# Preservation Planning, Beacons for a TDR

## *Three Cases on Archiving Emerging Media*

**Marjolein Steeman**
*Netherlands Institute for Sound and Vision*
*The Netherlands*
msteeman@beeldengeluid.nl

**Abstract – In order to preserve digital objects for the long term repositories need to choose a preservation strategy. For new emerging types of media this is a challenge. This paper describes how various cases occurred at the Netherlands Institute for Sound and Vision. It shows how preservation planning helps management in putting these matters in the right context and taking informed decisions based on knowing what we know now. It concludes with an overview of the content of a Preservation Plan, as has been implemented in practice.**

**Keywords – preservation planning, trusted digital repositories, new media, audiovisual heritage, AV archives**

**Conference Topics – Designing and Delivering Sustainable Digital Preservation; Exploring New Horizons.**

## I. Introduction

The oldest existing lighthouse of the Netherlands is called the Brandaris. It stands at the eastpoint of one of the isles in the North and dates from 1592. Many years it served as a beacon to guide ships from the far east, west and from the northeast to Amsterdam. And also to guide them on the way out.

This paper is called Preservation Planning, beacons for a TDR (trusted digital repository). It shares recent experiences on preservation planning at The Netherlands Institute for Sound and Vision (NISV). Especially with the emergent new media formats in our modern information society. It will present how preservation planning is put into practice in this institute and how it serves as a beacon that helps guiding the ingest of and access to media works in our repository.

## II. Preservable formats

Sound and Vision is an independent media institute that holds a heterogeneous collection including the public broadcast archives, education and science collections as well as amateur and independent works. The archive stores more than 1 million hours of digital AV material and also at least 20,000 objects and over 2.5 million photos. The institute is a museum, an archive and a knowledge institute. In 2016 the Data Seal of Approval was granted: a certificate for trustworthiness of repositories.

At Sound and Vision the complete archival storage contains 34 petabyte of files. These 34 petabyte are used by only a few different file types. Dpx and wav files (40% of the used capacity) are used to store our digitised film. A tiny part of the storage consist of tiff files, representing the photos. Wav-files (4% of the storage) are used for audio and mxf-files (55% of the storage) for video. Overall, Sound and Vision has only four preservable file formats in its repository.

Only content that is presented in or will be digitised to one of these preservable formats, qualifies for full preservation. Other formats are not accepted because the longevity can not be guaranteed. This is called a “just in case” policy.

The preservable formats have been described in detail in a Preservation Metadata Dictionary (PMD). This PMD is the first product of our preservation planning activities. It is used as a reference for new ingest: what technical metadata must be provided and what characteristics are allowed. Also: via a systematic mapping it records where the characteristics are documented in the repository systems. The PMD is conformant to level 1A of PREMIS and it


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

is recognised as essential information for NISV as a trusted repository [1]..

All principles and choices for execution of the business of sustainable digital preservation have been outlined in a policy document [2]. By documenting the current policy and the standards employed, it is possible to account to all parties that entrust their digital collections to Sound and Vision, and to offer the staff of Sound and Vision transparency and clarity on the rules and procedures that apply.

## III. Emerging media

But what if new media, new formats, new requirements come into play? To answer this question the following case is exemplary.

### *A. Webvideo*

In 2004 Vimeo, the first big webvideo platform arrived. Soon followed by YouTube in 2005. A few years later Sound and Vision did research on the options for archiving webvideo, followed by some internal projects and an exhibition in 2016. In 2018, the institute decided it was ready to store webvideo in its trusted archive, as the following terms had been met:

1. Our mission is comprehensive: "Sound and Vision wants to improve everyone's life in and through media by archiving, exploring and clarifying that media". Webvideo is definitely within scope.
2. We recognised the Internet is great for sharing, but it is not an archive: we sure must take on our role here.
3. We developed selection criteria for webvideo that should cover the new Dutch media landscape of webvideo.
4. Agreements were made with rightsholders of the videos on archival services and on publishing in specific context.
5. Tooling had been implemented to gather metadata from the web.
6. And last but not least a proposal was made for a new preservable format.

This was when preservation planning was allerted. To get a full understanding of this proposal, let us first give some context.

The current preservable format for video is an MXF-wrapper with D10-30 or D10-50 videocodec. D10 is an implemented MPEG-2 codec used in production workflows for digital television. It is an industry standard, well documented and widely supported. The MXF/D10 is transcoded to a proxy for viewing or dissemination.

But the codec uses a bitrate of 30 or even 50 Mbps. Where the webvideo comes in max. 2,5 Mbps. This means that transcoding all webvideo to MXF/D10 would inflate the size of the files. An unwelcome effect. Also: the MXF/D10 isn't lossless; it is lossy. Transcoding a lossy compressed file (webvideo) to another lossy codec is far from ideal for preservation.

Therefore webvideo team proposed the introduction of a new preservable format: an MP4-wrapper with an H264 codec. This seemed a plausible proposal. A lot of webvideo nowadays has exactly this format, so transcoding would then often not be needed.

However from a preservation point of view, one might question this option. The MP4/H264 might be widespread at the moment, but for how long? H265 with even better compression is coming up. Also: H264 defines the codec, but there are a range of other file-characteristics that may have implications for access or playout. And on top of that: again it is a lossy compression. Transcoding may have impact on the quality of the file, which is ofcourse undesired.

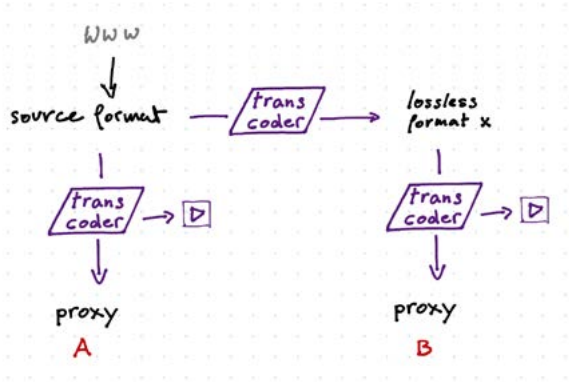


Figure 1 New scenarios A and B for webvideo.[1]

[1] all illustrations by M. Steeman/NISV licensed under CC BY-SA


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

So instead a scenario for a lossless format was made. A new preservable format that is trusted to stand the test of time. That is: to live long and, or migrate well. The suggestion (fig.1) is to archive the source format "as is", provided that our systems can create a proxy and support playout. If not, the source format is transcoded to a lossless format "x". From this a proxy can be created infinitely.

Once the source format has been accepted (Scenario A), there is no immediate need to transcode to a lossless format. The source format can be disseminated or the proxy itself as a standard derivation.

There will be a need to keep monitoring though. If the source format threatens to become obsolete then still a lossless archival master must be created (Scenario B). This will in fact depend on evolutions in the playout environments. As a starting point the internal transcoding software will act as a reference for the playout environment. It has been provisionally agreed to that when new versions of this software cease to support certain outdated formats, this calls for action.

Transcoding to a lossless format will probably also inflate the size of the file. But instead of inflating all files, this will only happen when it is relevant to do so. In other words: a "just in time" policy is applied instead of "just in case".

### *B. Getting our Bearing*

From just in case, to just in time. This is an essential addition to the NISV preservation policies. It opens up the archive for new media that so far were put aside on separate disks, where the risks of not being properly looked after are eminent. It also introduces a new operational practice, following scenario A or B. And it sheds light upon the issue of obsolescence. In particular how this risk must be monitored.

The scenarios were documented in a preservation plan, that was presented to the NISV preservation board. It was important to have their consent, before the consequences of the policy were worked out. Even more important: making this preservation plan, together with all internal stakeholders, indeed helped Sound and Vision to retrieve its bearing with respect to preservable formats. Preservation planning operated as a true beacon and put us back on track.

## VII. Preservation planning

The case ends with drawing up a "preservation plan" to underpin the new policy on preserving webvideo content. How does this relate to the latest standards on preservation planning?

### *A. Planets and OAIS*

Becker c.s. [3] make an important distinction between concrete preservation plans and high-level policies. It is claimed that a preservation plan is seen on a more specific and concrete level and Becker refers to the definition as was adopted by the Planets project: "A preservation plan defines a series of preservation actions to be taken by a responsible institution due to an identified risk for a given set of digital objects or records (called collection)" [4]. The preservation actions are specified, along with responsibilities and rules and conditions for execution on the collection.

The Planets preservation workflow consists of four phases:

1. Define requirements
2. Evaluate alternatives
3. Analyse results
4. Build preservation plan

In this view the preservation plan is right at the end of the process of working out all details. The definition speaks of 'preservation actions'. The preservation plan contains an executable workflow definition to perform a specific migration on a specific set of records or files. However, in the case of webvideo the preservation plan documented the recommendations to the board on how to approach this new preservation case. This implies a more generic plan, proposing new policy guidelines.

The Open Archival Information System (OAIS) [5] is a widely accepted reference model to become a so called Trusted Digital Repository. Preservation Planning is one of the entities of the OAIS functional model.

Preservation Planning is linked to the entity of Administration, that contains the services and

functions needed to control the operation of the other OAIS functional entities on a day-to-day basis. Administration functions include maintaining configuration management of system hardware and software. It is also responsible for establishing and maintaining Archive standards and policies.

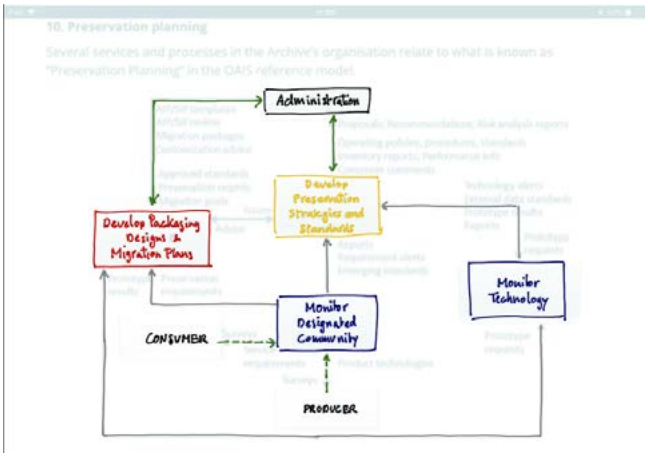


Figure 2 Functions within Preservation Planning according to OAIS.

Fig. 2 shows this relation between Preservation Planning and Administration more in depth, by unfolding Preservation Planning in the composite functions. Preservation Planning consists of four functions. The webvideo case seems to fit very well in one of these functions: developing preservation strategies and standards (yellow).

To develop packaging designs and migration plans (red) refers to more operational planning. This function is more in line with the concept of Planets. It delivers a detailed timetable of actions.

Both have a relationship with Administration but in a very different way. Developing strategies and standards relates to management that establish the policies and make decisions on scenarios or options. Where packaging designs and migration plans are input for System Configuration, the operational level of Administration.

In the workflow presented by Planets, the first three steps are said to be compliant with Develop Preservation Strategies and Standards. The outcome is provided to the Develop Packaging Designs and Migration Plans function as advice to create a detailed migration plan.

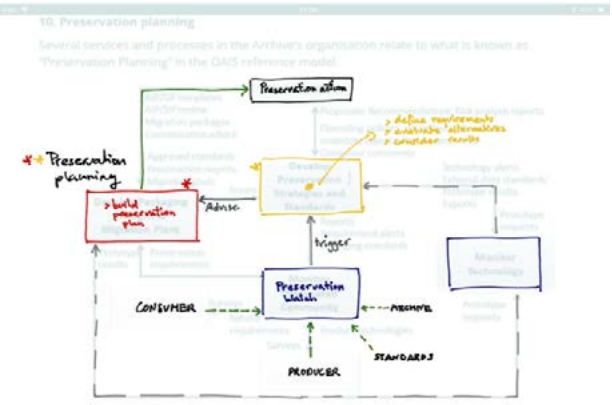


Figure 3 Planets workflow within OAIS.

It is evident that the Planets workflow is very straightforward. Preservation watch leads to testing and evaluating, resulting in an advise. A detailed plan is built and carried out by Preservation Action. Management is not involved explicitly. Policies seem already set and covered.

Given the experiences at Sound and Vision both planning functions are not necessarily part of the same workflow. The “Preservation Plan” that documented the additional policies on webvideo is the outcome of Develop Preservation Strategies and Standards. This plan is explicitly presented to Administration. A detailed action plan on a given set of digital objects would rather be referred to as “Migration plan”.

Using the metaphor of the beacon, “developing preservation strategies and standards” can very well be the lighthouse that guides the ships at the horizon. Where “packaging designs and migration plans” are like the mooring buoys that are placed to navigate between shallows or along the fairway at a particular location.

### *B. Triggering a Preservation Plan*

The two other functions of OAIS preservation planning, are monitoring functions (blue). First of the designated community (consumer, producer) and secondly of technology (file formats, standards, tooling etc). Both give input to the yellow Strategies and Standards and to the red Develop packaging designs and migration plans.

The difference between the two can be illustrated by an example. A topical issue right now is the fact


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

that production technology in broadcast environments is changing gears towards 4K. Makers create files in 4K. MXF/D10 might not be adequate for those producers. Our consumers might no longer be happy with an - in this respect - inferior standard.The following questions arise:

- what do we know about the production context?
- how widespread and fast is this change?
- will the broadcasters come up with a high resolution standard broadcast format?
- what formats will be conventional among makers?
- and who's deciding about new standards

These questions are addressed by the monitoring function of the designated community.

From a technology point of view the new emerging formats and codecs are studied by the OAIS-technology monitoring function. They ask questions like:

- open source? how is versioning done? what about backward compatibility?
- proprietary? are there licensing issues?
- how do new codecs perform in terms of transcoding speed?
- will our own infrastructure and tooling be able to adopt the new format?

With these two monitoring functions the repository builds up knowledge. The aim is that this knowledge is adequate to give a timely and substantiated advise on which preservable format to choose. The urgency of the issue in combination with the comprehensiveness of the knowledge, will trigger the preparation of a preservation plan to introduce this new format to management.

The two monitoring functions can trigger a preservation plan in several ways:

- Producer: new production technology, new collections
- Consumer: new requirements for playout
- Archive: new collection profile, priorities in budgets, outcomes of self assessment
- Standards: new opportunities or risk alert (obsoletion)

Monitoring implies an ongoing activity. The outcome is always temporary; based on current findings. But in terms of risk management the outcome must be assessed and sometimes calls for action. Then preservation planning must document the options and give advise, thus presenting a preservation plan. In some cases this will give rise to a specific migration, but certainly not necessarily.

## V. Requirements for preservation

Making the effort of drawing a preservation plan offers the opportunity to think through the preservation challenge as it emerges as exemplified in the webvideo case. This will be further illustrated by the following two other cases.

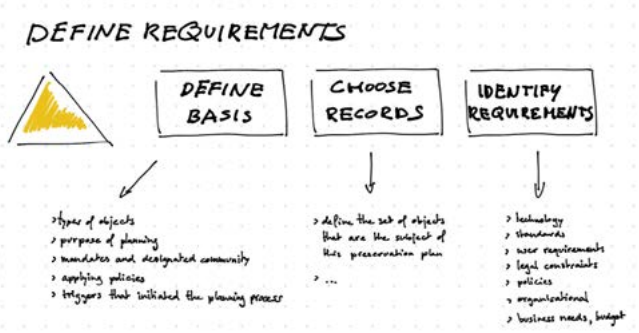


Figure 4 Steps within the first phase of the Planets workflow.

In terms of scope it will turn out that drawing a preservation plan has much similarity with very first phase of the Planets workflow, "define requirements". This will become apparent when the outline of the NISV preservation plan will be given, at the end of this paper. It is interesting to note that this phase is followed by the definition of alternatives and a Go/No-Go. Perhaps this is the parallel with presenting the plan to the NISV preservation board.

### *A. GIF - Graphical Interchange Format*

Recently it was decided that Sound and Vision wants to include GIF images to the collection. This triggered preparing a preservation plan.

First the technical aspects of the GIF-format were investigated. GIF was introduced in 1987 by CompuServe; it's history goes back to the start of the internet. It became popular because it used a very efficient compression technique. Many pictures could be downloaded rapidly, even on slow connections.


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

The extreme limitations of the GIF format and the restrictions of websites that display them played a vital role in the way GIFs were made, with makers tweaking the size and color palette as well as editing frame-by-frame to make the best-looking, smallest possible file. All that nuance can disappear if the archive is not careful to preserve both the GIF itself, and the context of its creation.

For instance: rendering a GIF on current browsers might not give the same result as the original. Some users (like exhibition curators) might even go so far as fully emulate old hardware to ensure that variables like CPU speeds or screen technology don't mess up the visual representation the artist intended.

To avoid this, the GIF may be transformed into a video file. But there is a significant risk that this may change the way the GIF appears, caused by misinterpretation of instructional metadata, or by the introduction of color shifts or even potentially compression artifacts through the process of encoding as video.

One final point of consideration when rendering GIFs from the early web: it is often the case that these GIFs play back at a faster rate today, as they were limited by the slow CPUs of the time of their creation. Employing emulation to view historic Animated GIFs in something close to a period specific CPU, operating system, and web browser is therefore often recommended.

This short introduction illustrates there are at least two options on preserving GIFs in a repository. First: one could add GIF as a preservable format. This would imply that the minimum set of metadata for GIF would be documented in the Preservation Metadata Dictionary (PMD) , together with a mapping to the NISV systems and table columns, where this metadata will be stored. There would be some research needed to define what technical metadata can assure that all the specifications to render the GIF properly, are covered. This will include some specifications of the suitable environment for rendering the GIF. Also some more insight must be given on possibilities (or necessity) of emulation.

Or, the other option is to ingest the GIF as a reference file and to create (or acquire) an MP4 that resembles the original GIF. For this option no additional preservable format is needed; the MP4 is treated as the archive master and will be preserved as any web video, as presented in the first case.

In both options, the main question is: how can we establish whether rendering the master file represents the original work? The only difference between both options is: do we assess this later, given the requirements at that point in time, or do we make this assessment now, at the moment we accept the MP4 as peer. Either way, the archive must define what significant properties it wants to preserve, for whom and with what costs.

These scenarios and their implications must (and will) be addressed in a preservation plan. As a basis the context of the plan will be described (triggered by collection policies, typology of the main designated community). The GIF-object will be explained followed by the requirements that must be met like the extension to the PMD or the procedure of consent to the acquired or created MP4.

### *B. Games*

For GIF, emulation was introduced as a way to render the original GIF, provided you simulate the original environment. For Games emulation is the only option, as there is no working substitute for the interactive feature of the game. After all a single standard format that can represent all possible interactive user experience does not exist.

In the NISV preservation plan on games the following three requirements are included, because these will have to be met in order to preserve games in the NISV repository.

Firstly the PMD should be extended with the new preservation format for Games (disk images that hold the original game-software). Find a way to document additional content like instruction videos. In a PREMIS-schema (fig. 5) is shown how this should be done. Several rights have to be managed too. The environment is added as a separate object.

For now Sound and Vision chooses not to archive environments but it must document the characteristics, to be able to create or emulate the environment when needed.


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

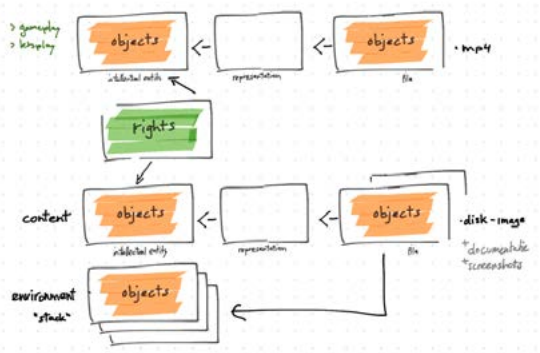


Figure 5 Schema of object categories for preservation metadata on games; based on PREMIS.

Secondly, consider that the policy on digital preservation doesn't support emulation yet. This should be added in the next update. This implies that:

- emulation must be added, next to migration
- the preservable format *disk image* should be added
- and preservation service levels as in what do we promise to preserve, should be redefined

Thirdly, to monitor the longevity of the games, NISV will organise a 5-year monitoring cycle. Once every 5 years it will check the rendering of disk-images. Is NISV still able to configure the hard and software that runs the game? Will new versions of emulators still do the job?

This brings to mind our just in time policy for webvideo: it is the same challenge: will new versions of the transcoder still be able to transcode the source files to the standard proxy? And: will this be an acceptable norm for our designated community? In a "just in time" policy NISV must somehow organise a trigger not to be too late!

## VI. The preservation plan

By creating a standard table of contents for the NISV Preservation Plan, better informed decisions by management are ensured. The Preservation Plan at NISV has 4 sections.

First the outline of the context of the plan. What triggered the plan. What risks are to be mitigated; for instance legal issues, legacy or increasing backlogs. Then specific goals of the plan and the foreseen impact on digital assets already in the Archive are to be addressed.

Secondly the collection itself is described. Which Designated Community is leading, and what will be the designated use; the nature and scale of the expected ingest, the 'significant properties' of the material, and notes on selection criteria or demarcation in agreement with other archives in the Netherlands.

Third it defines what requirements are to be met. Special attention is paid to preconditions or assumptions regarding technical issues, planning, internal users (availability, competences), and internal procedures to be redesigned, implemented or just applied.

Then, at the heart of the Preservation Plan are the scenarios, followed by a recommendation. The scenarios may differ in the outline of the preservation strategy, chosen preservation formats, implications for the metadata dictionary, technical requirements, and so on.

These Preservation Plans are discussed by the NISV preservation board and as a result may lead to assignments to implement tooling, prepare specific upgrades to IT infrastructure or start prototyping a new format. Also, the outcome may be the formulation of add-ons to the preservation metadata dictionary, or even to current preservation policies themselves.

## VII. Conclusion

With three cases it is shown how preservation planning at NISV plays a role in checking preparations for new ingest to standing preservation policy. And how it suggests updates to this policy. Preservation planning gives NISV archival management the opportunity to make deliberate choices on preservation. And the documentation makes these choices transparent.

The way NISV adopted preservation planning is consistent with OAIS. It differs from the implementation by the Planets project, although it certainly has corresponding elements. Especially the outline of the NISV Preservation Plan owes to the work done by this working group.

Also the way the two monitoring functions can trigger a preservation plan is very similar to Planets.


16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

NISV has combined the two functions accordingly. This "preservation watch" is in reality an abstract state of mind and sense of responsibility of all colleagues that have knowledge of audio visual technology, in house, outside at our DC's or in the field in general. Given the topicality of preservation issues Sound and Vision will mobilise this implicit knowledge by organising meetups on these issues.

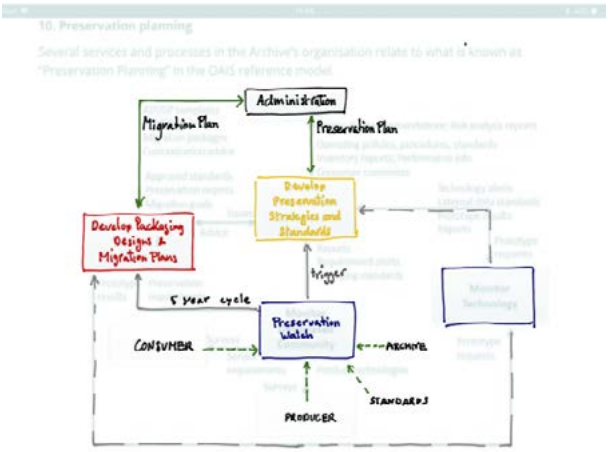


Figure 6 NISV adoption of Preservation Planning within OAIS.

The schema (fig. 6) shows how NISV has adopted preservation planning. The NISV preservation plans are triggered by risk alerts from Preservation watch. The plans, together with risk assessments, standards (like the PMD) and policies add up to the knowledge base of Administration. All operational preservation actions by Administration build on this knowledge base. Parallel on the drawing of preservation plans is the set up of migration plans. Preservation watch fosters this function by a cyclical process, like the five year cycle for the "just in time" policy.

Preservation planning is not the equivalent of a once every five year general policy on preservation. Neither it is reduced to the preparation of preservation actions on a specific set of objects. It stretches out over adjustments or add-ons to preservation policies on one side and the set up of concrete migrations on the other side. As some beacons will guide our main course with a reassuring light on the horizon, while other beacons will set out a strict direction that must be followed. Each will help us reach our preservation goals, even in poor weather or heading for unknown shores.

## Acknowledgment

NISV thanks Annemieke de Jong for her work on making digital preservation a leading principle for our institute. Her outstanding work on preservation policies and TDR certification could pave the way for a organisation-wide tackling of preservation issues and inspires many colleagues, at NISV, as well as elsewhere to take up the challenge of designing longevity for our collections.

## References

[1] M. Steeman, J. Schuurman, "Preservation metadata for DPX files: A case study" Journal of Digital Media Management, vol. 7, no. 1, pp. 33-45, Autumn 2018

[2] A. de Jong, "Digital Preservation Sound and Vision: Policy, Standards and Procedures," Netherlands Institute for Sound and Vision, 2016

[3] C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, H. Hofman, "Systematic planning for digital preservation: evaluating potential strategies and building preservation plans", International journal on digital libraries, vol. 10, issue 4, pp 133-175, 2009, http://www.ifs.tuwien.ac.at/~becker/pubs/becker-ijdl2009.pdf

[4] H. Hofman, Planets-PP subproject, C. Becker, S. Strodl, H. Kulovits, A. Rauber, "Preservation plan template. Technical report", The Planets project, 2008, http://www.ifs.tuwien.ac.at/dp/plato/docs/plan-template.pdf

[5] OAIS2012, CCSDS, "Reference Model for an Open Archival Information System (OAIS)", Magenta book, http://public.ccsds.org/publications/archive/650x0m2.pdf, 2012.