Memorandum

Title: Guide to P5 Preservation Plan

To: New Preservation Concept

From: Asbjørn Skødt

Date: 28. august 2020

Case no.: 19/06296

Table of contents

1. Introduction

The last step in the concept model pertains to the drafting of a preservation plan for the investigated content type and specific preservation plans for the file/data formats of the content type.

The purpose of preservation plans is to create a factual and documented basis for your archive's long-term preservation of digitally created data. Preservation plans document metadata, identificators, risks, validation requirements, preservation actions/solutions such as migration paths, and collate the methical assessments from the concept model's previous steps. Likewise, preservation plans registers choice of preservation level and level of monitoring.

2. What is a preservation plan?

A preservation plan transforms the policies and strategies, which management have created for your archive's operation, into guidance for real preservation actions. A preservation plan is the lowest and most detailed level for preservation.

The preservation plan primarily entails to how an AIP should be created and maintained over time, but it may also describe conditions relating to SIP and DIP. These affect and are vice versa affected by the AIP. The preservation plan must describe which data it delimits, the migration paths for data, storage, monitoring and the detailed data management.

3. The Danish National Archives' preservation plans for digitally born data

The DNA's preservation plans for digitally created data governs both content types and file formats – both not collections and archival fonds. The point of the preservation plans is to enable the DNA to assess and document how the preservation of specific data are to be tackled and under what conditions. It is crucial to remark, that these preservation plans does not include any kind of assessment of preservation-worth (whether to preserve or to destroy data), but they document and collate how we preserve the data.

Regarding storage, the DNA have the policy that digitally created data must be stored on multiple media types, multiple locations, multiple organisations, and that control of data integrity (checksums) must be performed yearly, which is why it does not make sense to specify these conditions in the preservation plans for specific content types or file formats.

3.1. Content type

Preservation plans for content types must be delivered in a report with free choice of disposition. The report must collate the assessments and tests applied in the previous steps of the concept model, and it must conclude on these as recommendations for the long term preservation of the content type's data. The content type preservation plan must be delivered together with specific preservation plans for the investigated file formats, and these must be delivered by applying a template (*see 8.1. Template*).

Preservation plans for content types are the last step of the concept model, and they will typically only be reformulated, if the content types is reinvestigated by use of the concept model. On the other hand, preservation plans for specific file formats may be reformulated when necessary (*see 7. Reformulation and monitoring*).

Figure 1. Relationship between file formats, content types and preservation plans in the concept model.
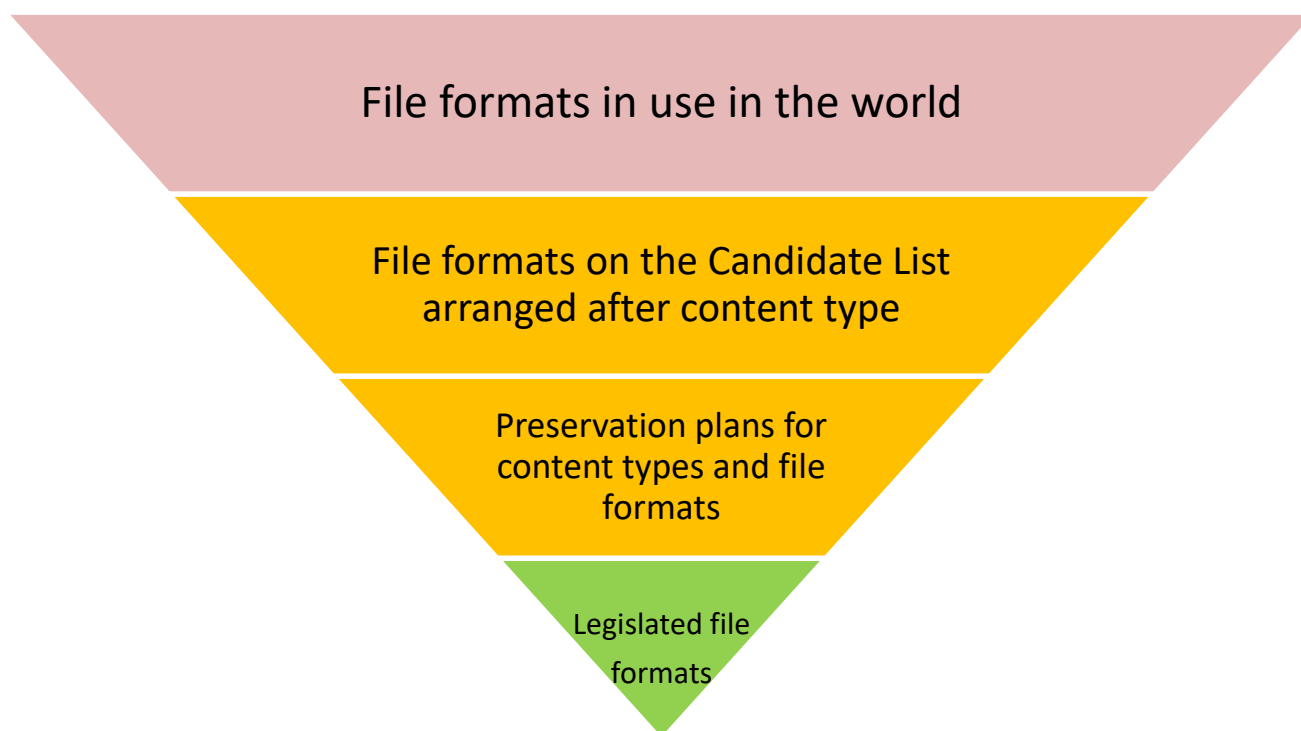
3.2. Data formats

Preservation plans for data/file formats are suggested collated in the spreadsheet "Table_Preservation Plans for Data Formats". The spreadsheet has three different sheets for registration of the preservation plans. These are "1. Tentative preservation plans" containing the plans produced by application of the concept model methods, "2. Active preservation plans" containing the legislated, and hence active, plans and "3. Deprecated preservation plans" containing previously legislated plans or plans for data, which are most likely not in usage by data producers anymore.

4. Preservation levels

The determination of preservation level for data is an important choice. The preservation level determines the level of preservation safety (i.e. data are migrated and continued accessibility are secured), which the DNA offers for these data. Previously, the DNA offered fundamentally the same level of preservation through application of a normalization paradigm, which has resulted in uniform and highly manageable data collections. Different preservation levels are, however, a new concept, which does not fully break with the normalization paradigm but offers instead the possibility of differentiated preservation of data, whereby the DNA does not necessarily offer migration or supported dissemination of data.

Maintenance is the promise, which the DNA gives to data producers, that we at any given time are able to disseminate the data in contemporary software renderers. We are able to give this promise, because we have IT systems and methods for the maintenance of data, so that data migrations happen if they are in risk of technological obsolescence or if new preservation goals for dissemination require it. Alternatively, the fulfillment of preservation goals may also happen through investment in other IT systems for dissemination.

Examples of preservation levels (under discussion):

1. <u>Optimal</u> = approved as a preservation format and maintained.
2. <u>Normal</u> = submissions of original file format and migrated file format. The latter are maintained.
3. <u>Observation</u> = only submission of original file format and maintenance.
4. <u>Minimal</u> = only submission of original file format and no maintenance.
5. <u>Deponi</u> = agreement with another organisation to preserve the data on our behalf.

5. Validation

We propose the two fields "Validation req." and "Validation tools". The requirements for validation contain the full specifications of which significant properties must be validated. Validation may include compression type, bit depth, macros and metadata. Validation of extensions must always be performed and is therefore not specified in the preservation plans, however if validation of magical signature must be applied, then this must be specified, because the magical signature is not always known.

Validation tools are the software applied for the validation e.g. Jpylyzer and/or manual validation methods such as visual samples.

6. Risks

The registration of risks connected to file formats are a key element for the managing of data collections and affect a number of other template fields in the preservation plan. Risks may be connected to e.g. storage, tools, loss in quality, licenses, technological obsolescence, standardization, validation, dependencies to other formats and dissemination prospects.

In and by far the registration of risks may have their starting points in the conclusions from the format and consequence assessments, if they have been completed. If not, the risks are to be assessed from the knowledge, which the staff member have immediately available or been able to collate at the time of registration. If new risks are identified over time in connection with monitoring, these are to be added as reformulations of preservation plans.

7. Reformulation and monitoring

Reformulation of preservation plans are conducted as part of the monitoring of file formats. Preservation plans are, in this way, not set to last forever, but they are instead objects of constant reiterations motivated by monitoring and changed preservation goals.

Monitoring means the continuous work with keeping track of the risks of technological obsolescence for file formats or that newer file formats are created, that potentially better preservation formats.

Monitoring can be performed through various sources such as literature, newsletters, conferences, networks, communities, sparring, consultancy and software reviews. Monitoring are performed in different intervals, and we propose the following classifications:

1. Ordinary = Monitoring are persistent through the abovementioned sources
2. Yearly = The file format are on the Candidate List and are discussed once yearly
3. Intensive = The file format are on the Candidate List and is currently being investigated
4. None = The file format has the lowest preservation level and is not monitored

In case the reformulation of the fields "Preservation level", "Convert to" and "Validation req." are necessary, then the active preservation plan must be transferred to the sheet "Deprecated preservation plans" and a new preservation plan must be created, based on copying of relevant consistent template data. The sheet "Tentative preservation plans" are proposals for new preservation plans and does not require to be transferred to "Deprecated preservation plans" upon reformulation.

8. Deliverables

The step P5 Preservation Plan leads to the following deliverables:

- A report describing the content type preservation plan should describe the preservation goals and the selection of significant properties for preservation in a report, which includes the recommendation of preservation formats.
- Specific file format preservation plans – both recommended and not recommended – delivered in a template. This information are then collated in the spreadsheet "Table_Preservation Plans for Data Formats".

8.1. Template

Apply the document "Template_Preservation Plan for Data Format" to fill the preservation plan. After filling the information, the information must be copied to the spreadsheet "Table_Preservation Plans for Data Formats". The spreadsheet will from now on be the place where the preservation plan is updated. Therefore, the template only has a purpose in connection to the initial concept model investigation where a large amount of information is collated and registered.

The template contains the following fields:

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