  - Describe the file formats typically used within the content type and which significant properties and corresponding use cases, that these file formats support.

3. Recommend preservation solution

Based on previous steps, an assessment of whether data immediately can be migrated with acceptable loss and submitted in one or more of your archive's preservation file formats must be made.

- The assessment must address whether data can immediately be migrated to a suitable preservation format by application of an existing migration tool.

The screening assessment must result in one of the following recommendations:

A. It is our assessment that data in the content type or specific format can always be submitted in existing preservation formats. If relevant, notify the data producer.
B. It is our assessment that data in the content type or specific format can always be submitted in existing preservation formats, but we also assess that the data are migrated with loss to an extent, which should be investigated further. Register the content type or specific format on the Priority List and if relevant notify the data producer.
C. It is our assessment that data in the content type or specific format cannot be submitted in existing preservation formats with acceptable loss. The content type or specific format must be registered on the Priority List.

3.1. Preservation level

The recommendation of preservation solution must be followed by a recommendation of preservation level. The recommendation must be given on the basis of a consideration of what is best to do under the current circumstances and available resources. The recommendation is a specific instruction for how to manage the data. The recommendation should be negotiated with relevant experts and managers in your organization.

One of the following preservation levels is suggested:

1. Keep at data producer or other institution (postpone submission or use proxy archive)
    o   Negotiate agreement
2. Submit original data
    o   Recommend a minimal solution where only copies of original data are submitted
3. Migrate to existing or temporary preservation format
    o   Instruct preservation solution compliant with legislation

3.2. Preservation plan

A preservation plan must be produced. Describe the content type, preservation level, its file formats and how the content type should be managed and register information for the specific file formats and how to preserve these. See Appendix 3.

If recommendation C was reached, then the preservation plan may be reformulated as a result of conducting the full Concept Model investigation.

4. Collate information in screening report

The information gathered during the screening items 2-3 must be collated in a screening report, which are approved by management and sent to relevant caseworkers handling the submission.

Proposed disposition for screening report:
1. Summary
    a. *Briefly describe the background, challenge and conclusion*
2. Screening of [insert name of content type]
    a. *See item 2 for content*
3. Recommendation of preservation solution
    a. *See item 3 for content*
4. (Possibly) Appendix: Preservation plan (if preservation level 1-3)
    a. *Describe the preservation of the content type*
    b. *Insert filled templates for file formats*
5. (Possibly) Appendix: Priority criterias (if preservation level 4)
    a. *Insert filled template from Appendix 2*

Case filing of the screening report should happen separately under the name "Screening of [insert name of content type]". Here also relevant information from mails and the data producer's notification should be registered.

Preservation plans for individual file formats must be copied into "Table_Preservation plans for data formats".

Appendix 1: Example of questionnaire for screening interview

Questions to data producer when screening content type and recommending preservation level.

1. Tell us about yourself and your organization
2. Why do you produce data of this content type? What is the purpose?
3. How many files of the content type do you have, and how large is their total size in TB?
4. How large a percentage of your total data, does the content type fill? Both quantity and TB.
5. Which file formats do you store the content type in?
6. How important is it for future reuse, that data are preserved in their original formats?
7. Which properties are important to preserve?
   a. Which properties for capturing the content of data are important?
   b. Which properties for rendering data are important?
   c. Which properties for the context of data (metadata) are important?
   d. Which properties for the structure of data (e.g. linked files and folders) are important?
   e. Which properties for the linking of relationships (e.g. external files and information) are important?
8. Do you think, the content type can be immediately converted to one or more of our preservation formats?
9. Which file formats would you recommend to preserve the content type in, so they can be reused in e.g. 50 years from now?

Appendix 2: Template for registering content type in Priority List

Assessment of priority criterias must include a reason given to the clarifying questions listed in the table below. Conduct meeting with relevant colleagues to determine assessment.

The information must then be registered in the Priority List. The assessment contributes to forming a uniform decision basis for prioritizing the investment of analytical resources in developing new preservation plans for content types.

Table 1: Template for assessing priority criterias

Criteria	Clarifying question	Answer	Reason
Urgent	Must a decision concerning a new preservation plan be made rapidly? It may be urgent if data should be submitted to your archive before administrative deletion or	Yes (2)/No (0)	

	submission processing time limitations. If urgent, it might be relevant selecting a lower preservation level to speed the process up		
Relevance	Is data preservation-worthy?	Yes (2)/No (0)	
Prevalence	Is the content type's data prevalent among data producers? If the content type is not prevalent, it might be relevant selecting a lower preservation level to speed the process up	Greatly (2)/To some degree (1)/Limited (0)	
Quantity	Is the total data quantity, which your archive expect to receive over a 5 year period relatively large? (e.g. >10GB)	Greatly (2)/To some degree (1)/Limited (0)	
Reuseability	Would data submissions improve future reuse?	Greatly (2)/To some degree (1)/Limited (0)	
Focus	Do data have a heightened societal and political focus?	Greatly (2)/To some degree (1)/Limited (0)	
Maturity	Are the data formats established and relatively mature for preservation	Greatly (2)/To some degree (1)/Limited (0)	
Comments	Describe any potential information related to the data.	Free text	

Appendix 3: Template for preservation plan

Fill the template separately for every file format, which the report has screened.

Table 2: Preservation plan template for data format

Start date	State start date for preservation plan
End date	State end date for preservation plan
Content type	State the data area, which the format relates to e.g. "spreadsheet"
Full name	The entire name of the format (without year and version number)
Other names	Other ways of writing the name, both formal and informal
Version	The version number of the standard this preservation plan covers
File extensions	The file format extensions
Description	Brief description of the characteristics of the file format
Standard	The name of the standard this preservaton plan covers
Publication year	Year of publication for the standard
Publisher	Name of the publisher(s) of the standard
License	How is the file format regulated? E.g. "Open", "Proprietary" etc.
Other standard	If another standard for the file format is also valid
Pronom PUID	ID in the technical registry PRONOM
MIME type	What is the file format's MIME type?
Signature	What is the signature (magical number) of the format?
Format assessed	Has the file format been assessed by use of the concept model? Yes/No
Preservation level	What preservation level is chosen? Maximal/Middle/Minimal/None
Convert to	Which file formats must the file format be converted to before submission?
Convert from	Which file formats are the file format converted to before submission?
Validation req.	Which significant proper are required for validation of migration?
Validation tool	Which tool(s) should be used for validation?
Risks	Which risks exist in the long term preservation of the file format?
Monitoring	How should the file format be monitored? Ordinary/Yarly/Intensive/None
Comments	Possible other comments e.g. next steps