https://doi.org/10.2352/issn.2168-3204.2021.1.0.10

EN 17650 – The new standard for digital preservation of cinematographic works

Siegfried Foessel, Heiko Sparenberg; Fraunhofer IIS; Erlangen; Germany

Abstract

EN 17650 is a proposed new European Standard for the digital preservation of cinematographic works. It allows organizing of content in a systematic way, the so called Cinema Preservation Package (CPP).

The standard defines methods to store content in physical and logical structures and describes relationships and metadata for its components. The CPP uses existing XML schemes, in particular METS, EBUCore and PREMIS to store structural, descriptive, technical and provenance metadata. METS XML files with their core metadata contain physical and logical structures of the content, hash values and UUIDs to ensure data integrity and links to external metadata files to enrich the content with additional information. The content itself is stored based on existing public and industry standards, avoiding unnecessary conversion steps.

The paper explains the concepts behind the new standard and specifies the usage and combinations of existing schemes with newly introduced metadata parameters.

Motivation

The advent of Digital Cinema changed not only the distribution and the projection/display of cinematographic works, but also the entire production workflow; and with it the types of preservation items. Most cinematographic works today are completely digital, and are no longer recorded and stored on analog film.

This arrived together with a manifold of new digital formats during the production and mastering process. Whereas the Digital Cinema Package (DCP) [1] is a stable format for the distribution to the cinema and the equivalent to the film distribution reel, a common standard for the preservation of assets used during production, mastering or in general around a digital cinematographic work did not exist. This was the motivation in 2014 to propose a new standard for the preservation of cinematographic works. The proposal was added to the "Rolling plan for ICT standardization" [9] and granted as new CEN project in 2019. The work on the standard started at the end of 2019 with a small project team under the supervision of CEN TC457 and is now under ballot as proposed European Standard prEN17650 [2].

Problem

Digital workflows in the media industry are adopting quickly new technical achievements and change the used file formats very often. This is in contrast with the wish of archivists for a long-term preservation format, description and structure. The question is: How can digital formats for cinematographic works be stored in such a way, that they can be interpreted reliably also in 10, 20 or 50 years from now. The new standard tries to answer this question and offers methods to help in this preservation task. It does not answer the question on which media type the files should be stored and carried. It does also not prevent the content owner to store the assets in formats not appropriate for long-term preservation.

Approach

Structure

The first requirement for the preservation standard was the definition of a clear package structure that allows storing the content in a well-defined physical order and in parallel in logical collections. This was achieved by using and adopting the METS schema [3][4]. To allow splitting and exchanging parts of the cinematographic work a hierarchical system of METS files and sub-packages was defined (see figure 1).

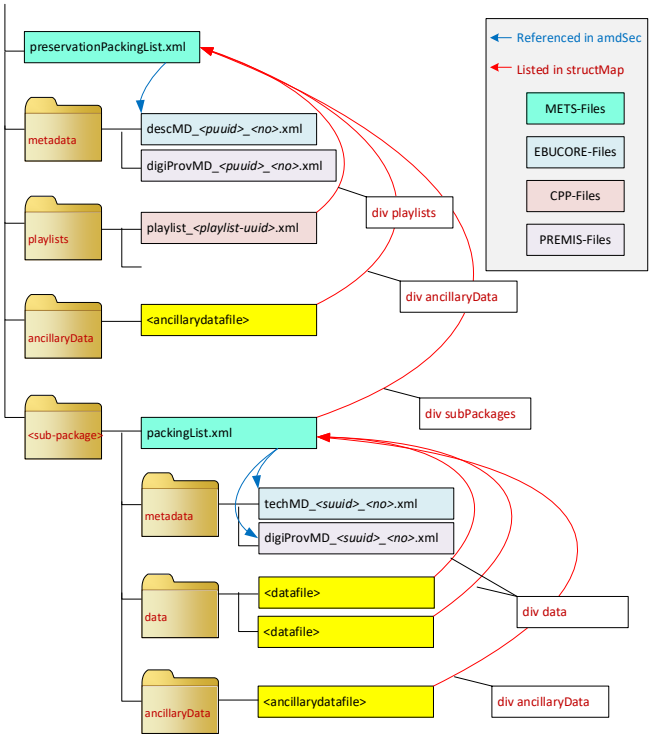

Figure 1: Basic structure elements in a CPP

The METS files (named *packingList.xml* in the sub-packages and *preservationPackingList.xml* in the root folder) finally link descriptive, provenance and technical metadata to the content files. The cinematographic work in a CPP is organized in multiple sub-packages with the main assets in *data* folders. The data files can be image sequence files (in an imagePackage), sound files (in a soundPackage), timedText files (in a timedTextPackage), composed files like MP4 files (in an audiovisualPackage), a set of files belonging together as a package like DCP, DPP or IMF package (in a componentizedPackage) or extra files (in an extraPackage). Multiple asset types are defined. The *packingList.xml* file in the root sub-package folder is of type METS, list all asset files and link them to technical and provenance metadata files in the *metadata* folder. Optionally ancillary data files can be added in the *ancillaryData* folder. The reason for the organization in sub-packages was that in daily operations only parts of a cinematographic work are processed or moved. In consequence, not all parts of the CPP should be touched or scanned when a content element is changed, added or removed. Especially the re-generation of hash values can lead to time-consuming workloads.

The METS files are used to list all files with their checksums, create logical and physical views on the content and link data files to other metadata files. Technical, descriptive and provenance metadata are not embedded directly in the METS files. One reason is that METS itself does sometimes not offer good metadata descriptions and the option embedding of other schemes increases the complexity for schema checking and complicates manual editing. The same structure – metadata, ancillary data and assets - is repeated on the top level of the package, but instead of referencing data files in the *data* folder, the *preservationPackingList.xml* file references the *packingList.xml* files in the sub-Packages, generating a hierarchical system of METS files. In the *preservationPackingList.xml* file also files for descriptive metadata related to the complete cinematographic work are referenced.

One addition in the root of a CPP is the use of so-called playlists in the *playlist* folder that allows generating pre-composed content. This is not only used for a playable piece of content e.g. a composition of image sequence and sound files, but can also be used simply for a composition of files belonging together. The playlist binds individual data files together whereas the logical view in the METS *preservationPackingList.xml* file binds subpackages together.

The reader may detect structure similarities to the Information Packages of the E-ARK project [5]. This similarity is by design. Some ideas are adopted from the E-ARK project and adapted to the needs of the cinema and movie industry. This adoption is also related to the second requirement in the project. Existing standards should be used as much as possible; it was not intended to reinvent the wheel. For this reason multiple standards and projects were investigated to test the applicability in the use case "cinematographic works".

Metadata

As a result of the investigation of metadata standards, the following formats are proposed in the CPP:

- METS for structural metadata, hash values and linkage of files
- EBUCore [6] for technical metadata
- EBUCore for descriptive metadata (based on EN15744)
- PREMIS [7] for provenance metadata

All proposed metadata standards could be adopted with its existing schema descriptions and inherent extension mechanism. Means, no new schema definition was necessary.

EBUCore was selected for the description of the technical metadata for two reasons. First, it already offers a large set of technical metadata parameters for the video and movie industry. In addition, some tools exists like MediaInfo that can automatically generate basic technical metadata files. Second, EBUCore has an easy extension mechanism that allows adding new metadata parameters without redefining the standard (see example in figure 2). In the standard a full set of technical metadata parameters are defined together with its mandatory or optional availability.

For the descriptive metadata the selection was not so obvious. With EN15744 [8], a minimum set of metadata for cinematographic works already exists. However, the standard describes only the data elements as ontology, but not the language or syntax for an implementation. This led to different implementations in the film archive world. The lack of examples also led to some confusion and differences in its use. During the project the Deutsche Kinemathek worked together with the Zuse Institute in Berlin on a new EN15744 implementation based on EBUCore for their internal purposes. This implementation was contributed to the project team that adapted it in a more generalized way for the CPP (see figure 3).

Simplified Example of EBUCore technical data:

```
<ebucore:format formatName="imageFormat">
  <ebucore:imageFormat>

     <!—EBUCore existing parameters -->
     <ebucore:width>1998</ebucore:width>
     <ebucore:height>1080</ebucore:height>
     <ebucore:aspectRatio>
        <ebucore:factorNumerator>185</ebucore:factorNumerator>
        <ebucore:factorDenominator>100</ebucore:factorDenominator>
     </ebucore:aspectRatio>

     <!—EBUCore extensions -->
     <ebucore:technicalAttributeString typeLabel="imageCodecStandardReference">JPEG 2000</ebucore:technicalAttributeString>
     <ebucore:technicalAttributeString typeLabel="colourReference">ITU-R Rec. BT.709-6</ebucore:technicalAttributeString>
     <ebucore:technicalAttributeUnsignedInteger typeLabel="componentNumber">3</ebucore:technicalAttributeUnsignedInteger>

  </ebucore:imageFormat>
</ebucore:format>
```

Figure 2: Example of technical metadata with EBUCore

Excerpt of EBUCore descriptive metadata:

```
<ebucore:coreMetadata>

            <!-- EN15744 'title'-->
            <ebucore:title typeLabel="originalTitle">
              <dc:title xml:lang="en">New York City</dc:title>
            </ebucore:title>
            <ebucore:alternativeTitle typeLabel="releaseTitle">
              <dc:title xml:lang="de">New York City - Die Stadt, die niemals schläft</dc:title>
            </ebucore:alternativeTitle>

            <!-- 15744 'production company' -->
            <ebucore:publisher>
              <ebucore:organisationDetails>
                <ebucore:organisationName>Rising Filmproduction, Berlin</ebucore:organisationName>
              </ebucore:organisationDetails>
              <ebucore:role typeLabel="productionCompany"/>
            </ebucore:publisher>

            <!-- 15744 'cast' -->
            <ebucore:contributor>
              <ebucore:contactDetails contactId="http://d-nb.info/gnd/1098357507">
                <ebucore:name>Wegner, Ulrike</ebucore:name>
              </ebucore:contactDetails>
              <ebucore:role typeLabel="cast"/>
            </ebucore:contributor>

            <!-- 15744 'credits' -->
            <ebucore:contributor>
              <ebucore:contactDetails contactId="http://d-nb.info/gnd/119069342">
                <ebucore:name>Wiesinger, Lothar</ebucore:name>
              </ebucore:contactDetails>
              <ebucore:role typeLabel="credits"/>
              <ebucore:role typeLabel="director"/>
            </ebucore:contributor
…
```

Figure 3: Desciptive metadata in EBUCore syntax

IDs

METS uses a lot of internal IDs for referencing and cross linking of elements. For a CPP however also global IDs are beneficiary, especially to offer such IDs to archive management systems. In Figure 1 the use of UUIDs is shown in the filenames as puuid (preservation package UUID) or suuid (sub-package UUID). In addition the sub-package name contains the suuid. This allows an easy identification of identical sub-packages or preservation packages in the archive system and their related metadata. Each time the data elements change a new UUID will be generated.

Data formats

The CPP so far is agnostic to the used content formats in the *data* folder. However, recommendations for formats are listed in the standard. In general, well-defined formats should be used, like ISO, ITU, SMPTE, IETF standards etc. that can be referenced in the metadata files and for which open software implementations are available. Special placeholders in the metadata files are available for these references. In case proprietary formats have to be stored and archived, a description of the format should be stored together with the data files in the related *ancillaryData* folder.

Results

At the begin of the project the focus was more on the physical structure of the preservation package. During the project, requests from the archives were received to add also logical views and to give more guidance on the metadata usage. Especially for the technical metadata many new metadata parameters were added to the standard. A complete overview of components in a cinema preservation package is shown in figure 4. It also adds an additional optional folder, called *checkerReports*. This folder is reserved for post-analysis files after the CPP was created, containing e.g. checks if hash values are still correct, schema checks or consistency checks. These files are not included in the file lists inside the METS files. The folder is considered more a reserved storage place for such files.

The result of the work is a complete standard for describing a Cinema Preservation Package. A related Technical Report, which is already available in draft version, gives additional information and guidance how to implement the standard and offers additional explanations to the structure. Various types of content combinations are described as reference for concrete implementations in the technical report.

Figure 4 also shows all components of a Cinema Preservation Package together with the used schemes for the metadata. Actual work is ongoing with the development of an open source reference software that will enable a better understanding of the standard. At the time of the publication of this paper the standard inquiry deadline is passed, so that the project team will process the comments to the proposed standard.

Conclusions

The reuse of existing standards and metadata schemes led to many discussions inside the technical committee TC457 responsible for this project and the project team. At the end it needed a lot of training into existing standards and schemes, as the specific requirements created complex dependences. But with the now proposed standard a good solution was found which can easily be extended for the future and is based on already existing formats in the archive world.

The authors wants to thank the European Commission for the project support in the CEN TC457 for preservation of cinematographic works, in addition Karin Bredenberg for her guidance to the E-ARK project and the METS description, the EBU for their support in the EBUCore, Deutsche Kinemathek and Zuse Institute for their support in descriptive metadata implementations.

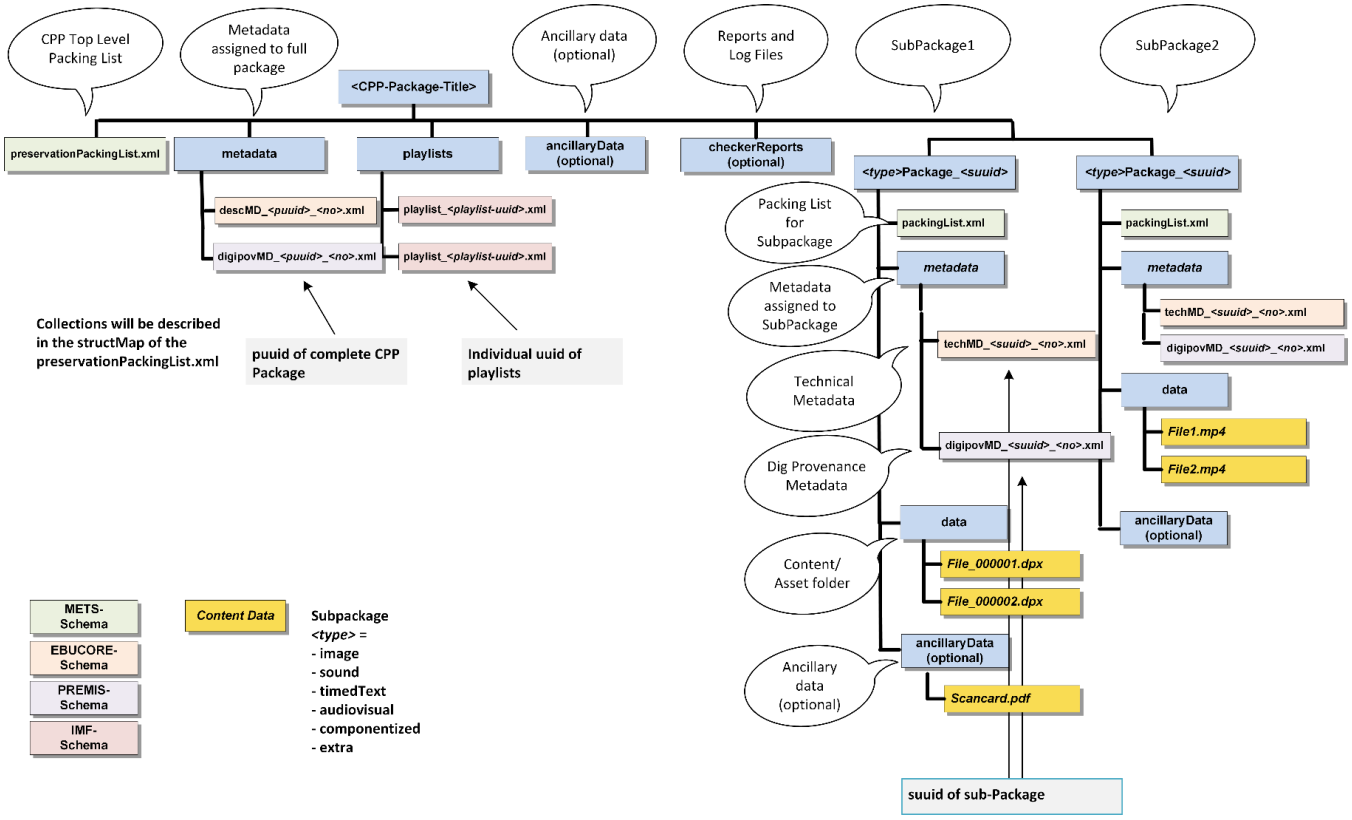

Figure 4: Components of a CPP with its related metadata

References

[1] DCI Digital Cinema Initiatives LLC, "Digital Cinema System Specification Version 1.4", Online on 19th May 2021, https://www.dcimovies.com/specification/DCI_DCSS_v1-4_20-July-2020.pdf

[2] CEN prEN17650: "Digital preservation of cinematographic works", https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:70021,2368598&cs=1292AD421DDA2852EDCB5C69C35B403A2, Feb. 2021.

[3] METS, "Metadata Encoding & Transmission Standard", http://www.loc.gov/standards/mets/, Feb 2021.

[4] METS Primer V1.6, https://www.loc.gov/standards/mets/METSPrimer.pdf, Feb 2021.

[5] E-ARK Project, "E-ARC CSIP Common Specification for Information Packages", https://earkcsip.dilcis.eu/pdf/eark-csip.pdf

[6] EBUCore, "Metadata Specifications", https://tech.ebu.ch/MetadataEBUCore, Feb 2021.

[7] PREMIS, "Preservation Metadata Maintenance Activity", https://www.loc.gov/standards/premis/, Feb. 2021.

[8] EN15744, "Film identification — Minimum set of metadata for cinematographic works", http://filmstandards.org/fsc/index.php/EN_15744, Feb 2021

[9] Website European Union, "Rolling plan for ICTR standardization", https://digital-strategy.ec.europa.eu/en/policies/rolling-plan-ict-standardisation, as of 23.05.2021

Author Biography

Siegfried Foessel is head of the department Moving Picture Technologies at Fraunhofer IIS, Germany, and professor for the technology study program at HFF film school in Munich. His previous research projects include: digital cinema cameras, certification test plan for digital cinema, image coding and post-production tools for digital cinema and immersive audio. Siegfried is member and chair of various standardization bodies. He is FKTG president, SMPTE fellow and IEEE senior member.

Heiko Sparenberg is head of the research group Digital Cinema at Fraunhofer IIS, Germany. His main research activities include utilization of scalable media for improving data workflows, compression technologies based on machine learning approaches as well as the development of post-production software for professional movie productions.