THE EYE FILMMUSEUM DIGITAL FILM REGISTRATION MODEL

By Nils van der Meer, Anne Gant and Walter Swagemakers

November 2018

Table of contents

1 FOREWORD

The cinema industry has witnessed a digital revolution in recent years, as the production and presentation of film in the broadest sense of the word (including feature films, documentaries, experimental films and amateur films) has switched from analogue to digital. Cultural institutions that collect films, such as the Eye Filmmuseum, museums and archives, are now having to deal with an rapidly growing influx of film material that is 'born digital'.

In addition to this trend, a massive volume of analogue film has been digitised in recent years. In order to guarantee the sustainable conservation of digital film objects and to ensure that they remain usable in the long term, it is absolutely vital that curators of film collections adopt a registration system that is suited to this type of material in their collections. A digital film registration model is a vital link in the digital sustainability chain, from ingest through long-term storage to access.

In order to manage this digital work flow, digital formats need to be standardised and work flows need to be automated. This is possible only if a digital film registration model is adopted, for both catalogues and media asset management systems. This paper explains the model adopted by the Eye Filmmuseum in Amsterdam, in the hope that other archives and museums facing similar challenges will find this information useful.

This paper has been produced as part of the Sustainability Programme operated by the Dutch Digital Heritage Network. The members of the network wish to share their knowledge with the audiovisual community. If you have any queries about the contents of this paper, please contact Walter Swagemakers (Senior Project Manager Collections, Eye Filmmuseum, Amsterdam) at: walterswagemakers@eyefilm.nl.

2 INTRODUCTION

This article describes a model for registering digital AV objects. Our goal is to present the model in such a way that other organisations who are faced with similar problems can use it, too. We do not justify the model or explain the choices we made in developing it. We acknowledge that there are a wide range of alternative options in this respect. It is not our intention to set a standard. Rather, we wish to offer the reader a set of practical guidelines.

The description of our model uses various terms from the Open Archival Information System (OAIS) reference model, in the hope that readers are familiar with the terminology. The model as set out in this article uses the Archival Information Package (AIP) as defined in the catalogue and the digital archive. We assume that the reader is familiar with the OAIS model and for this reason do not attempt to explain it.

3 BACKGROUND

We decided not to adopt the FRBR (Functional Requirements for Bibliographic Records) model. This is because a digital copy – such as a Digital Cinema Package (DCP) or a Digital Distribution Master (DCDM) – is capable of containing more than one expression of a single cinematographic work. For this reason, the FRBR model is not a particularly suitable model for registering digital films.

Here at the Eye Filmmuseum, it used to be customary for digital AV material to be described in the catalogue in the same way as analogue AV material. However, there are a number of additional problems involved in describing digital AV material as compared with analogue AV material:

- A digital film object is **complex**.
  It may consist of multiple files or copies in a range of different formats. This means that the structure needed in order to describe these objects is considerably more complex than that needed for an analogue film object.

  The management and use of such objects is also more complex, given that the validity of a digital film object is not easy to guarantee. For example, an entire digital object may be damaged simply because one of its component parts is either missing or damaged. Additional automation is needed in order to manage this complexity.

- There is the problem of **multiplicity**.
  Given the ease with which a new digital film object can be produced on the basis of an existing object (for example, in a different format or for a specific purpose), there is a need to describe and manage a much larger number of objects.

- A digital film object has a **more complex range of metadata**.
  If digital film objects are to be managed in a digitally **sustainable** manner, it is absolutely vital to ensure that a careful record is kept of the technical metadata. Digital objects come with far more technical metadata than analogue objects. As a result, recording and verifying a digital object is a much more labour-intensive process than recording an analogue object.

- A digital film object is **modifiable**.
  Digital film objects are much easier to modify than their physical counterparts. This has two important consequences: firstly, any procedures involving digital film objects must be planned in such a way that they do not alter either the film object itself or the metadata on the object. Secondly, the technical metadata on the digital object is not always reliable (for example, the modified date of a file in a file system may change when it is copied). This means that a mechanism needs to be found for safeguarding the relevant metadata. Where a digital object is converted into a different format as part of a digital sustainability programme, it is also important to preserve the metadata on the preceding format(s), even if the original digital object no longer exists.

- There is a much greater **diversity** of digital objects.
  All sorts of related digital objects may be supplied in addition to the primary object (such as a

feature film, documentary or art film). These include promotional materials, stills, teasers, trailers, scenarios and press kits, for example. There is a much larger quantity of related objects than used to be the case with analogue film objects. All these various objects are also related to each other in various ways and need to remain related in the long term. They also need to be described and managed.

As we were working on the model, we decided, for reasons of efficiency, to host and manage the digital archive ourselves. This has huge advantages for the various work flows involved, given that it more or less puts an end to the practical problems posed by the size of the files. Certain types of digital AV materials are so large that, in the past, physical carriers (such as HDs or LTO tapes) were used to exchange data with an external digital archive instead of uploading and downloading them. This made it very difficult to automate the work flows.

4 MODEL: ARCHIVAL INFORMATION PACKAGE (AIP)

The specific circumstances at Eye mean that we cannot rely solely on open standards. In certain respects, open standards are still far from mature, are not well supported by software and are hardly ever used by film-makers and producers. In other words, our model needed to be capable of supporting a range of different formats, both now and in the future.

## 4.1 Structure of the model

4.1.1 Controlled inheritance

We used a system of controlled inheritance. Inheritance ensures that properties that are shared by all digital AV objects also share a common definition, and that properties that are restricted to certain specific types of AV objects are defined only for those particular types of object. The term 'controlled inheritance' means that a clear choice is made in relation to each type of object, as to which properties are inherited by the type in question.

The following inheritance is defined. The name used by Eye is given in brackets.

- AV object (copy) *abstract*
    - Analogue AV object (analogue copy)
    - Digital AV object (digital copy) *abstract*
        - Rendition
          A single playable file
        - CPL(Composition Playlist)
        - DCP (Digital Cinema Package)
          A DCP includes one or more CPLs
          A DCP consists of a number of directories in a predetermined directory structure
        - Image sequence
          An image sequence consists of one image file per frame, (optionally) distributed over a number of directories for each reel (also known as an 'act'). An image sequence may also be associated with a number of sound mixes and one or more subtitles.
          A DCDM (Digital Cinema Distribution Master) is a specific type of image sequence.
        - Additional digital AV material
          This can be all sorts of different things: separate sound mixes for specific purposes; separate subtitle files for specific purposes; the digital representation of an AV exhibition; working materials, etc.

For example, all AV objects have a title, which means that the title is defined at the level of the AV object. All objects derived from the AV object explicitly inherit this field. Only a CPL has a CPL name, which means that the latter is defined at CPL level. A codec is defined at the level of the digital AV object: in other words, all digital AV objects have a codec.
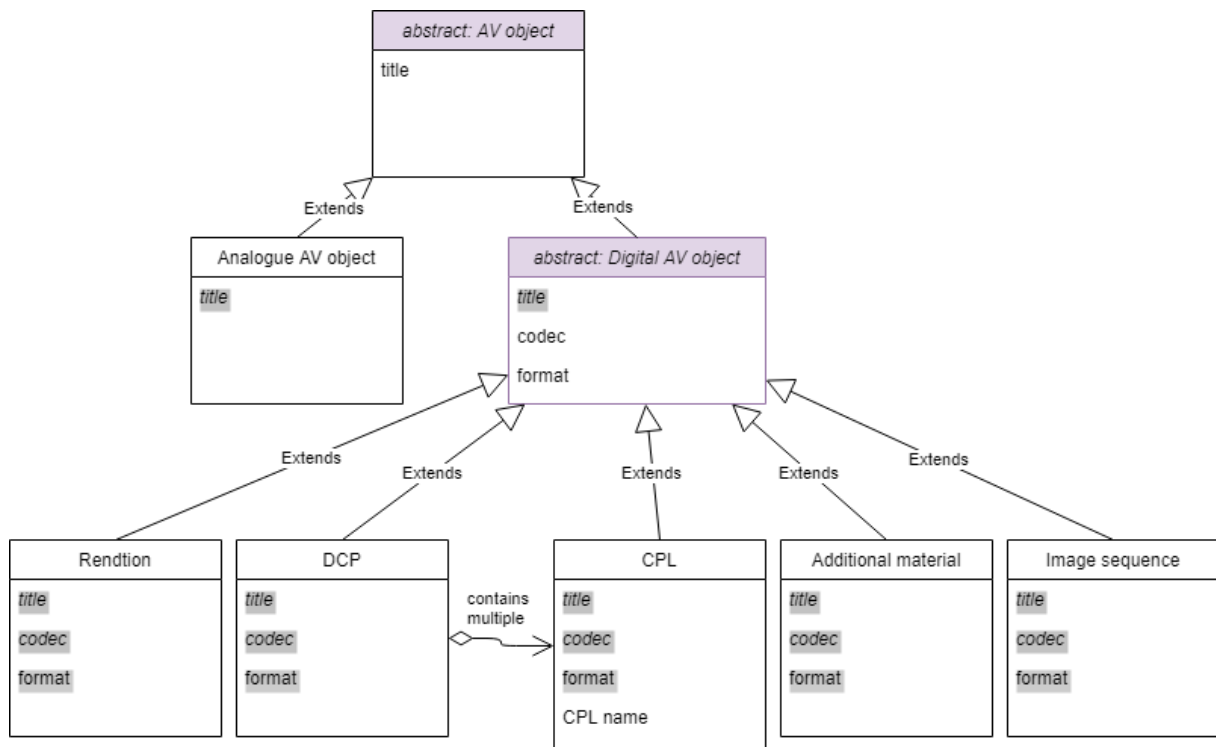
Figure 1: The inheritance of an AV object. *The AV object and digital AV object are abstract types. These are merely definitions of fields, whereas records of renditions, CPL, etc. can actually be defined. The fields shaded in grey are inherited fields.*

A CPL is a selection of component parts from either one DCP or (very occasionally) several DCPs. For practical reasons, we decided to model a CPL in the same way as a DCP, i.e. as a separate digital AV object. In practice, we often see that the DCP is encrypted, and that only the CPL is known. It is the CPL that is played and not the DCP. Although it is not advisable to use an encrypted DCP as an AIP, the model nonetheless needs to cater to this situation as it does occur from time to time.

4.1.2 Repeatable field blocks

Digital AV objects are complex objects consisting of a number of different component parts. These can include subtitles in several languages and a number of sound systems that all form part of one and the same AV object. This means that the representation of a digital AV object is equally complex.

We used repeatable blocks of fields to model this. In other words, we defined 36 fields to represent a single sound system of an AV object. More than one block of 36 fields can be defined for the same AV object. Obviously, we do not always need all 36 fields: in many cases, many of these fields are not present.

To illustrate the complexity and use of repeatable field blocks, below follows an example of a partial representation of a DCDM containing a number of different subtitles and sound systems:

Field block: Subtitle1
- Language: Dutch
- Type: switchable (on/off)
- Format: .xml
- Frame rate (fps): 25

Field block: Subtitle2
- Language: English
- Type: switchable (on/off)
- Format: .xml
- Frame rate (fps): 25

Field block: Sound1
- Sound system: Dolby 5.1
- Soundtrack language: French
- Commentary language: Dutch
- Frame rate (fps): 25
- Codec: PCM

Field block: Sound2
- Sound system: LtRt
- Soundtrack language: French
- Commentary language: Dutch
- Frame rate (fps): 25
- Codec: PCM

The following blocks are defined:

- Sound
  Used for describing a single sound variant. Variants may differ in terms of language, but also in terms of the sound system, frame rate (fps) or purpose (i.e. DVD, TV, cinema, etc.) of the sound mix.
- Subtitle
  Used for describing a single subtitle variant. Variants may differ in terms of language, but also in terms of the frame rate (fps), as subtitles often contain time codes.
- Image sequence field block
  Image sequences are often divided into reels or acts. An image sequence field block is used to represent a reel.
- Carrier info
  A digital AV object can migrate from one carrier to another several times during its lifetime. The 'Carrier info' field block is used to represent this. The field block states the date from which (and optionally, the date until which) a digital AV object was stored on a given carrier (or, optionally, is still stored). The 'Carrier info' field block must contain sufficient information to enable the AV object to be located on the carrier.
- Colour specification
  This field block is not repeatable. It is used for grouping (and keeping) together fields relating to the colour specification.

The appendix contains a list of all the fields in the model, together with their OAIS categories.
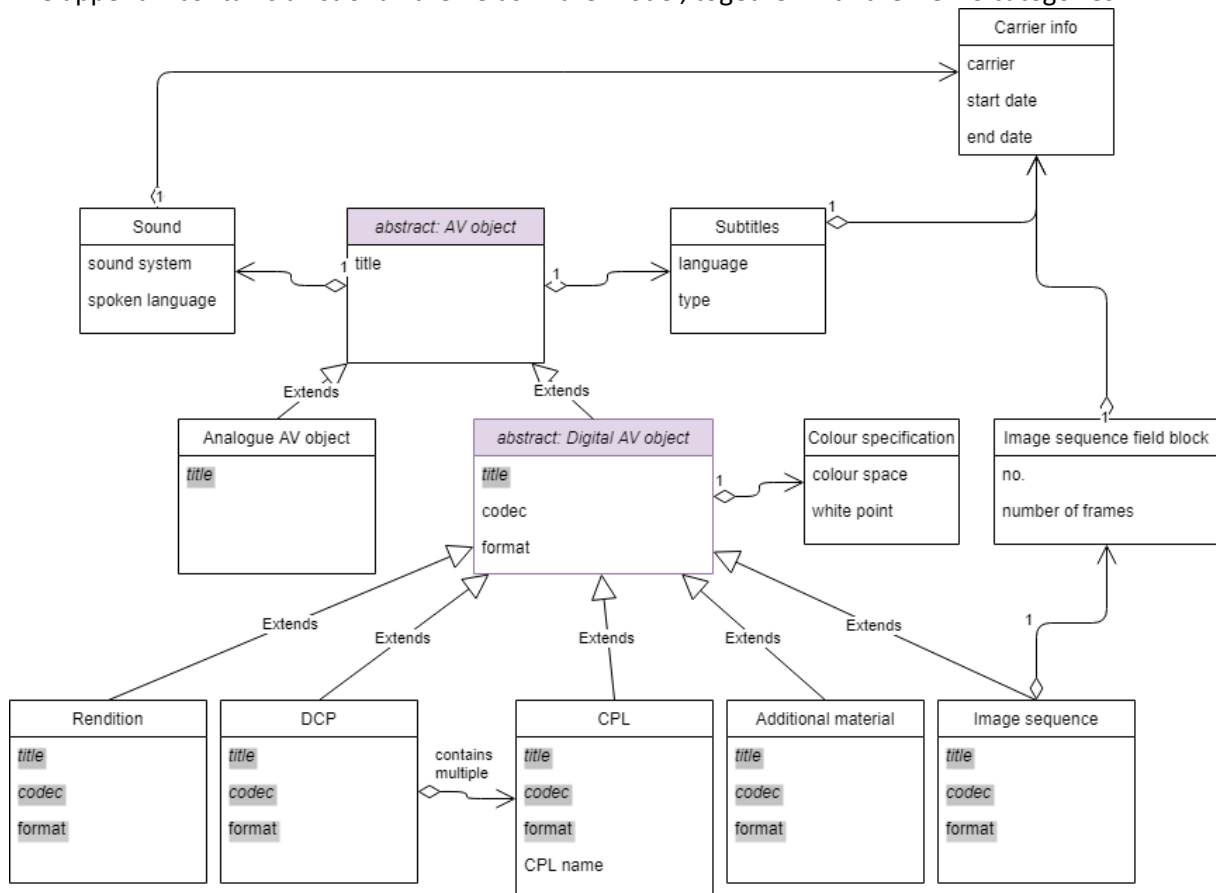
Figure 2: The structure of an AV object. *An AV object contains sound field blocks and subtitle field blocks. By virtue of the principle of inheritance, all objects derived from the AV object contain the same field blocks. The figure also shows that the sound and subtitle field blocks have their own Carrier info field blocks.*

4.2 Designated community

The designated community for the model consists of professionals who are familiar with digital AV material. This means that they need to be familiar with the meaning of terms such as codec, format, DCP and CPL, as well as with the software used for playing (and, optionally, processing) AV material.

This community may be subdivided into a number of groups:

- users who wish to play the material;
- users who wish to sell a copy of the material;
- users who wish to restore the material (in the case of digitised material);
- users who wish to preserve the material for the future.

4.3 Content information

The content information consists of the data object and the representation information. The data object can comprise one or more files stored in a specific directory structure. The representation information is intended to give the members of the designated community sufficient information for them to be able to interpret the data object. It consists of both structural and semantic information.

4.3.1 Data object

There are four types of data object:

1 **Rendition**
   A single playable file.

2 **Image sequence**
   A DCDM is a specific type of image sequence used in the model. An image sequence consists of one image file per frame. Although these image files may be stored in the same directory, the image sequence may also be divided over a number of reels (also referred to as 'acts'). In the latter case, each reel has its own directory containing the image files for that reel. An image sequence may also be associated with sound mixes and one or more subtitles.

3 **DCP**
   A DCP consists of one or more **CPL**s. A DCP has a specific directory structure.

4 **Additional digital AV material**
   This can be all sorts of different things: separate sound mixes for specific purposes; separate subtitle files for specific purposes; the digital representation of an AV exhibition; working materials, etc.

DVDs, Blu-rays, Digibetas and other digital videotapes are not described as digital AV objects, but rather as analogue AV objects, as the carrier and the content are inextricably linked to each other, as is the case with an analogue AV object.

4.3.2 Structural representation information

The first level of representation information is structural information, i.e. the translation of the 'bit stream' of a data object into discrete files in a practical structure. This information is managed in the digital archive, thus ensuring that the bits in storage are converted into files in directories.

The digital archive includes a specific feature for image sequences, to circumvent the performance problems caused by storing and retrieving large numbers of files. The image files for each reel or act are stored in uncompressed form in a TAR (tape archive) file. By doing this for each individual reel or act, users can retrieve specific parts of image sequences (rather than whole sequences). If there is no longer any room available on the tape, a reel can be subdivided into a number of TARs. This is something that the system does automatically. When a request is received for the subdivided reel, both tapes are retrieved and the reel is fully restored.

The same approach is used for Additional digital AV material. By making use of TAR files, it is possible to restore the entire directory structure in which the material is stored in the digital archive.

4.3.3 Semantic representation information

The purpose of the semantic representation information is to enable the data object to be interpreted, so that the designated community can use it. The information does not need to be very extensive: simply the type of data object and the format may well be enough information for a member of the designated community. This is because AV data objects themselves contain large amounts of metadata, which the software can read (see the next section).

More information is needed, however, on certain types of data objects. In the case of an image sequence (i.e. a format in which an image is stored for each frame in a specific format such as DPX or TIFF), the software cannot be used to generate a reliable reading of the frame rate (fps), even though the latter is needed in order to correctly interpret the data object. For the purpose of this article, the frame rate (fps) is treated as a form of representation information.

By definition, therefore, the semantic representation information is information that cannot be derived from the data object itself. The following fields are the main types of semantic representation information fields. The letters (cl) in brackets mean that a controlled list is used. A full list of fields is given in the appendix.

Field	Description
Image/Sound (cl)	States whether the AV object contains image only, image and sound, or sound only. This field is needed because it is not possible to decide on the basis of the technical metadata alone what exactly the object contains. For example, a rendition may come with an audiotrack without there actually being any sound.
Type (cl)	Specifies the type of data object. It may state, for example, that the data object is an encrypted DCP.
Format (cl) and Format version	The format of the container, generally expressed as an extension (.mov, .mp4, .wav). There is no obligation to specify the version of the format.
Frame rate (fps) (cl)	Frame rate for the AV object.
Codec (Image) (cl) and Codec ID (Image)	States how the AV object is coded (for example, Prores and 4444, with the following codec: ID ap4h). A single digital AV object may come with more than one codec.
Type of image sequence field block (cl)	In the case of an image sequence, an image sequence field block is of a particular type. It is generally an act, but it could also be a leader or an insert.
Sound	Sound is a repeatable field block because a digital AV object can contain more than one sound mix.
Sound.Format (cl)	The format of the sound container, generally expressed in the form of an extension.
Sound.Sound system (cl)	The sound system, e.g. Dolby 5.1 or LtRt. It is not always possible to generate this information automatically, as each channel may consist of a separate WAV file. If there are two WAV files, it may not be clear whether these are for LtRt or Stereo. A single sound mix may contain more than one sound system.

Field	Description
Sound.Frame rate (fps) (cl)	The frame rate for which the sound mix is intended.
Sound.Codec (cl) and Sound.Codec ID	States how the sound of the AV object is coded. This is usually PCM. In some cases, there is a separate codec for each audio channel, in which case the codec names are separated by a slash (/).
Subtitles	Subtitles are a repeatable field block as a digital AV object may contain more than one set of subtitles.
Subtitles.Type subtitles	States the type of subtitles used. For example, they may be burned-in subtitles if they are embedded in the image (the term is borrowed from analogue material) or switchable (on/off) subtitles if there is an option to choose subtitling when the object is played. Alternatively, there may be a separate subtitle file.
Subtitles.Frame rate (fps)	States the frame rate at which the subtitles are intended to be played.

4.3.4 Software

A specific type of software is needed in order to use the data object. This software may also be regarded as forming part of the representation information, although this is really a case of overspecification as far as the designated community is concerned. The members of the community know exactly what type of software they wish (and are able) to use.

We use the followed software at Eye:

- DaVinci Resolve
- Scratch (Assimilate)
- EasyDCP Player
- EasyDCP Creator
- Quicktime
- VLC
- Adobe Premier Pro
- Adobe Media Encoder
- MPEG Streamclip
- JES Deintelacer

4.4 Descriptive information

The purpose of the descriptive information is to enable the designated community to access the data object in the catalogue. Many of the fields are copies of the technical metadata for the data object in question. This enables the members of the designated community to decide what they wish to do with the data object without having to consult the object itself. As a further consideration, it must be possible to find the data object in the catalogue. This means that the 'cinematographic work' component of the AV object is essential: it is the description of the contents of the data object.

The descriptive information also contains all other information with the exception of the checksums that are part of the fixity information, as users must be able to use this information to retrieve the data object.

The following fields are the main descriptive information fields. The letters (cl) in brackets mean that a controlled list is used. A full list of fields is given in the appendix.

Field	Description
Cinematographic work	Contains a link to the cinematographic work in this AV object. If it refers to a compilation of works, the compilation must be recorded as a cinematographic work.
Title of copy	The title of the AV object. This is generally the same as the title of the cinematographic work, but sometimes the title of the AV object is not in the same language as the title of the cinematographic work.
Availability	Calculated field showing whether the AV object is currently available. Takes loans into account (in the case of analogue material).
Suitability for screening	Calculated field showing whether the AV object is suitable for screening. There are three possible values: 1 Suitable 2 Limited suitability (with explanation) 3 Not suitable (with explanation)
Colour aspect (cl)	Refers to the colour as perceived by the viewer. An AV object may be in 24-bit colour, even though what the viewer sees is a black-and-white film. This is a repeated field: obviously, a film can contain scenes in black-and-white as well as scenes in colour.
Image ratio (cl)	The image ratio of the image on the screen. This is an important field on account of digital AV objects containing letterbox, pillarbox or windowbox image.
Language of credits (cl)	Repeated fields with language information.
Language of titles (cl)	Repeated fields with language information.
Language of intertitles (cl)	Repeated fields with language information.
Playing time (calculated) (h:m:s:f)	Calculated field for image sequences that calculates the aggregate playing time based on the total number of frames and the frame rate (fps).

Field	Description
Playing time (h:m:s:f)	Where it concerns an image sequence, this may differ from the calculated playing time as inserts etc. are not necessarily included.
Completeness (status) (cl)	Shows whether the AV object is a complete representation of the cinematographic work.
CPL - encrypted (cl)	Shows whether the CPL is in an encrypted DCP.
Projection instructions	Text field containing information for projection of the CPL or rendition.
Sound.Function, use (cl)	Shows the type of sound mix, e.g. Endmix or Dialogue/voiceover.
Sound.Purpose (cl)	Indicates whether the sound is intended for TV, cinema, DVD, etc.
Sound.Commentary language (cl)	Shows the language of the commentary (if the sound mix contains a commentary).
Sound.Soundtrack language (cl)	Shows the languages used in the soundtrack (if the sound mix contains a soundtrack).
Sound.Dubbing language (cl)	Shows the languages in which the film is dubbed (if the sound mix contains dubbing).
Subtitles.Subtitle language (cl)	Shows the languages (Subtitles is a repeated field block!) in which subtitles are available.

4.5 Reference information

The reference information consists of unique identifiers (IDs) for the content information. These IDs are included both in the digital archive and in the catalogue, and are also part of the descriptive information.

Three repeatable field blocks, i.e. Sound, Subtitles and Image Sequence, also have their own IDs. This makes it easy to identify each separate field block, and to retrieve it from the digital archive as a separate Dissemination Information Package (DIP).

Field	Description
ID	
Player url	For renditions, this link can be used to view the preview in the digital archive.
Carrier info	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.
Carrier info.Carrier	Refers to the carrier described in the catalogue. A carrier can be an LTO (Liner Tape-Open) tape or a hard disk. The digital archive is also regarded as a carrier.
Carrier info.Start date	The date since when the digital AV object has been stored on the carrier.
Carrier info.End date	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).
Carrier info.In use yes/no (cl)	Shows whether the digital AV object is still stored on the carrier.
Carrier info.Link	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).
Carrier info.Object ID used by external carrier	The identifying characteristic under which the digital AV object is stored on an external carrier.
Object ID of image sequence field block	
Object ID of sound	
Object ID of subtitles	
Previous numbers	Field used for storing old IDs that are no longer is use. In the case of Eye, for example, this would state the ID previously used for recording the data object in the catalogue, in the days when digital AV objects were recorded in the same way as analogue AV objects.

4.6 Provenance information

The provenance information documents the history of the content information. In other words, it describes the changes in both the digital AV object and the representation information.

The Carrier info described in the reference information reflects both the way in which the digital AV object is currently stored and its storage history. Three repeatable field blocks, i.e. Sound, Subtitles and Image Sequence, come with their own information on storage and storage history. This makes it easy to identify each separate field block, and to retrieve it from the digital archive as a separate Dissemination Information Package (DIP).

The catalogue used by Eye always keeps a record of previous versions. This means that the model does not need to have a separate feature for recording the provenance of the representation information.

The main provenance information fields are as follows:

Field	Description
Carrier info	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.
Carrier info.Carrier	Refers to the carrier described in the catalogue. A carrier can be an LTO tape or a hard disk. The digital archive is also regarded as a carrier.
Carrier info.Start date	The date since when the digital AV object has been stored on the carrier.
Carrier info.End date	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).
Carrier info.In use yes/no (cl)	Shows whether the digital AV object is still stored on the carrier.
Carrier info.Link	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).
Carrier info.Object ID used by external carrier	The identifying characteristic under which the digital AV object is stored on an external carrier.
Reason for alteration	Field containing textual information on an alteration affecting a batch.
Cropped/uncropped (cl)	Cropping is an image processing technique used on the digital AV object before it is ingested.
Historical migration data	Field containing information on carrier migration before the carrier info field block was used.

4.7 Context information

The context information describes how the material was acquired (or produced). In the catalogue, an 'acquisition' describes how the material was acquired, i.e. whether as a gift or a purchase, whether it has been digitised and how the digitisation was funded.

If a digital AV object is self-produced, conservation links are used to keep a record of its genealogy.

The main context information fields are as follows:

Field	Description
Acquisition	Link to the acquisition of which this digital AV object formed part.
Deaccession destination (Corporation)	If the digital AV has been disposed of, this field contains a link to the corporate entity that now owns the object.
Deaccession destination (Person)	If the digital AV has been disposed of, this field contains a link to the person who now owns the object.
Born-digital / digitised (image) (cl)	States whether the image in the digital AV object was born-digital or digitised.
Born-digital / digitised (sound) (cl)	States whether the sound in the digital AV object was born-digital or digitised.
Copy of	The source of the digital AV object if it was derived from another AV object. The genealogy of digitised AV objects is generally as follows: analogue copy --> image sequence -> rendition. The sound may come from a different analogue copy from the image. The same applies when the rendition is made.
Original of	Shows which digital AV objects have been derived from this digital AV object.
CPL	The link between the way in which the DCP is recorded and the way in which the CPLs are recorded in the DCP.
DCP	Contains a link to the DCP of which this CPL forms part.

4.8 Fixity information

The fixity information documents the integrity of the data object. Checksums for the data objects are stored in the digital archive. In the case of data objects stored in the form of TARs, the checksum for the TAR file is computed and stored.

A record is also made of any frames missing from image sequence field blocks and frames that are out of sequence in the catalogue, so as to make sure that this problem does not crop up every time they are used.

4.9 Access rights information

The access rights information makes clear what may and may not be done with the digital AV object. In part, this is acquisition information, as it relates to the contractual terms on which the gift or purchase was made, as the case may be.

The access rights information also specifies whether the cinematographic work is in the public domain or is under copyright or whether it is an orphan work. This also affects the use that may be made of the material in question.

The access rights information fields are as follows:

Field	Description
Acquisition	Link to the acquisition of which this digital AV object formed part.
Archive restriction for screening (cl)	This is an internal restriction, indicating for example that a better version is available.
Copy restriction (cl)	This is an external restriction, i.e. may the material be used, either with or without the rights-holder's consent? Should you take a good look at the contractual terms before using the material?
Rights arrangements	Where a special arrangement has been made with the rights-holders for the material to be used for a specific purpose, this information is set out in a document known as a 'rights arrangement'. This is then linked to the material to which it applies.

5 CHOICES MADE IN DEVELOPING THE MODEL

During the process of designing the model, certain choices were made, in relation to both the structure of the model and the fields incorporated in it.

5.1 Structure

The structure for the model was chosen in accordance with the various types of digital AV objects. The range of types was based on shared properties, which led to a division of object types into two basic categories: playable digital AV objects (rendition) and digital AV objects used for conservation (image sequence). Once this division had been made, two further categories remained: DCPs, which have a number of highly specific properties, and Additional material (given that an extremely wide range of digital material is produced during the production of a film, and the model also needed to cater to this).

5.2 Fields

We looked at the following aspects in order to decide which fields to include in the model:

- fields used with analogue AV objects and which are also relevant to digital AV objects;
- the technical metadata already available that can be determined by automatic means (using the MediaInfo library);
- the fields needed for the various work flows involving digital AV objects;
- the fields required for a sustainable form of digital storage.

Obviously, there is a degree of overlap. The format, for example, is usually an aspect that can be determined by automatic means, but it is also needed for a number of work flows and for sustainable storage (if a given format is no longer used, material in that format must be transcoded. This is known as a 'transformation' in OAIS terms).

6 PROCESS: INGEST FUNCTION

6.1 Submission Information Packages

Digital AV material that needs to be ingested arrives in all sorts of different formats and structures. It may be stored on hard disk or on LTO tape, and very occasionally it arrives in the form of a download. In other words, the Information Packages (IPs) are not standardised, unless we have digitised the material in question ourselves. Submission Information Packages need to be made out of the material before it can be ingested.

Based on the IPs that have been submitted, the digital AV objects are manually renamed and stored in a predefined directory structure. The directory structure contains a number of SIPs for each cinematographic work; this is because a work may consist of a range of renditions, a DCP, sound mixes, etc. The files and directories are named in such a way that all the files in the ingested SIPs of a cinematographic work have a unique name. The name contains additional information, so that, once the film has been ingested, the various AIPs have been more or less fully recorded.

6.2 Structure of Submission Information Packages

The directory structure of the files in a single cinematographic work is as follows:

- ➢ **WORKFLOWCODE**_FREETEXT {_**SOURCEID**{_**SOURCEID**}
  - ✳ Film
    - ✦ Image sequence
      - ⌑ SEQ{**NR**}_**COLOURSPACE_COLOURGAMUT_WHITEPOINT_FPS**
        - • {**ACTTYPE_**}R**NO**
    - ✦ DCDM_**COLOURGAMUT_WHITEPOINT_FPS**
      - ⌑ IMAGE
        - • {**ACTTYPE_**}R**NO**
      - ⌑ AUDIO
        - • **SOUNDSYSTEM**
          - ‣ {**ACTTYPE_**}R**NO**
      - ⌑ SUBTITLES
        - • **LANGUAGECODE**{-**LANGUAGECODE**}*
          - ‣ FREETEXT_ **LANGUAGECODE**{-**LANGUAGECODE**}*SUB_R**NO**.XML
    - ✦ Renditions
      - ⌑ FREETEXT**_CODEC_ASPECTRATIO_IMAGERATIO_COLOURGAMUT_WHITEPOINT_SOUNDSYSTEM{#LANGUAGECODE}*{- SOUND SYSTEM{#LANGUAGECODE}*}*_LANGUAGECODESUBTITLE{-LANGUAGECODESUBTITLE}*_LANGUAGECODEOPENINGCREDITS{-LANGUAGECODEOPENINGCREDITS}*_LANGUAGECODECLOSINGCREDITS{-LANGUAGECODECLOSINGCREDITS}***
    - ✦ DCP
      - ⌑ PKL-NAME
        - • Directory structure of DCP
    - ✦ Audio

- ⌑ **FPS_SOUND SYSTEM_FUNCTIONUSE_PURPOSE{_SOURCEID}**
  - • R**NO**
- ✦ Subtitles
  - • **LANGUAGECODE**{-**LANGUAGECODE**}*
    - ⁘ **FPS**
      - ▫ File name with reel number
      - ▫ {**TIFF**}
        - ⁘ tiff subtitle files
          There is a tiff directory only if there are tiff subtitle files.
- ✦ Remaining material
  - ⌑ Free directory structure
- ✹ Film-related
  - ✦ Free directory structure
- ✹ Exhibition
  - ✦ Free directory structure

The items in bold print are codes, most of which refer to controlled lists.The ISO639-3 standard is used for language codes. Items listed in brace brackets ({}) are optional. An asterisk means that the item in question may be repeated a number of times or left out.

NO is a serial number, e.g. R**NO** stands for the reel number.

FREETEXT is variable text in a file name that does not contain any information for the ingest process.

SOURCE_ID refers to the ID in the catalogue for the AIP that provided the basis for this SIP. In other words, it is used for digitised film objects (in which case the SOURCE_ID is the ID of the analogue AV object). If there are two SOURCE_IDs in the name of the directory, this means that a different AV object is the sound source for the image in question. This information is used to complete the 'Copy of' field in the Context Information.

WORKFLOWCODE can have three different values (at present):

- • BL
  backlog work flow: the project in which the former 'analogue' registration of the digital AV object is replaced by registration with the aid of the new model
- • SC
  scanity work flow: our own digitisation work flow
- • DB
  born-digital work flow: this is the work flow used to process born-digital material.

7 PROCESS: ARCHIVAL STORAGE FUNCTION

This is what the digital archive does. The digital archive works with a tape robot containing LTO tapes. The digital archive software uses LTFS (linear tape file system) so that the data can be accessed in the same way as files on a hard disk.

7.1 Data integrity

The digital archive uses various methods of guaranteeing the data integrity of the data objects during the ingest process. The digital archive at Eye stores the data objects on LTO tapes. The write head is followed by a read head, which reads and compares the bits as soon as they have been recorded.

The checksum for the data object is calculated before it is stored. Once the whole object has been stored, a restore is performed and a checksum computed for the result of the restore. This is then compared with the previous value. Only if the two checksums are the same is the object removed from temporary storage in the digital archive.

8 PROCESS: DATA MANAGEMENT FUNCTION

The catalogue has features for processing data. These can be used for completing the registration after the ingest process. Once the registration has been completed, the 'Registration status' field is then set to 'Registration completed'.

For reasons of efficiency, the catalogue contains a feature that is capable of processing in a single run a set of AIPs that have been ingested in one batch (AIPs are ingested per cinematographic work, which means that they have the same values for certain fields). The following fields can be allocated in batches:

Field	Type of information
Cinematographic work	Descriptive information
Title of copy	Descriptive information
Language: comment on language	Descriptive information
Start of opening credits (h:m:s:f)	Descriptive information
Language of titles	Descriptive information
Language of credits	Descriptive information
Language of intertitles	Descriptive information
Image ratio	Descriptive information
Colour aspect	Descriptive information
Sound.Commentary language	Descriptive information
Discrepancy in description of cinematographic work	Descriptive information
Structure (editing)	Descriptive information
Title and intertitles	Descriptive information
Description of contents	Descriptive information
Completeness (status)	Descriptive information
Completeness (summary)	Descriptive information
Born-digital / digitised (image)	Context information
Born-digital / digitised (sound)	Context information
Acquisition	Context information
Comment on acquisition/deaccession	Context information
Copy restriction	Access rights information
Rights arrangements	Access rights information

Once batch processing has been completed, the individual records of the AIPs are opened so that the user can process an AIP's individual properties.

In other words, the work flow is as follows: Receipt of material ⇨ Production of SIPs ⇨ Ingest ⇨ Batch processing of AIPs ⇨ Individual processing of AIPs.

9 PROCESS: DATA ACCESS FUNCTION

The catalogue has a search facility that allows the user to search through all the fields described in this document. The user must also be able to find an AIP on the basis of the information given on the cinematographic work. It must be possible to find the AIPs using the following fields on the work:

- Cinematographic work's ID
  Category
  Fiction/Non-fiction/Hybrid
- Genre
  GTAA genre
- Year of production
- Country of production
- Production company
- Titles
- Director
- Cast
- Crew
- Content information
  All sorts of text fields for a variety of target groups and platforms.
  These include the key words from the GTAA (Common Thesaurus for AV Archives).

The digital archive also has a small set of searchable fields, so that users can also search for the AIPs in the digital archive. These fields are synchronised with the fields in the catalogue.

It is easy to use the search result from the catalogue to open the corresponding AIPs in the digital archive. The next step is to retrieve a Dissemination Information Package (DIP) in the digital archive. Users can also retrieve part of an AIP (for example, one reel from an image sequence, or a single sound mix), provided that the part in question has its own ID. If the part is stored in the form of a TAR, the relevant TAR is unpacked by the digital archive.

The digital archive has an additional transcoding feature for renditions. A software suite called FFmpeg is used for this purpose. Transcoding profiles can be defined in the digital archive, and FFmpeg then performs the transcoding.

10 IMPLEMENTATION: POINTS TO BEAR IN MIND

The following points should be borne in mind in implementing the model.

10.1 DCP and CPL

As we have already said, a CPL (Content Play List) is a selection of items from a DCP (Digital Cinema Package). The most comprehensive way of modelling this is by modelling all the constituent parts of the DCP, and ensuring that the CPL refers to those parts that are used by the CPL. Any change that is made to the DCP is then automatically made to the CPL. The only problem is that this is something of a rarity in practice and building this feature into the model would tend to make it much more complex. This is why we decided not to take this into account.

As a further consideration, we also use these parts of the model for cinema screenings at Eye. Many of the DCPs supplied for projection are encrypted, and only the CPL that is available for use is actually known. For this reason, we decided that our model should allow a CPL to be described without a DCP. Thus, virtually all the fields relating to the CPL are defined, whereas a DCP has only a very small set of fields that are used only when the DCP is ingested.

10.2 Image sequences, sound mixes and subtitles

Formats used for sustainable archiving purposes generally use one 'photo' per frame. DPX and TIF are common formats. Although there is no point in describing these individuals files, it does make sense to describe these files as an ordered set (sequence).

The same applies to individual files in a sound mix. There is no point in describing the individual files in the mix. In certain cases, this also applies to subtitles (which may consist of a series of separate frames).

There is an additional problem in that storing and retrieving large numbers of files can easily cause performance problems at the storage facility. This is why the sequences are stored in the form of a single TAR file, which is then saved in the digital archive. This also ensures that the level of the description, i.e. the sequence, corresponds with the file that has been saved.

Sound mixes consisting of a number of sound files are also saved in the digital archive in the form of TAR files. We decided to use TAR because it is an open-source format that is free of rights.

10.3 Multiformat

Our own day-to-day needs make it impracticable for us to make use of open standards only. In certain respects, open standards are still rather immature, are not particularly well supported by software and only rarely used by film-makers and producers. One example of an open standard is digitised material with a low frame rate. Only a small number of formats are capable of handling a low frame rate.

10.4 Repeatability and inheritance

Because Eye's catalogue has no problem in defining repeatable fields, it was fairly easy for us to design the model so that it can describe renditions that have multiple formats. We therefore made extensive use of this feature. The model may be more difficult to use with systems that do not allow repeatable fields to be defined.

Our catalogue also supports controlled inheritance. If your catalogue does not support controlled inheritance, it may be worth adjusting the categorisation of object types so that the model works better with the catalogue.

The list of fields indicates the type of data object to which each field applies. No information is given on inheritance: it is up to the person implementing the model to decide whether or not such information should be provided.

10.5 External digital archive

We have our own digital archive. This has a number of advantages: if the digital archive is hosted by an external organisation, it is more difficult to create a link between the catalogue and the digital archive and to exchange information between them. This may result in an information deficit in both systems, i.e. a lack of technical metadata in the catalogue and a lack of management data in the digital archive.

The sheer size of digital AV objects means there is an added disadvantage in using an external digital archive: in many cases, it is impractical to download film material because downloading simply takes too long. This undermines the availability of the material in the digital archive.

10.6 Cinematographic work: versions, variants, trailers and teasers

Here at Eye, we had already adopted a practice, in describing analogue AV objects, of storing information on versions and variants not with the cinematographic work itself, but with the analogue AV object. In other words, information on variants in terms of language and content was stored with the object.

This is in close keeping with the practice in relation to digital AV objects. A DCP or DCDM can contain a number of variants (and perhaps even several versions) of a cinematographic work. The model described in this document makes allowance for this, for example, by including a field headed 'Discrepancy in description of cinematographic work'.

Similarly, we do not create a separate item for a trailer of teaser. Instead, we use a field that indicates whether the AV object in question is a trailer or teaser for a cinematographic work. Although this feature is not included in the model described in this document, it could be included.

11 GLOSSARY

Term	Definition
DCDM	Digital Cinema Distribution Master A complete master representation that can serve as the basis for the production of derivatives in all sorts of different formats.
DCP	Digital Cinema Package A digital AV object specifically intended for cinema screening. The DCP itself is not playable. A DCP consists of one or more CPLs.
CPL	Content Play List A recipe for playing the items in the DCP.
Image sequence	A digital AV format consisting of 1 digital image per frame. Image sequences are often divided into reels or acts. To a certain extent, this is a throwback to analogue times. At the same time, the division makes it easier for software to perform certain operations on the image sequence in question.
Inheritance	Means that certain properties inherent to a generic term (hyperonym) are included in a more specific term. For example, the terms 'ape' and 'human' are inherited from the generic term 'mammal'. 'Ape' and 'human' are more specific terms than 'mammal'. All mammals have a certain gestation period: this is a property defined at the hyperonymic level, i.e. at the level of mammals. It is a property inherited by apes and humans.
Catalogue	The system in which information on the objects in the archive is managed.
Digital archive	The system in which digital objects are stored.
KDM owner	Key Delivery Message owner The KDM is used to access a CPL in an encrypted DCP. This is often for a limited period. Information on the KDM owner is recorded so that, if necessary, a new KDM can be requested at the end of the period in question.
Sound mix	A mix of the sound for a film. A sound mix has a sound system (e.g. stereo or Dolby 5.1). A sound mix generally consists of a number of WAV files, which must be played on the right channel.

12 LIST OF FIELDS

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Context	Acquisition and deaccession	Acquisition	Type of item	Mandatory		Image sequence DCP Additional material Rendition	Link to the acquisition of which this digital AV object formed part.	
Context	Acquisition and deaccession	Deaccession destination (Corporation)	Type of item			Image sequence DCP Additional material Rendition	If the digital AV has been disposed of, this field contains a link to the corporate entity that now owns the object.	
Context	Acquisition and deaccession	Deaccession destination (Person)	Type of item			Image sequence DCP Additional material Rendition	If the digital AV has been disposed of, this field contains a link to the person who now owns the object.	
Context	Acquisition and deaccession	Date of deaccession	Date			Image sequence DCP Additional material Rendition		
Context	Acquisition and deaccession	Method of deaccession	Type of item			Image sequence DCP Additional material Rendition		Deleted, do not use: Deaccessioned, NA, Moved to new storage, Transferred, Destroyed
Access rights	Acquisition and deaccession	Archive restriction for screening	Type of item			CPL Additional material Rendition	This is an internal restriction, indicating for example that a better version is available.	Conservation source, Not a preferred copy (better copy available), Unique copy
Context	Acquisition and deaccession	Own production	Calculated	Mandatory		Image sequence DCP Additional material Rendition	Calculated field indicating whether we have produced the digital AV object ourselves.	
Context	Acquisition and deaccession	KDM owner	Type of item			CPL		
Access rights	Acquisition and deaccession	Copy restriction	Type of item			Image sequence CPL Additional material Rendition	This is an external restriction, i.e. may the material be used, either with or without the rights-holder's consent? Should you take a good look at the contract terms before using the material?	No information, Restriction: for internal use only, this information should not be communicated externally, Restriction: for internal use only, provided that the donor has given its prior consent, Restriction: see terms of exchange; ask Collection Coordinator for further information, Restriction: specific contract terms;

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
								ask Collection Coordinator for further information
Context	Acquisition and deaccession	Comment on acquisition/deaccession	Text area			Image sequence DCP Additional material Rendition		
Access rights	Acquisition and deaccession	Rights arrangements	Type of item			Image sequence CPL Additional material Rendition	Where a special arrangement has been made with the rights-holders for the material to be used for a specific purpose, this information has been set out in a document known as a 'rights arrangement'. This is then linked to the material to which it applies.	
Representation	Image	3D	Type of item			Image sequence CPL Rendition	States whether the data object must be played in 3D, and how the 3D effect is achieved.	Yes, Anaglyph, Yes, DCP, Yes, DCP, Anaglyph, Yes, DCP, Polarisation, Yes, frame packing, Yes, line by line, Yes, system unknown, Yes, Polarisation, Yes, side by side, Yes, side by side, Anaglyph, Yes, side by side, Polarisation, Yes, top and bottom, Yes, top and bottom, Anaglyph, Yes, top and bottom, Polarisation, No
Descriptive	Image	Aspect ratio	Type of item			Image sequence CPL Rendition		1,21:1 Normal image 1,25:1 (from ingest), 1,31:1 (from ingest), 1,32:1 (from ingest), 1,33:1 Full image, 1,37:1 Normal image, 1,66:1 Widescreen, 1,75:1 Widescreen, 1,78:1 Widescreen, 1,85:1 Widescreen / Flat, 1,90:1 Full Container, 2,33:1, 2,35:1 Cinemascope, 2,39:1 Scope, 2,55:1 Cinemascope, NA, Unknown
Descriptive	Image	Image compression	Type of item			Image sequence CPL Rendition		lossless, lossy

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Image	File size for all image sequence field blocks	Compound			Image sequence	This is a calculated field that adds up all the image sequence field blocks. The file size always consists of a value and a unit.	
Descriptive		File size for all image sequence field blocks.Unit	Type of item			Image sequence		GB, KB, MB, TB
Descriptive		File size for all image sequence field blocks.Size	Integer			Image sequence		
Descriptive	Image	Bit depth of image	Text			Image sequence CPL Rendition		
Descriptive	Image	ChromaSubsampling (422/4444/etc)	Text			Rendition		
Representation	Image	Codec (Image)	Type of item			Image sequence CPL Rendition	States how the AV object is coded (for example, Prores and 4444, with the following codec: ID ap4h). A single digital AV object may come with more than one codec.	0D01030102060100, 0D01030102060100-0000000000000000, 422, 422 HQ, 422 LT, 4:2:2@Main, 4444, AAC, AVC, Baseline@L3.1, D-Cinema 2k, DNxHD 175, DNxHD 175x, DNxHD 185, DPX, DV, IMX 30, IMX 50, JPEG, JPEG 2000, LC, Main@L4.1, Main@L4.2, Main@L5, No restrictions, PCM, ProRes, Raw, SDPX, Test newlistCodec, VC-3, YUV
Representation	Image	Codec ID (Image)	Text			Image sequence Rendition	See Codec	
Context	Image	Date of creation	Date			Image sequence	Date on which the image sequence was created	
Provenance	Image	Cropped/uncropped	Type of item			Image sequence Rendition	Cropping is an image processing technique used on the digital AV object before it is ingested.	cropped, uncropped
Descriptive	Image	Image ratio	Type of item			Image sequence CPL Rendition	The image ratio of the image on the screen. This is an important field on account of digital AV objects containing letterbox or pillarbox image.	1,21:1 Normal image, 1,25:1 (from ingest), 1,31:1 (from ingest), 1,32:1 (from ingest), 1,33:1 Full image, 1,37:1 Normal image, 1,66:1 Widescreen, 1,75:1 Widescreen, 1,78:1 Widescreen, 1,85:1 Widescreen / Flat, 1,90:1 Full Container, 2,33:1, 2,35:1 Cinemascope, 2,39:1 Scope, 2,55:1 Cinemascope, NA, Unknown

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Image	Colour aspect	Type of item	Mandatory	repeated	Image sequence CPL Rendition	Refers to the colour as perceived by the viewer. An AV object may be in 24-bit colour, even though what the viewer sees is a black-and-white film. This is a repeated field: obviously, a film can contain scenes in black-and-white as well as scenes in colour.	Coloured-in, Coloured-in (manual), Colour, NA, Unknown, Stencilling, Tinting, Toning, Black-and-white
Representation	Image	Colour specification	Compound			Image sequence CPL Rendition	Specifies aspects such as White point (cl), Log/lin, Grading (cl), Colourspace and Colour Gamut. In some (but not all) cases, this information may be derived from the data object itself.	
Representation	Image	Colour specification.Colour gamut	Type of item			Image sequence CPL Rendition		ACES, LOG, Unknown, P3, REC2020, Rec709, RGB, XYZ
Representation	Image	Colour specification.Colourspace	Text			Image sequence CPL Rendition		
Representation	Image	Colour specification.Graded/ungraded	Type of item			Image sequence CPL Rendition		graded, ungraded
Representation	Image	Colour specification.Log/lin	Text			Image sequence CPL Rendition		
Representation	Image	Colour specification.Comment on colour specification	Text			Image sequence CPL Rendition		
Representation	Image	Colour specification.White point	Type of item			Image sequence CPL Rendition		Other, D50, D55, D60, D65, DCI-P3, None, Unknown
Descriptive	Image	Resolution (width x height)	Compound			Image sequence CPL Rendition		
Descriptive		Resolution (width x height).Resolution: width	Integer			Image sequence CPL Rendition		
Descriptive		Resolution (width x height).Resolution: height	Integer			Image sequence CPL Rendition		
Descriptive	Image	Resolution (xK)	Type of item			Image sequence CPL Rendition		2K, 4K, 6K, 8K, HD, HDReady, SD, ULTRA HD
Descriptive	Image	Scan type	Type of item			Image sequence CPL Rendition		interlaced, progressive
Descriptive	Image	Scan sequence	Text			Rendition		
Descriptive	Image	Language of credits	Type of item		Repeated	Image sequence CPL Rendition		
Descriptive	Image	Language of titles	Type of item		Repeated	Image sequence CPL Rendition		

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Image	Language of intertitles	Type of item		Repeated	Image sequence CPL Rendition		
Descriptive	Image sequence field block	Number of frames	Integer			Image sequence		
Descriptive	Image sequence field block	Act number	Integer	Mandatory		Image sequence		
Descriptive	Image sequence field block	File size of image sequence field block	Compound			Image sequence	States the sum total of the sizes of all individual frame files. The file size always consists of a value and a unit.	
Descriptive	Image sequence field block	File size of image sequence field block.Unit	Type of item			Image sequence		GB, KB, MB, TB
Descriptive	Image sequence field block	File size of image sequence field block.Size	Integer			Image sequence		
Descriptive	Image sequence field block	.cin/.dpx/etc sequence first file	Text			Image sequence	File name of first file in the sequence	
Descriptive	Image sequence field block	.cin/.dpx/etc sequence last file	text			Image sequence	File name of last file in the sequence	
Provenance	Image sequence field block	Carrier info	Compound		Repeated	Image sequence	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.	
Provenance	Image sequence field block	Carrier info.Start date	Date			Image sequence CPL Additional material Rendition	The date since when the digital AV object has been stored on the carrier.	
Provenance	Image sequence field block	Carrier info.Carrier	Type of item			Image sequence CPL Additional material Rendition	A carrier can be an LTO (Liner Tape-Open) tape or a hard disk. The digital archive is also regarded as a carrier.	
Provenance	Image sequence field block	Carrier info.End date	Date			Image sequence CPL Additional material Rendition	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).	
Provenance	Image sequence field block	Carrier info.In use yes/no	Type of item			Image sequence CPL Additional material Rendition	Shows whether the digital AV object is still stored on the carrier.	Yes, No, NA, Unknown

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Provenance	Image sequence field block	Carrier info.Link	Url			Image sequence CPL Additional material Rendition	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).	
Provenance	Image sequence field block	Carrier info.Object ID used by external carrier	Text			Image sequence CPL Additional material Rendition	The identifying characteristics under which the digital AV object is stored on an external carrier.	
Fixity	Image sequence field block	missing .cin/.dpx/etc frame number	Text			Image sequence		
Reference	Image sequence field block	Object ID of image sequence field block	Sequence	Mandatory		Image sequence		
Descriptive	Image sequence field block	Comment on image sequence field block	Text			Image sequence		
Fixity	Image sequence field block	Out of sequence files	Text			Image sequence		
Descriptive	Image sequence field block	Playing time (calculated) (h:m:s:f)	Time			Image sequence	Calculated playing time of the image sequence field block, based on the frame rate (fps) and the number of frames.	
Representation	Image sequence field block	Type of image sequence field block	Type of item	Mandatory		Image sequence	In the case of an image sequence, an image sequence field block is of a particular type. It is generally an act, but it could also be a leader or an insert.	Act, Generic, Insert, Leader, Additional material
Reference	Management	Previous numbers (Eye)	text			Image sequence DCP Additional material Rendition	Field used for storing old IDs that are no longer in use. In the case of Eye, for example, this would state the ID previously used for recording the data object in the catalogue, in the days when digital AV objects were recorded in the same way as analogue AV objects.	
Descriptive	Condition	Condition report	Type of item		Repeated	Image sequence DCP Additional material Rendition	Link to a report describing the condition of the AV object.	
Descriptive	Condition	Projection instructions	Text area			CPL Rendition	Text field containing information for projection of the CPL or rendition	
Descriptive	Condition	Condition of copy	Text area			Image sequence CPL Additional material Rendition	Text field summarising the condition of the AV object.	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Context	Conservation links	Born-digital / digitised (image)	Type of item			Image sequence CPL Additional material Rendition	States whether the image in the digital AV object was born-digital or digitised.	Born-digital, Digitised, Unknown
Context	Conservation links	Born-digital / digitised (sound)	Type of item			Image sequence CPL Additional material Rendition	States whether the sound in the digital AV object was born-digital or digitised.	Born-digital, Digitised, Unknown
Context	Conservation links	Copy of	Type of item		Repeated	Image sequence CPL Additional material Rendition	The source of the digital AV object if it was derived from another AV object. The genealogy of digitised AV objects is generally as follows: analogue copy --> image sequence -> rendition. The sound may come from a different analogue copy from the image. The same applies when the rendition is made.	
Context	Conservation links	Original of	Type of item		Repeated	Image sequence CPL Additional material Rendition	Shows which digital AV objects have been derived from this digital AV object.	
Descriptive	Sound	Number of acts (sound)	Integer			Image sequence Additional material	A sound field block contains information on the sound mix of the entire digital AV object, i.e. there is no sound field block for each act.	
Descriptive	Sound	Number of audio channels	Integer			Image sequence CPL Additional material Rendition		
Descriptive	Sound	File size (sound)	Compound			Image sequence Additional material Rendition	States the sum total of the sizes of all individual sound files. The file size always consists of a value and a unit.	
Descriptive	Sound	Sound file size.Unit	Type of item			Image sequence Additional material Rendition		GB, KB, MB, TB
Descriptive	Sound	Sound file size.Size	Integer			Image sequence Additional material Rendition		
Representation	Sound	Bit depth of sound	Text			Image sequence CPL Additional material Rendition		

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Sound	Codec	Type of item			Image sequence CPL Additional material Rendition	States how the sound of the AV object is coded. This is usually PCM. In some cases, there is a separate codec for each audio channel, in which case, the codec names are separated by a slash (/).	AAC, PCM, PCM / PCM, PCM / PCM / PCM, PCM / PCM / PCM / PCM
Representation	Sound	Codec ID	Text			Image sequence Additional material Rendition	See Codec.	
Descriptive	Sound	Completeness (status)	Type of item			Image sequence CPL Additional material Rendition		Complete, Virtually complete, Fragment, Incomplete, Short at act transitions, Unknown
Context	Sound	Date of creation	Date			Image sequence Additional material		
Descriptive	Sound	Purpose	Type of item			Image sequence CPL Additional material Rendition		airline, cinema, Blu-ray, DVD, internet, Unknown, TV
Provenance	Sound	Carrier info	Compound		Repeated	Image sequence Additional material	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.	
Provenance	Sound	Carrier info.Start date	date			Image sequence CPL Additional material Rendition	The date since when the digital AV object has been stored on the carrier.	
Provenance	Sound	Carrier info.Carrier	Type of item			Image sequence CPL Additional material Rendition	A carrier can be an LTO (Liner Tape-Open) tape or a hard disk. The digital archive is also regarded as a carrier.	
Provenance	Sound	Carrier info.End date	Date			Image sequence CPL Additional material Rendition	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).	
Provenance	Sound	Carrier info.In use yes/no	Type of item			Image sequence CPL Additional material Rendition	Shows whether the digital AV object is still stored on the carrier.	Yes, No, NA, Unknown
Provenance	Sound	Carrier info.Link	Url			Image sequence CPL Additional material Rendition	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Provenance	Sound	Carrier info.Object ID used by external carrier	Text			Image sequence CPL Additional material Rendition	The identifying characteristic under which the digital AV object is stored on an external carrier.	
Representation	Sound	Encoding software	text			Image sequence Additional material Rendition	This may contain extra information o the contents of the file, particularly where errors have been found in certain software. We used to have files that no longer passed our quality control procedure because they had been made with an old version of the software.	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Sound	Format	Type of item	Mandatory		Image sequence CPL Additional material	The format of the container, generally expressed as an extension.	½ inch VK30, 1/2 inch, 13x8cm, 1/4 inch video, 16mm, 17,5mm, 18x24mm, 1i. B aka BCN, 1 i. C (AMPEX&SONY), 22mm, 28mm, 30mm, 30mm (Pathéorama), 35mm, 3,5x5inch, 44,5mm, 4,5x6cm, 5x5cm, 6,25mm, 63mm, 65mm, 68mm, 6x6cm, 70mm, 8mm, 8x8cm, 9,5mm, AC3 (.ac3), AIFF (.aif), Apple uncomp. (.mov), BetaLP, Betamax, BetaSP, Blu-ray, CD, Cineon (.cin), Compact cassette, D10-30 (.mxf), D10-50 (.mxf), D15-50 (.mxf), Dat, DCP - encrypted, DCP non-encrypted, Digibeta, Digital video, Digital video D2, Digital video D3, Digital video D4, DNxHD (.mov), .dpx, DPX (.dpx), DS8, Double 8, DVCAM, DVD, ED-Beta, Flash (.flv), H264 (.mov), HDCAM, Hi-8, Jpeg2K (.jp2), Jpeg2K (.mov), Jpeg2K (.mxf), Laserdisc, Mini-DV, .mov, .mp4, Mpeg1 (.mpg), Mpeg2 (.mpg), Mpeg2 (.mxf), Mpeg4 (.mp4), .mxf, NVT, Unknown, other file format, Prores 422 (.mov), Prores 4444 (.mov), S16mm, S8mm, Single 8, SuperBetamax, S-VHS, S-VHS-C, .tif, .tiff, TIFF (.tif), U-matic, unknown file format, VCR (Philips), VHS-HQ, VHS-LP, VHS-SP, Video, Video 2000, Video8, .wav, Wav (.wav), .xml
Descriptive	Sound	Function, use	Type of item	Mandatory		Image sequence CPL Additional material Rendition		Audio description, Dialogue music and effects, Dialogue / voiceover, Final mix, Digitised sound, Music and effect, Music track, Unknown, clear, Printing Master, Sound stems
Representation	Sound	Audio channels used	Text			Image sequence CPL Additional material Rendition	States which audio channels have actually been used (and must therefore be played).	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Sound	Sound system	Type of item		Repeated	Image sequence CPL Additional material Rendition	The sound system, e.g. Dolby 5.1 or LtRt. It is not always possible to generate this information automatically, as each channel may have a WAV file of its own. If there are two WAV files, it may not be clear whether these are for LtRt or Stereo. A single sound mix can contain more than one sound system.	4-channel, 5.1, 5.1 Merged, 5.1 Remaster, 5.1 Remaster merged, 6-channel, 6-channel Dolby, 6-channel Dolby splitsurround, 7.1, Bilateral, Cyan track, Decoded, Decoded and Printing Master, Decoded SR, Dolby, Dolby A, Dolby Atmos, Dolby digilanguage, Dolby Prologic, Dolby Prologic 2, Dolby SR, Double bilateral, DTS, LtRt, Mono, Mono with Academy filter, Multi bilateral, Unknown, Push-pull, RCA, Rivatone, SDDS, Shuttered unilateral, SRD, SRD/DTS, Stereo, Stereo merged, Stereo printing master and Stereo decoded, Twin unilateral, Unilateral, Visatone, Western Electric
Representation	Sound	Layout of audio channels	Text			Image sequence CPL Additional material Rendition	How are the audio channels distributed (for example, if Dolby 5.1 and 7.1 are used)?	
Reference	Sound	Object ID for sound	Sequence	Mandatory		Image sequence Additional material		
Descriptive	Sound	Comment on sound	Text			Image sequence CPL Additional material Rendition		
Context	Sound	Post-production company	Text			Image sequence Additional material		
Descriptive	Sound	Sampling rate	Integer			Image sequence CPL Additional material Rendition		
Representation	Sound	Frame rate (fps)	Type of item			Image sequence Additional material Rendition	The frame rate for which the sound mix is intended.	0, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 23.976, 24, 25, 26, 27, 28, 29, 30, 48, 50, 60, 8, 9
Descriptive	Sound	Playing time per act (h:m:s)	text			Image sequence Additional material	Text field showing the playing time of the sound mix for each act.	
Descriptive	Sound	Playing time (h:m:s)	Time			Image sequence CPL Additional material Rendition	Aggregate playing time of the sound mix.	
Descriptive	Sound	Commentary language	Type of item		Repeated	Image sequence CPL Additional material Rendition	Shows the language of the commentary (if the sound mix contains a commentary).	List of languages (too long to list here)

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Sound	Soundtrack language	Type of item		Repeated	Image sequence CPL Additional material Rendition	Shows the languages used in the soundtrack (if the sound mix contains a soundtrack).	List of languages (too long to list here)
Descriptive	Sound	Dubbing language	Type of item		Repeated	Image sequence CPL Additional material Rendition	Shows the languages in which the cinematographic work is dubbed (if the sound mix contains dubbing).	
Context	Related material	Sound summary	Calculated	Mandatory		Image sequence Rendition	Calculated field that shows what sound mixes are available for this cinematographic work. Particularly relevant for restoration and the production of renditions and DCPs.	
Context	Related material	Subtitle summary	Calculated	Mandatory		Image sequence Rendition	Calculated field that shows what subtitle variants are available for this cinematographic work. Particularly relevant for restoration and the production of renditions and DCPs.	
Context	Identification	CPL	Type of item	Mandatory	Repeated	DCP	The link between the way in which the DCP is recorded and the way in which the CPLs are recorded in the DCP.	
Descriptive	Identification	CPL – encrypted	Type of item			CPL		Yes, No, NA, Unknown
Descriptive	Identification	CPL name	Text			CPL	The name of the CPL in accordance with the standard for CPL names.	
Context	Identification	DCP	Type of item			CPL	Contains a link to the DCP of which this CPL forms part.	
Descriptive	Identification	Cinematographic work	Type of item	Mandatory	Repeated	Image sequence CPL DCP Additional material Rendition	Contains a link to the cinematographic work in this AV object. If it refers to a compilation of works, the compilation must be recorded as the work in question.	
Reference	Identification	ID	Sequence	Mandatory		Image sequence CPL DCP Additional material Rendition		
Descriptive	Identification	PKL name	Text			DCP		
Reference	Identification	Player url	Url			Rendition	For renditions, this link can be used to view the preview in the digital archive.	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Identification	Title of copy	Compound	Mandatory		Image sequence CPL DCP Additional material Rendition	The title of the AV object. This is generally the same as the title of the cinematographic work, but sometimes the title of the AV object is not in the same language as the title of the work.	
Descriptive	Contents and structure	Discrepancy in description of cinematographic work	Text			Image sequence CPL Rendition		
Descriptive	Contents and structure	Completeness (summary)	Text			Image sequence CPL Rendition		
Descriptive	Contents and structure	Completeness (status)	Type of item			Image sequence CPL Rendition	Shows whether the AV object is a complete representation of the cinematographic work.	Complete, Virtually complete, Fragment, Incomplete, Short at act transitions, Unknown
Descriptive	Contents and structure	Description of contents	Text area			Image sequence CPL Rendition	Description of the contents of the AV object. Particularly relevant in the case of unidentified material.	
Descriptive	Contents and structure	Structure (editing)	Text			Image sequence CPL Rendition	States whether this is a specific version in terms of editing (e.g. certain acts are in a different order, or there is a different final act).	
Descriptive	Contents and structure	Title and intertitles	Text area			Image sequence CPL Rendition	States the title and intertitles, if relevant.	
Descriptive	Material properties	Number of Frames	Integer	Mandatory		Image sequence	Calculated field for the total number of frames in the image sequence (based on the image sequence field blocks).	
Descriptive	Material properties	File size	Compound			Rendition	File size of the rendition file. The file size always consists of a value and a unit.	
Descriptive	Material properties	File size.Unit	Type of item			Rendition		GB, KB, MB, TB
Descriptive	Material properties	File size.Size	Integer			Rendition		
Descriptive	Material properties	File size of all files	Compound			DCP	File size of the entire DCP. The file size always consists of a value and a unit.	
Descriptive	Material properties	File size of all files.Unit	Type of item			DCP		GB, KB, MB, TB

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Material properties	File size of all files.Size	Integer			DCP		
Context	Material properties	Date of creation	Date			CPL DCP Additional material Rendition		
Provenance	Material properties	Carrier info	Compound	Mandatory	Repeated	Image sequence DCP Additional material Rendition	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.	
Provenance	Material properties	Carrier info.Start date	Date			Image sequence CPL Additional material Rendition	The date since when the digital AV object has been stored on the carrier.	
Provenance	Material properties	Carrier info.Carrier	Type of item			Image sequence CPL Additional material Rendition	A carrier can be an LTO (Liner Tape-Open) tape or a hard disk. The digital archive is also regarded as a carrier.	
Provenance	Material properties	Carrier info.End date	Date			Image sequence CPL Additional material Rendition	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).	
Provenance	Material properties	Carrier info.In use yes/no	Type of item			Image sequence CPL Additional material Rendition	Shows whether the digital AV object is still stored on the carrier.	Yes, No, NA, Unknown
Provenance	Material properties	Carrier info.Link	Url			Image sequence CPL Additional material Rendition	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).	
Provenance	Material properties	Carrier info.Object ID used by external carrier	Text			Image sequence CPL Additional material Rendition	The identifying characteristic under which the digital AV object is stored on an external carrier.	
Representation	Material properties	Encoding software	text			Image sequence CPL Rendition	This may contain extra information o the contents of the file, particularly where errors have been found in certain software. We used to have files that no longer passed our quality control procedure because they had been made with an old version of the software.	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Material properties	Format	Type of item	Mandatory		Image sequence CPL Rendition	The format of the container, generally expressed as an extension	½ inch VK30, 1/2 inch, 13x8cm, 1/4 inch video, 16mm, 17,5mm, 18x24mm, 1i. B aka BCN, 1 i. C (AMPEX&SONY), 22mm, 28mm, 30mm, 30mm (Pathéorama), 35mm, 3,5x5inch, 44,5mm, 4,5x6cm, 5x5cm, 6,25mm, 63mm, 65mm, 68mm, 6x6cm, 70mm, 8mm, 8x8cm, 9,5mm, AC3 (.ac3), AIFF (.aif), Apple uncomp. (.mov), BetaLP, Betamax, BetaSP, Blu-ray, CD, Cineon (.cin), Compact cassette, D10-30 (.mxf), D10-50 (.mxf), D15-50 (.mxf), Dat, DCP - encrypted, DCP non-encrypted, Digibeta, Digital video, Digital video D2, Digital video D3, Digital video D4, DNxHD (.mov), .dpx, DPX (.dpx), DS8, Double 8, DVCAM, DVD, ED-Beta, Flash (.flv), H264 (.mov), HDCAM, Hi-8, Jpeg2K (.jp2), Jpeg2K (.mov), Jpeg2K (.mxf), Laserdisc, Mini-DV, .mov, .mp4, Mpeg1 (.mpg), Mpeg2 (.mpg), Mpeg2 (.mxf), Mpeg4 (.mp4), .mxf, NVT, Unknown, other file format, Prores 422 (.mov), Prores 4444 (.mov), S16mm, S8mm, Single 8, SuperBetamax, S-VHS, S-VHS-C, .tif, .tiff, TIFF (.tif), U-matic, unknown file format, VCR (Philips), VHS-HQ, VHS-LP, VHS-SP, Video, Video 2000, Video8, .wav, Wav (.wav), .xml
Descriptive	Material properties	Frame rate HFR Yes/No	Type of item			CPL Rendition		Yes, No, NA, Unknown
Provenance	Material properties	Historical migration data	Text			Image sequence DCP Additional material Rendition	Field containing information on carrier migration prior to the introduction of Carrier info.	
Representation	Material properties	Comment on material properties	Text			Image sequence CPL DCP Additional material Rendition	Additional material properties that are relevant but are not covered by other fields.	
Context	Material properties	Post-production company	Text			Image sequence DCP Rendition		

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Material properties	Frame rate (fps)	Type of item			Image sequence CPL Rendition	Frame rate for the AV object.	0, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 23.976, 24, 25, 26, 27, 28, 29, 30, 48, 50, 60, 8, 9
Descriptive	Material properties	Playing time (calculated) (h:m:s:f)	Time			Image sequence	Calculated field that calculates the aggregate playing time based on the total number of frames and the frame rate (fps).	
Descriptive	Material properties	Playing time (h:m:s:f)	Time			Image sequence CPL Rendition	Where it relates to an image sequence, this may differ from the calculated playing time as inserts etc. are not necessarily included.	
Descriptive	Material properties	Start of closing credits (h:m:s:f)	Time			CPL		
Representation	Material properties	Format version	Text			Image sequence	Details on the format.	
Provenance	Subtitles	Carrier info	Compound		Repeated	Image sequence CPL Additional material Rendition	Carrier info is a repeatable field block containing information on the carrier on which the digital AV object (or a part of it identified by a field block) is now stored, or used to be stored in the past.	
Provenance	Subtitles	Carrier info.Start date	Date			Image sequence CPL Additional material Rendition	The date since when the digital AV object has been stored on the carrier.	
Provenance	Subtitles	Carrier info.Carrier	Type of item			Image sequence CPL Additional material Rendition	A carrier can be an LTO (Liner Tape-Open) tape or a hard disk. The digital archive is also regarded as a carrier.	
Provenance	Subtitles	Carrier info.End date	Date			Image sequence CPL Additional material Rendition	If the digital AV object is no longer stored on the carrier, or if the carrier no longer exists, this field shows the date on which the object was deleted (if this is known).	
Provenance	Subtitles	Carrier info.In use yes/no	Type of item			Image sequence CPL Additional material Rendition	Shows whether the digital AV object is still stored on the carrier.	Yes, No, NA, Unknown
Provenance	Subtitles	Carrier info.Link	Url			Image sequence CPL Additional material Rendition	Contains a hyperlink to the digital AV object (if it can be accessed by means of a URL).	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Provenance	Subtitles	Carrier info.Object ID used by external carrier	Text			Image sequence CPL Additional material Rendition	The identifying characteristic under which the digital AV object is stored on an external carrier.	
Representation	Subtitles	Electronic subtitle file – file name	Text			CPL Additional material Rendition	States the file name of these subtitles if they are in a separate subtitle file.	
Representation	Subtitles	Format	Type of item		Repeated	Image sequence Additional material Rendition	The format of the subtitles.	.890, .pac, .rar, .srt, .stl, .txt, .xml
Reference	Subtitles	Object ID for subtitles	Sequence			Image sequence CPL Additional material Rendition		
Descriptive	Subtitles	Comments on subtitles	Text			Image sequence CPL Additional material Rendition		
Representation	Subtitles	Frame rate (fps)	Type of item		Repeated	Image sequence CPL Additional material Rendition	States the frame rate at which the subtitles are intended to be played.	0, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 23.976, 24, 25, 26, 27, 28, 29, 30, 48, 50, 60, 8, 9
Descriptive	Subtitles	Language of subtitles	Type of item		Repeated	Image sequence CPL Additional material Rendition		
Representation	Subtitles	Type of subtitles	Type of item			Image sequence CPL Additional material Rendition	States the type of subtitles used. For example, they may be burn-in subtitles if they are embedded in the image (the term is borrowed from analogue material) or switchable (on/off) subtitles if there is an option to choose subtitling when the object is played. Alternatively, there may be a separate subtitle file.	dialogue list, burn-in, separately playable subtitle file, rough subtitle file, switchable (on/off)
Provenance	Record management	Reason for alteration	Text			Image sequence DCP Additional material Rendition	Field containing textual information on an alteration affecting a batch.	
Descriptive	Status fields	Availability	Calculated	Mandatory		Image sequence CPL DCP Additional material Rendition	Calculated field showing whether the AV object is currently available. Takes account of loans.	

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Descriptive	Status fields	Suitability for screening	Calculated	Mandatory		Image sequence CPL DCP Additional material Rendition	Calculated field showing whether the AV object is suitable for screening. There are three possible values: 1 Suitable 2 Limited suitability (with explanation) 3 Not suitable (with explanation)	
Provenance	Status fields	Registration status	Type of item			Image sequence CPL DCP Additional material Rendition		Number only allocated (storage), Registration incomplete, Registration incomplete (storage), Registration complete
Descriptive	Language	Comment on language	Text area			Image sequence CPL Additional material Rendition		
Representation	Type	Image/Sound	Type of item	Mandatory		Image sequence CPL DCP Additional material Rendition	States whether the AV object contains image only, image and sound, or sound only. This field is needed because it is not possible to decide on the basis of the technical metadata alone what exactly the object contains. For example, a rendition may come with an audiotrack without there actually being any sound.	I, I/S, S, NA, Unknown
Descriptive	Type	DCP standard	Type of item			DCP		interop, smpte
Descriptive	Type	Function	Type of item	Mandatory		Image sequence CPL DCP Additional material Rendition		Conservation element, Unknown, Projection copy, Working materials.
Descriptive	Type	Is DCDM	Type of item	Mandatory		Image sequence		Yes, No, NA, Unknown

Type	Block	Name of field	Type	Mandatory	Repeated	Applies to	Description	Controlled list values
Representation	Type	Type	Type of item	Mandatory		Image sequence CPL DCP Additional material Rendition	Specifies the type of data object. It may state, for example, that the data object is an encrypted DCP.	Analogue video, Camera negative, CPL, CPL in DCP version file, DCDM, DCP, DCP - encrypted, DCP - non-encrypted, Digital video, Digital Intermediate, Digital Master, Duplicate negative / Intermediate negative, Film-related material, Fine Grain Master – black-and-white, Sound, Sound-negative, Sound-positive, Image sequence (born-digital), Image sequence (digital intermediate), Intermediate positive - colour, Internegative, Low-contrast copy, Negative, NA, Reverse copy, Reverse original, Unknown, Subtitles, Perfo tapes, Playable Rendition, Positive, Remainder, Technical file, Exhibition material