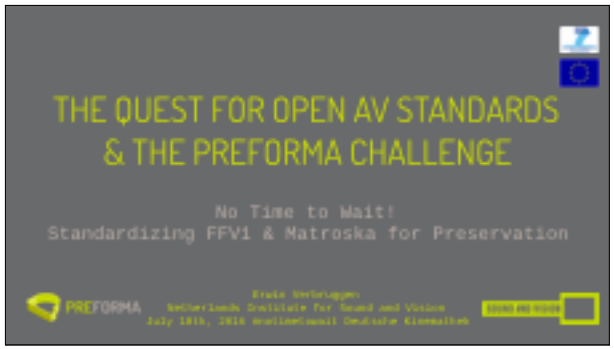


1

Hello, my name is Erwin Verbruggen and I work at the Netherlands Institute for Sound and Vision.

I am here to tell you about the PREFORMA project and how it relates to us all being here.

2

Two Efforts Merging

Story of MKV / FFV1 comes from two sides:
1. Archivists, specialists, users => Story will be told by Peter
2. Archivists supported by European Commission => Story that I'm here to tell

(c) Olexandra Lytvyn

3

The PREFORMA Project “Empowers memory institutions to gain full control over the technical properties of digital content intended for long-term preservation”

The PREFORMA Project

The main objective of the project is to give memory institutions full control of the process for testing the conformity of files to be ingested into their archives. This shall be obtained by developing a set of tools which will enable the testing process to happen within an iteration that is under full control of the memory institutions.

PREFORMA is a Pre-Commercial Procurement project. This means that the research and development work will be realised by the suppliers selected through a tender procedure, which is integral part of the Pre-Commercial Procurement project.

The procurement, following the rules for tenders in public sector, matches the memory institutions professional knowledge and the supplier’s skills in development and promotion of products.

The total costs of the procurement amounts to €2,805,000

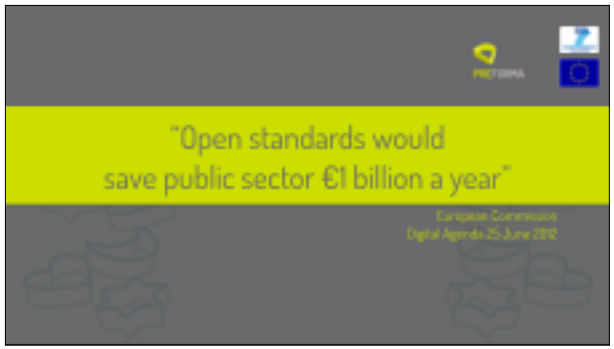
4

14 partners from nine European countries
includes memory institutions, research centres, and universities

This list of participating institutions includes very different types of organisations, both in size, set-up and mission. The conformance checker should validate an audiovisual file format that fits both the preservation strategy of each of these four institutions and the umwelt that they perform in: the number of connected organisations that deliver to and depend on these archives' services.

The PREFORMA project procures the development of a conformance checker for an audiovisual preservation format. The requirements for this conformance checker will be based on the needs of four different audiovisual collections:

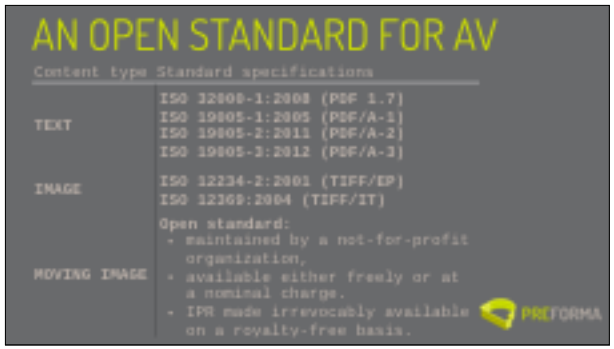
- Sound & Vision, a large broadcast archive
- Greek Film Centre, a small film archive
- National Library of Sweden, a large national library
- Girona Record management services, a small city archive



5

From press release + report (2013): Against lock-in: building open ICT systems by making better use of standards in public procurement.

“The Commission today releases a new policy to help public authorities avoid dependence on a single ICT supplier. Following the recommendations in this new "against lock-in" approach could save the EU's public sector more than €1.1 billion a year. For example, open tendering procedures can attract increased numbers of bidders with better value bids (doubling the number of bidders typically lowers contract size by 9%).”



The case against MXF (Lustrell)

For PREFORMA, fitness for long-term preservation depended on four criteria:

#1. A format that can capture content in an uncompressed or mathematically lossless encoding, and retain as many original properties as possible (ie. bit-depth, resolution, …) because this is a requirement for implementing an effective preservation strategy. This meant ruling out all well-adopted lossy codecs (ie. D10).

#2. A format that is Free/Libre (cf EIFv1)
This was a key issue for the PREFORMA project. Software developed within the scope of the project must be licensed under a specific copyleft license (i.e. “GPLv3 or later and MPLv2 orlater”) to ensure that the tools for checking archive files will be available over very long periods.
=> This requirement ruled out the MXF wrapper, since the current licensing of the required MXF specifications do not allow for implementing a validator under the given license. Moreover the specifications for MXF are behind a paywall and thus limit their collaborative use.

#3. A format that is well documented
The questions we asked ourselves were: is there a specification available? Is it accessible and maintained? Otherwise, you can’t write an implementation and create false or correct test files.

#4. A format that has been adopted

7

Currently, these four memory institutions basically store audiovisual reproduction and access files, i.e. AVI, MOV wrappers with MPEG2 and MPEG4 streams, DCP and MXF wrappers with IMX-D10/30/50.

These files are not suitable for long term preservation workflows because:
the files use lossy compression, which complicates future migration and transcoding of the file (risk for the quality of the file)
the files use proprietary formats and encodings, which complicates future availability of software for reading and writing the files. (risk for the accessibility of the file)

None of the PREFORMA partners uses files that fully meet the state-of-the-art requirements for preservation formats as described comprehensively by the Library of Congress

8

How Sound and Vision chose its Standards

Sound and Vision is the business archive of the Dutch broadcasting corporations and one of the largest audiovisual archives in Europe.

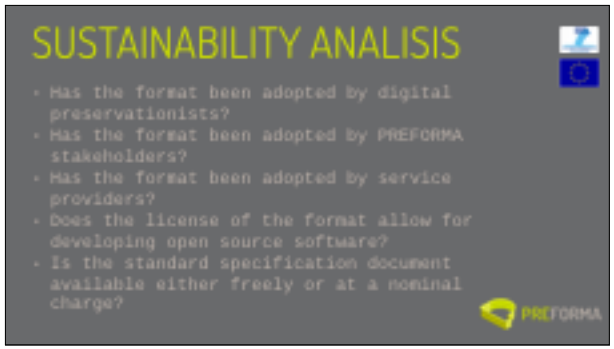
We make available tens of thousands of hours of archive footage online for various end-user services, including dedicated services for the creative industries, education and research. The institute operates as a visitor attraction aimed at the general public and is visited by over 250,000 people annually.

—-

The audiovisual files that memory institutions receive need to be highly adapted to the (post-)production and distribution workflow of the domain in which they are produced. Files from broadcasters are made for streaming and efficient editing. Files from film production have extensive functionality for encrypted access to the content. These properties often complicate long term preservation of the content.

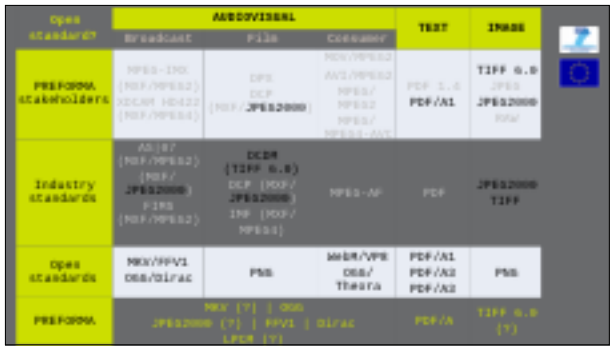
Workflows in audiovisual collections often rely on professional broadcast and film production software for efficient and reliable processing of audiovisual files. Consequently, collections prefer using the corresponding proprietary formats for managing these files.

Audiovisual standards leave a lot of room for interpretation, which means that different production systems, in turn, often only accept or work with a specific interpretation of a standard.



9

Cost for acquiring necessary specifications and access to the standardization process of MXF was too high for an open source project, in particular in comparison with MKV. Moreover, PREFORMA could not tender for MXF because of Swedish Law that obliges public services to tender for open formats as defined in the EIFv1.
PREFORMA considered that working on the documentation and standardisation during the course of the project would be more feasible with the MKV community than with SMPTE for MXF. Given that there was already a validator available for MKV, the project expected that the investment in this container would pay off more.
We chose for standards where we could imagine that the existing shortcomings could be solved within the scope of the project. We evaluated the tenders in two stages. So we have chosen the project that offered the best strategy for solving the existing shortcomings.
Even if we would like to switch from MKV to MXF now, MXF does not permit to wrap FFV1 essences. From what we've been told it would be possible, but it would require some money to pay for a codec id registration.

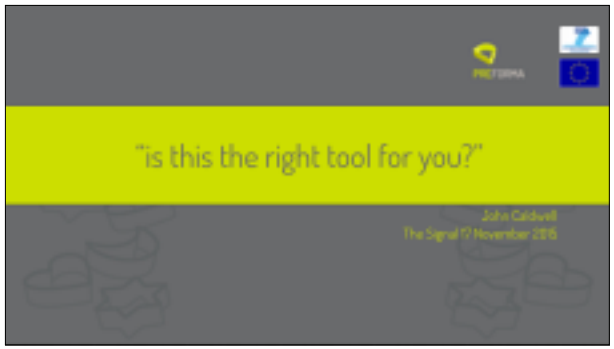


10

What these criteria lead to:
None of the standard specifications of text, image our audiovisual material matched all four criteria. After discussion in the project board about the preferred way to make a compromise, the project board decided to insist on the first two criteria and see how the shortcomings in the two remaining criteria could be solved within the scope of the project.
Concrete: To what extent can a lack of documentation be solved in the project? To what extent can the project demonstrate the viability of the format?

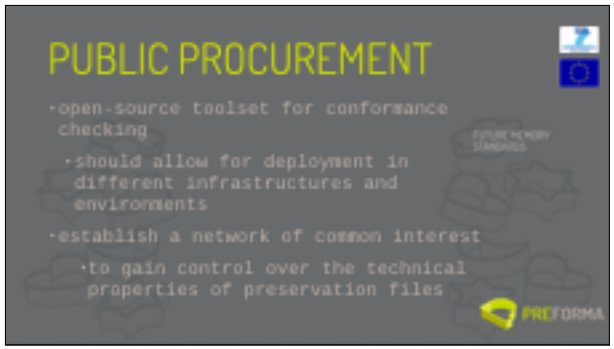
The result of this exercise for audiovisual ended up in a set of two containers (MKV and Ogg) and three video/image codecs (JPEG2000, FFV1 and Dirac). Tenderers have been invited to make an R&D project proposal for a combination of a container and codec from this list.

Please note that the procurement procedure had two steps before the actual development could start. Proposals have been evaluated first on the criteria of the Invitation to Tender. PREFORMA invited two tenderers to make a design. The two designs have been evaluated based on how they dealt with the shortcomings of the file formats.



11

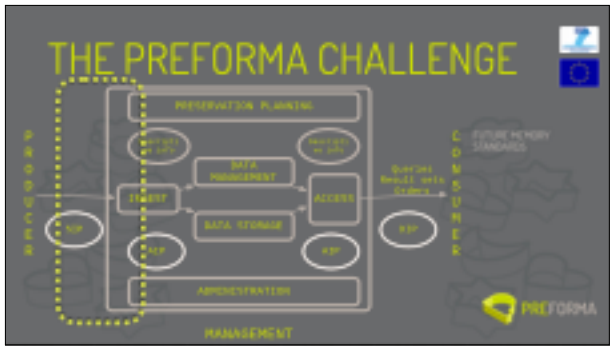
"There are lots of tools out there, from checksum validators to digital forensics suites and wholesale preservation solutions. Many people feel it's important to have the latest and greatest […], but in this impulse for the new and the now, we sometimes forget to ask the big question: For your institution, is tool X good enough?"



12

PREFORMA research and development activities aim to empower memory institutions to gain full control over the technical properties of preservation files. This is achieved through the development of an open-source conformance checker and the establishment of a healthy ecosystem around an open-source ‘reference’ implementation for specific file formats.

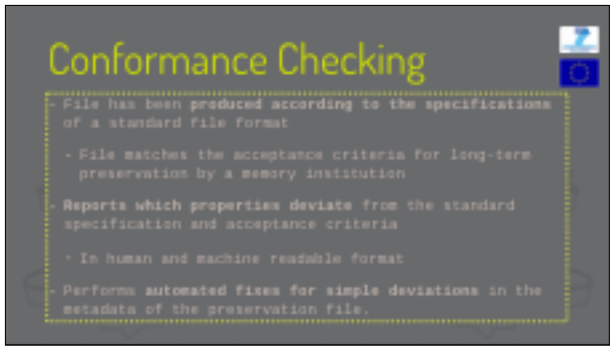
All software developed in each Open Source project will be provided under two specific Open Source licenses, namely: both Mozilla Public License “MPL v2.0 or later” and under GNU General Public license 3.0 “GPLv3 or later”. Further, related to each Open Source project, all digital assets , associated with the Open Source project , including all synthetic files developed for the file format , will be provided under the Creative Commons (CC) license Attribution - ShareAlike 4.0 International (CC BY - SA 4.0) and hosted on the open platform.

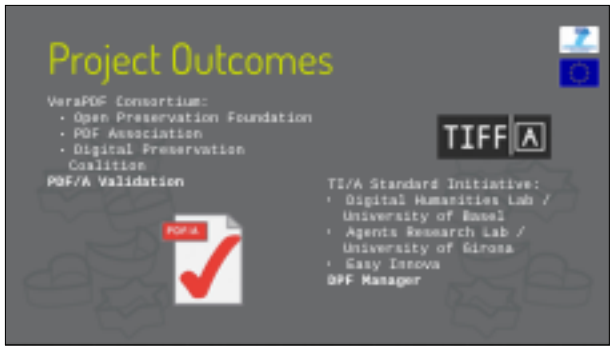


The solution proposed by the PREFORMA project is intended for use within the OAIS Reference Framework, focusing in particular on the OAIS ingest functions, which represent the first step in the digital curation and preservation workflows.

PREFORMA aims to work on one specific step in the OAIS process, i.e. file validation.

Developed in an iterative process with multiple releases and with a number of experiments with ‘real’ data sets (files) from memory institutions during each iteration


Conformance Checking
- File has been produced according to the specifications of a standard file format
- File matches the acceptance criteria for long-term preservation by a memory institution
- Reports which properties deviate from the standard specification and acceptance criteria
- In human and machine readable format
- Performs automated fixes for simple deviations in the metadata of the preservation file.



15

The first design phase lasted four months

Six suppliers were selected (two for each media type) and the suppliers of the three best designs were requested to proceed to the current prototyping phase.

The question of an open source reference format was rather straight-forward for documents and images. The VeraPDF consortium works with industry partners on the PDF/A standard. Supplier Easy Innova developed the TIFF/A standard initiative arround which it develops its conformance tools.

The MediaArea team has worked on several software tools used in preservation data conformance, including MediaInfo, a tool to extract metadata from audio and video files, and Bay Area Video Coalition's QCTools project, a tool that performs quality control checks and analysis on digitised analog video.

MediaArea is one of three teams that have moved forward from Phase 1 (design) into Phase 2 (software prototyping). This second phase will last 22 months. The research conducted during Phase 1 is available on PREFORMA’s Open Source Portal. In collaboration with Open Preservation Foundation and Easy Innova, MediaArea will develop a software framework for checking media files.

In Phase 1, MediaArea was tasked with researching community standards and developing file conformance checks for FFV1 (a lossless video codec), Matroska Multimedia Container, and Linear Pulse-Code Modulated Audio (LPCM) to support efforts of long-term preservation in memory institutions.

In Phase 2, MediaArea will be developing MediaConch, an extensible, open source software project consisting of an implementation checker, policy checker, reporter, and fixer that targets preservation-level audiovisual files. It will provide detailed and batch-level file conformance checking via an adaptable and flexible application program interface accessible by the command line, a graphical user interface, or a web-based shell.

Our prototype phase objectives and workplan include collaborations with partners from Matroska, FFmpeg, and libav. MediaArea will also work closely with Artefactual Systems in


Thank You.
everbruggen@beeldengeluid.nl
twitter: @erwinverb - @benglabs
PREFORMA