FILE FORMATS

v1.10.0

This document forms part of the Ministry of Education and Culture’s Open science and digital cultural heritage entity.
Licence
Creative Commons Suomi CC-BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/)

Users of this Specification are entitled to distribute the report, i.e. copy, circulate, display publicly and perform publicly, and modify it under the following conditions:

- The Ministry of Education and Culture is appointed the Original Author (not, however, so that notification would refer to a licensee or means by which the Specification is used as supported by the licensor).
- The user is not entitled to use the Specification commercially.
- If the user makes any modifications to the Specification or uses it as the basis for their own works, the derivative work shall be distributed in the same manner or under the same type of licence.
CONTENTS

1. INTRODUCTION ........................................................................................................................................ 5

1.1. Digital Preservation Services ................................................................................................................. 5
1.2. Recommended and Acceptable for Transfer File Formats ................................................................. 5
1.3. Technical Characteristics of File Formats ............................................................................................. 6
1.4. Format Registry ........................................................................................................................................ 7
1.5. Updates to this Specification .................................................................................................................... 7

2. RECOMMENDED FILE FORMATS ...................................................................................................... 8

2.1. Text ........................................................................................................................................................ 8
2.1.1. Comma Separated Values (CSV) ....................................................................................................... 8
2.1.2. Electronic Publications (EPUB) ......................................................................................................... 8
2.1.3. Extensible Hypertext Markup Language (XHTML) ........................................................................ 9
2.1.4. Extensible Markup Language (XML) ................................................................................................ 9
2.1.5. Hypertext Markup Language (HTML) .............................................................................................. 9
2.1.6. Open Document Format (ODF) ......................................................................................................... 10
2.1.7. PDF for Long-term Preservation: PDF-Archive (PDF/A) ................................................................. 11
2.1.8. Plain Text ............................................................................................................................................ 11

2.2. Audio .................................................................................................................................................... 11
2.2.1. Audio Interchange File Format (AIFF), LPCM-encoded .................................................................. 11
2.2.2. Broadcast Wave Format (BWf) ......................................................................................................... 12
2.2.3. Free Lossless Audio Codec (FLAC) .................................................................................................. 12
2.2.4. Linear Pulse-Code Modulation (LPCM) .......................................................................................... 12
2.2.5. MPEG-4 AAC – Advanced Audio Coding (AAC) ........................................................................... 13
2.2.6. Waveform Audio Format (WAV) ...................................................................................................... 13

2.3. Video ..................................................................................................................................................... 13
2.3.1. Digital Picture Exchange (DPX) ......................................................................................................... 13
2.3.2. FF Video Codec 1 (FFV1) .................................................................................................................. 14
2.3.3. JPEG 2000 Sequence ........................................................................................................................ 14
2.3.4. MPEG-4 AVC – Advanced Video Coding (AVC) ............................................................................. 14

2.4. Still Image ............................................................................................................................................ 15
2.4.1. Digital Negative (DNG) .................................................................................................................... 15
2.4.2. Joint Photographic Experts Group (JPEG) ........................................................................................ 15
2.4.3. Joint Photographic Experts Group JPEG 2000 (JP2) ....................................................................... 16
2.4.4. Portable Network Graphics (PNG) .................................................................................................... 16
2.4.5. Scalable Vector Graphics (SVG) ....................................................................................................... 16
2.4.6. Tagged Image File Format (TIFF) .................................................................................................... 17

2.5. Web Archive ....................................................................................................................................... 17
2.5.1. Web ARCHive Format (WARC) ..................................................................................................... 17

2.6. Geospatial data .................................................................................................................................... 17
2.6.1. Geographic Tagged Image File Format (GeoTIff) ............................................................................ 17
2.6.2. Geography Markup Language (GML) ............................................................................................. 17
2.6.3. Keyhole Markup Language (KML) .................................................................................................. 18

2.7. Databases .......................................................................................................................................... 18
2.7.1. Software Independent Archiving of Relational Databases (SIARD) .................................................. 18

2.8. Research Data .................................................................................................................................... 18
2.8.1. SPSS Portable (POR) ....................................................................................................................... 18

2.9. Scientific Software ............................................................................................................................... 19
2.9.1. Matlab .............................................................................................................................................. 19
2.9.2. Hierarchical Data Format (HDF5) ................................................................................................... 19

3. FILE FORMATS ACCEPTABLE FOR TRANSFER ..................................................................... 20

3.1. Text .................................................................................................................................................... 20
3.1.1. Microsoft Office Suite ....................................................................................................................... 20
3.1.2. Portable Document Format (PDF) .................................................................................................... 20
3.2. Audio .......................................................... 21
  3.2.1. Audio Interchange File Format (AIFF-C) .................................................. 21
  3.2.2. Moving Pictures Expert Group (MPEG) MPEG-1 layer-3, MPEG-2 layer-3 (MP3) .................................................. 21
  3.2.3. Windows Media Audio (WMA) ........................................ 22
3.3. Video .......................................................... 22
  3.3.1. Digital Video and its Variants (DV) .................................................. 22
  3.3.2. Moving Pictures Expert Group (MPEG-1, MPEG-2) .................................................. 22
  3.3.3. Windows Media Video (WMV) .................................................. 23
3.4. Still Image .................................................. 23
  3.4.1. Encapsulated Postscript (EPS) .................................................. 23
  3.4.2. Graphics Interchange Format (GIF) .................................................. 24
4. CONTAINER FORMATS ........................................ 25
5. TECHNICAL METADATA ........................................ 28
  5.1. ADDML .................................................. 28
     5.1.1. Applying the Schema .................................................. 28
     5.1.2. Example .................................................. 28
  5.2. AudioMD .................................................. 30
     5.2.1. Applying the Schema .................................................. 30
     5.2.2. Example .................................................. 31
  5.3. VideoMD .................................................. 32
     5.3.1. Example .................................................. 34
  5.4. MIX .................................................. 35
     5.4.1. Applying the Schema .................................................. 35
     5.4.2. Example .................................................. 36
6. CONTROLLED VOCABULARY .................................. 37
  6.1. Character Sets for Text Files .................................................. 40
  6.2. File Format Details .................................................. 40
ANNEX A. SUMMARY OF EVALUATION OF FILE FORMATS .................. 42
ANNEX B. HANDLING NON-RECOMMENDED AND NON-ACCEPTABLE FOR TRANSFER FORMATS .................. 44
  B.1. Requirements and Recommendations .................................................. 44
  B.2. Practical Guidelines .................................................. 45
ANNEX C. LOCAL XML SCHEMA FILES .................................. 48
REFERENCES .................................................. 51

This is an unofficial translation of the national file formats specification of Finland. Should this translation conflict with the Finnish version, the Finnish version will have higher priority.
1. INTRODUCTION

This document specifies the file formats, and technical metadata requirements, in which the national Digital Preservation Services (DPS) archives and ingests digital resources. The specification concentrates on digital content that is managed as files; recording medium is not within its scope.

1.1. Digital Preservation Services

DPS refer to the national services produced for the digital preservation of cultural heritage resources and research data. In this specification, a partner organization refers to an organization, department, or other entity using the DPS for the digital preservation of digital content.

Digital preservation refers to the reliable preservation of digital information for several decades or even centuries. Hardware, software, and file formats will become outdated, but the information must remain understandable. Reliable digital preservation requires active monitoring of information integrity and anticipation of various risks. Metadata, which describes for example the resource itself, its provenance and rights related to it, has a key role in preservation.

The DPS produced for cultural heritage resources guarantees the preservation of essential national information resources held in libraries, archives, and museums. Digital cultural heritage resources cover both digitized and born-digital information resources: publications falling within the bounds of legal deposit, government publications belonging to the national cultural heritage, and other digital information resources worth preserving created by organizations operating under the Ministry of Education and Culture.

The DPS produced for research data ensures the availability and preservation of digital research resources. This DPS supports a permanent and coordinated approach to support the management of research resources. The aim is to ensure the verifiability and repeatability of research at various stages of the life cycle and to make the results easy to use. This enables research results to be reused, evaluated, utilized in decision-making and secured by increasing digital data for future generations of researchers.

1.2. Recommended and Acceptable for Transfer File Formats

DPSs recognize two kinds of file formats: Recommended and acceptable for transfer. Recommended file formats in the former group are those that are considered to remain usable for a long time, whereas acceptable for transfer formats are popular (a significant amount of content is currently stored in these formats) but not archivable.

If the file format of a resource to be preserved is neither recommended nor acceptable for transfer, the file format must be normalized to a recommended file format, if possible. The partner organization is usually responsible for normalization, but the centralized DPS may offer support (e.g., information about applicable migration tools). Before normalizing the content, the DPS should be contacted in order to ensure that the most appropriate target format is used. It is important that files are not normalized just to acceptable-for-transfer formats, as this would necessitate a second migration to a recommended format. Each migration poses a risk to the data being preserved, so the number of migrations should be minimized. Further, new digitized content and born-digital content should be produced using a recommended file format. If there are several versions of a recommended file format, the latest approved version should be preferred.

The DPS and the partner organization utilizing it shall agree on how the migration process shall be carried out as a part of the preservation plan. The plan should describe, for example, what happens to the acceptable for transfer file format if it is migrated to a recommended format by DPS staff. Further, the plan should also specify one or more recommended file formats for each transfer format.

The correct selection of file formats is important for successful digital preservation. Maintenance of the selection criteria and the list of approved file formats is one of the key tasks of the DPS. In the selection process, file formats based on open and official standards are favored. If there are no such standards available, de facto – or industry standards have also been accepted.
This document specifies recommended and acceptable for transfer file formats for text, audio and video resources, still images, web archive containers, statistical information, and geospatial data. It should be noted that some file formats may not match this division of file formats. For example, PDF is regarded as a text format, but a PDF file may contain a still image. The actual content of submitted files shall be taken into account in a preservation plan. A report published by Library and Archives Canada was utilized in this specification [LAC], among other international sources.

Other DPS specifications include a specification metadata and packing the content for DPS [FI_META] and a specification for APIs for transferring data to and from the DPS [FI_API].

This specification uses the following terms as specified by IETF (Internet Engineering Task Force) in [RFC_2119].

- MUST
- MUST NOT
- REQUIRED
- SHOULD
- SHOULD NOT
- MAY
- OPTIONAL

### 1.3. Technical Characteristics of File Formats

The DPS imposes common restrictions on all preserved files that apply to all file formats, although not all limitations are always applicable. Especially:

1. Files MUST NOT use password protection or any other encryption techniques. Using them makes it difficult to preserve the file and, in some cases, will even prevent its preservation.
2. Files MUST NOT use DRM (Digital Rights Management) technologies. These may cause problems even to the bit-level preservation.
3. A digital signature on a file MUST NOT prevent the processing of the file. Using digital signatures MAY make digital preservation troublesome and, in some cases, even prohibits it.
4. Files MUST NOT be compressed. Many file formats are compressed and such compression is permitted, but separate packing (zip, etc.), for example, is prohibited.
5. A file MUST NOT lack any external components required for its presentation. These include, for example, fonts and other linked content that affect the presentation of a file. In addition, XML schemas are treated as external components, which MUST be included into a SIP as described in Error! Reference source not found., unless the schema is generally known by the DPS ingest.

In addition to the above common restrictions, the technical metadata in the submission information package shall provide detailed information about the preserved digital resources.

In addition to the mandatory technical metadata specified, there are media type-specific properties that are not relevant for other media types, but are essential or at least helpful when files in the media type in question are rendered. For instance, an audio file has various technical properties specific to audio files, such as sampling frequency. In digital preservation the role of technical metadata in general is to improve the usability and archivability of data.

Every SIP MUST contain at least the following technical metadata for each file:

1. Essential metadata which cannot be automatically derived from the file itself. For example, the character set of a text file shall be provided in technical metadata.

---

1 It is possible to deviate from these restrictions for specific reasons, but only so that the exceptions are agreed in writing in the DPS contract. Exceptions are always handled on a case by case basis by negotiating with the partner organization before signing the DPS contract.
2. Metadata related to the properties which has been specified as essential in the preservation plan. For instance, if aspect ratio needs to be preserved in migration, the aspect ratio of the original resource must be specified in technical metadata.

3. Embedded metadata that should be preserved in the migration process. For instance, if an EXIF file header contains essential metadata that may be lost in a migration to an archivable format, such metadata must be replicated in technical metadata.

Each file format description gives information on what metadata schema for the technical metadata must be used in the information packet. The DPS defines the elements and attributes of the metadata schema that are mandatory in the DPS precisely. In some cases, an unknown value in a mandatory element or attribute may be accepted. The partner organization must agree with the DPS on the use of unknown values during the deployment stage.

1.4. Format Registry

Format registry is a part of the DPS that contains metadata about file formats. Common format registries include PRONOM® and the website hosted by the Library of Congress on file format sustainability. Such registries are essential information sources for DPS to manage a variety of file formats, although there are weaknesses in their coverage and sometimes the file format information is not up-to-date.

Different format registries share some of the metadata elements, such as file format name, version, identifier, predecessor, and successor. For example, the file format name can be JPEG, version 1.00, identifier image/jpeg, and the successor JPEG version 1.01. However, there are also registry-specific metadata elements, which makes metadata exchange between registries more difficult. Further, there are no registry-independent exchange formats for this purpose.

Using both the metadata acquired from format registries and information supplied by producers (e.g., which applications they are using and for how long), the DPS will be able to find out when a certain file format becomes obsolete. This event will activate the process to revise the preservation plan in cooperation with the partner organization. After this, the decision is made on whether to migrate digital objects in the outdated file format to a new format. It the target format has not been specified, it is necessary to decide which file format to use. The target format shall be identified using the PRONOM Persistent Unique Identifier (PUID) whenever possible.

1.5. Updates to this Specification

This specification will be revised annually and it will be developed further based on the needs of partner organizations.

The three-level version number of the definition describes the level of changes that have been made to the specification. If the first digit is updated, it means significant changes to partner organizations’ systems that are integrated to the DPS. The middle digit reflects situations where changes may affect partner organizations’ systems. An update to the third digit describes situations where only minor corrections or clarifications have been made. For example, updates to examples in the specification are minor corrections.

The persistent identifier of the latest version of this specification is urn:nbn:fi-fe2020100578096

---

4. https://www.digitalpreservation.gov/formats/content/content_categories.shtml
2. RECOMMENDED FILE FORMATS

This section contains the specification of the recommended file formats for the national digital preservation services. A file format is accepted as a recommended format if its intelligibility can be guaranteed over the long term.

File formats are enumerated alphabetically and according to media type. If both container formats and ordinary file formats have been specified for a media type, only recommended file formats may be embedded in containers.

2.1. Text

2.1.1. Comma Separated Values (CSV)

CSV (comma separated values) is a file format for storing spreadsheet data to a text file. In CSV files a record is stored on one line, which is divided into fields by using separator characters.

Acceptable versions:

[ RFC_4180 ]; PRONOM: x-fmt/18

Vocabulary in the PREMIS format
Name semantic unit:

text/csv

Permissible character sets:

cf. section 6.1

Metadata schema for the mandatory technical metadata:

ADDML

2.1.2. Electronic Publications (EPUB)

EPUB is a hardware- and software-independent container format for electronic publications maintained by the International Digital Publishing Forum (IDPF) (versions 2.0.1–3.0.1) and by WC3 (the latest version 3.2). According to the standard, the structural metadata is included in the container. Text formatting and layout are created in the reading device.

Acceptable versions:

EPUB version 2.0.1: [EPUB 2.0.1] PRONOM: fmt/483
EPUB version 3.0.0: [EPUB_3.0.0]; PRONOM: fmt/483
EPUB version 3.0.1: [EPUB_3.0.1]; PRONOM: fmt/483
EPUB version 3.2: [EPUB_3.2]; PRONOM: fmt/483

Vocabulary in the PREMIS format
Name semantic unit:

application/epub+zip

Metadata schema for mandatory technical metadata

DPS does not require any technical metadata for EPUB files, but a producer may submit such information in the metadata format it prefers.
2.1.3. Extensible Hypertext Markup Language (XHTML)

XHTML is an XML adaptation based on HTML 4.0; the latest version recommended by W3C is version 5.

Acceptable versions:

- XHTML version 1.0; [XHTML_1.0]; PRONOM: fmt/102
- XHTML version 1.1; [XHTML_1.1]; PRONOM: fmt/103
- XHTML version 5.0; [HTML5]; PRONOM: fmt/471
- XHTML version 5.1; [HTML5.1]
- XHTML version 5.2; [HTML5.2]

Vocabulary in the PREMIS formatName semantic unit:

application/xhtml+xml

Permissible character sets:

cf. section 6.1

Metadata schema for mandatory technical metadata

DPS does not require any technical metadata for XHTML files, but a producer may submit such information in the metadata format it prefers.

2.1.4. Extensible Markup Language (XML)

XML is a cross-platform markup language based on SGML. XML, which is maintained by W3C, describes the logical structure of a document, but not its layout.

If a preserved XML file uses XML schemas, the partner organization MUST agree with DPS regarding the schema availability when ingesting content to the DPS.

Acceptable versions:

- XML version 1.0; [XML_1.0]; PRONOM: fmt/101
- XML version 1.1; [XML_1.1];

Vocabulary in the PREMIS formatName semantic unit:

text/xml

Permissible character sets:

cf. section 6.1

Metadata schema for mandatory technical metadata

DPS does not require any technical metadata for XML files, but producer may submit such information in the metadata format it prefers.

2.1.5. Hypertext Markup Language (HTML)

HTML is a simple hypertext markup language based on SGML. It has become the standard markup language for producing www pages. HTML describes the content of a document. The DPS recommends both HTML file format version 4.01, which is standardized by ISO as ISO/IEC 15445, or version 5.X [HTML5, HTML5.1, HTML5.2].

Acceptable versions:

- HTML version 4.01; [ISO_15445]; PRONOM: fmt/100
- HTML version 5.0; [HTML5]; PRONOM: fmt/471
- HTML version 5.1; [HTML5.1]
- HTML version 5.2; [HTML5.2]
Vocabulary in the PREMIS format
Name semantic unit:
  text/html

Permissible character sets:
  cf. section 6.1

Metadata schema for mandatory technical metadata
  DPS does not require any technical metadata for HTML files, but producer may submit such information in the metadata format it prefers.

2.1.6. Open Document Format (ODF)

ODF is an open container format developed by OASIS, which can be used to present documents produced by office applications, such as spreadsheets, diagrams, slide shows, and text documents. The ODF file format version 1.0 has been standardized by ISO as ISO/IEC 26300:2006 and version 1.2 as ISO/IEC 26300:2015. Technically, an Open document is a ZIP compressed archive, which contains several files and directories. The content, layout, metadata, and settings of a document are separated into individual XML files.

The most common file formats conforming to the ODF standard are:

- .odt, text document
- .ods, spreadsheet
- .odp, slide show presentation
- .odg, drawing
- .odf, equation

Acceptable versions:

- .odt version 1.0; [ISO_26300:2006]; PRONOM: fmt/136
- .odt version 1.1; PRONOM: fmt/290
- .odt version 1.2; [ISO_26300:2015]; PRONOM: fmt/291
- .odt version 1.3; [ODF_1.3]
- .ods version 1.0; [ISO_26300:2006]; PRONOM: fmt/137
- .ods version 1.1; PRONOM: fmt/294
- .ods version 1.2; ISO_26300:2015]; PRONOM: fmt/295
- .ods version 1.3; [ODF_1.3]
- .odp version 1.0; [ISO_26300:2006]; PRONOM: fmt/138
- .odp version 1.1; PRONOM: fmt/292
- .odp version 1.2; ISO_26300:2015]; PRONOM: fmt/293
- .odp version 1.3; [ODF_1.3]
- .odg version 1.0; [ISO_26300:2006]; PRONOM: fmt/139
- .odg version 1.1; PRONOM: fmt/296
- .odg version 1.2; ISO_26300:2015]; PRONOM: fmt/297
- .odg version 1.3; [ODF_1.3]
- .odf version 1.0; [ISO_26300:2006].
- .odf version 1.2; [ISO_26300:2015]
- .odf version 1.3; [ODF_1.3].

Vocabulary in the PREMIS format
Name semantic unit:
  application/vnd.oasis.opendocument.text
  application/vnd.oasis.opendocument.spreadsheet
  application/vnd.oasis.opendocument.presentation
  application/vnd.oasis.opendocument.graphics
  application/vnd.oasis.opendocument.formula
**Metadata schema for mandatory technical metadata**

DPS does not require any technical metadata for ODF files, but producer may submit such information in the metadata format it prefers.

### 2.1.7. PDF for Long-term Preservation: PDF-Archive (PDF/A)

PDF/A is an archivable version of the PDF file format. A PDF/A file contains all information required for rendering of the document. Characteristics that may undermine digital preservation are removed from PDF/A files. Such characteristics include scripts, audio, video, and encryption. The DPS recommends PDF/A versions 1a and 1b, which have been standardized by ISO as ISO 19005-1, versions 2a, 2b, and 2u, which have been standardized as ISO 19005-2, and versions 3a, 3b ja 3u, which have been standardized as ISO 195005-3.

**Acceptable versions:**

- PDF/A-1a; [ISO_19005-1]; PRONOM: fmt/95
- PDF/A-1b; [ISO_19005-1]; PRONOM: fmt/354
- PDF/A-2a; [ISO_19005-2]; PRONOM: fmt/476
- PDF/A-2b; [ISO_19005-2]; PRONOM: fmt/477
- PDF/A-2u; [ISO_19005-2]; PRONOM: fmt/478
- PDF/A-3a; [ISO_19005-3]; PRONOM: fmt/479
- PDF/A-3b; [ISO_19005-3]; PRONOM: fmt/480
- PDF/A-3u; [ISO_19005-3]; PRONOM: fmt/481

**Vocabulary in the PREMIS format**

Name semantic unit:

- application/pdf

**Metadata schema for mandatory technical metadata**

DPS does not require any technical metadata for PDF/A files, but producer may submit such information in the metadata format it prefers.

### 2.1.8. Plain Text

A plain text file contains only text without any formatting or other layout information. The characters that can be used in text files are specified in character sets.

**Acceptable versions:**

- [RFC_2046]; PRONOM: x-fmt/111

**Vocabulary in the PREMIS format**

Name semantic unit:

- text/plain

**Permissible character sets:**

- cf. section 6.1

**Metadata schema for mandatory technical metadata**

If a text file contains table-based information, its structure should be described using ADDML [ADDML].

### 2.2. Audio

#### 2.2.1. Audio Interchange File Format (AIFF), LPCM-encoded

AIFF is an audio file format developed by Apple Inc. A recommended AIFF file may contain only lossless PCM encoded audio. For AIFF files, one does not have to make a distinction between container and soundtrack, as both are treated as one in the DPS.
Acceptable versions:
   AIFF 1.3; [AIFF]; PRONOM: x-fmt/135

Vocabulary in the PREMIS format
   Name semantic unit:
   audio/x-aiff

Metadata schema for the mandatory technical metadata:
   AudioMD

2.2.2. Broadcast Wave Format (BWF)
BWF is an exchange format for audio files developed by the EBU (European Broadcasting Union), based on the WAV file format. The BWF is an extension to WAV, which guarantees compatibility with WAV files. A BWF file may only contain LPCM encoded audio for it to be considered recommended for preservation. An essential feature of the BWF is the Broadcast extension chunk (Bext), which may be used for storing metadata. The International Association of Sound and Audiovisual Archives (IASA) recommends the BWF file format for archiving. For BWF files, one does not have to make a distinction between container and soundtrack, as both are treated as one in the DPS.

Acceptable versions:
   BWF Version 2; [BWF]; PRONOM: fmt/527

Vocabulary in the PREMIS format
   Name semantic unit:
   audio/x-wav

Metadata schema for the mandatory technical metadata:
   AudioMD

2.2.3. Free Lossless Audio Codec (FLAC)
FLAC is a lossless compression method for audio files developed by the Xiph.org Foundation. It is free and open to the public.

Acceptable versions:
   FLAC 1.2.1; [FLAC]; PRONOM: fmt/279

This file format can be included in the following containers:
   Matroska [MATROSKA]

Vocabulary in the PREMIS format
   Name semantic unit:
   audio/flac

Metadata schema for the mandatory technical metadata:
   AudioMD

2.2.4. Linear Pulse-Code Modulation (LPCM)
PCM (Pulse-Code Modulation) is a digital representation of an analog signal, where the signal is sampled in uniform intervals. Linear Pulse-Code Modulation (LPCM) [LPCM] is a special type of PCM, where the quantization levels are linearly uniform, which is typically used in soundtracks in container formats.

This file format can be included in the following containers:
   Audio Video Interleave [AVI]
   DV (raw) [IEC_61834, SMPTE_314, SMPTE_370]
Material Exchange Format [SMPTE_377]
Motion JPEG 2000 [MJ2]
Matroska [MATROSKA]
Quicktime MOVie [MOV]

**Vocabulary in the PREMIS format**
Name semantic unit:
- audio/L16, audio/L8, audio/L20, audio/L24

**Metadata schema for the mandatory technical metadata:**
AudioMD

---

### 2.2.5. MPEG-4 AAC – Advanced Audio Coding (AAC)

AAC is a lossy file format. It was designed by Fraunhofer Gesellschaft as a replacement to the MP3 file format. MPEG-4 AAC is an updated version of the earlier MPEG-2 AAC standard. The MPEG-4 AAC file format (MPEG-4 Part 3) has been standardized as ISO/IEC 14496-3. It forms a part of the MPEG-2 and MPEG-4 specifications.

**Acceptable versions:**
- MPEG-4 Part 3; [ISO_14496-3]

**This file format may be included in the following containers:**
- MPEG-2 part 1 Transport Stream [ISO_13818-1]
- MPEG-4 part 14 [ISO_14496-14]
- Material Exchange Format [SMPTE_377]
- Quicktime MOVie [MOV]

**Vocabulary in the PREMIS format**
Name semantic unit:
- audio/mp4

**Metadata schema for the mandatory technical metadata:**
AudioMD

---

### 2.2.6. Waveform Audio Format (WAV)

The WAV container format was developed by IBM and Microsoft. A WAV file may only contain LPCM encoded audio for it to be considered recommended for preservation. For WAV files, no distinction between container and soundtrack needs to be made, as both are treated as one in the DPS.

**Acceptable versions**
- WAV; [WAV_LOC, WAV IBM MS]; PRONOM: fmt/141

**Vocabulary in the PREMIS format**
Name semantic unit:
- audio/x-wav

**Metadata schema for the mandatory technical metadata:**
AudioMD

---

### 2.3. Video

#### 2.3.1. Digital Picture Exchange (DPX)

In the Digital Picture Exchange (DPX) file format, video is stored as subsequent DPX images without packaging. This file format is commonly used when digitizing movies. File format is standardized as an SMPTE standard [SMPTE_268].
Acceptable versions:

DPX 2.0; [SMPTE_268]; PRONOM: fmt/541

Vocabulary in the PREMIS formatName semantic unit:

image/x-dpx

Metadata schema for the mandatory technical metadata:

MIX

2.3.2. FF Video Codec 1 (FFV1)

FFV1 is a lossless compression method for video developed since 2003 by the FFmpeg project. In 2015 the format was adopted to the IETF standardization process as part of the Cellar working group. The latest stable version is FFV1.3, published in 2013. This version has only minor changes compared to versions FFV1.0 and FFV1.1.

Acceptable versions:

FFV1 version 3 [RFC_9043]

This file format can be included in the following containers:

Matroska [MATROSKA]

Vocabulary in the PREMIS formatName semantic unit:

video/x-ffv

Metadata schema for the mandatory technical metadata:

MIX

2.3.3. JPEG 2000 Sequence

In JPEG 2000 Sequence, video files are stored as sequences of JPEG 2000 still images, which are packaged in Material Exchange Format (MXF), in Motion JPEG 2000 (MJ2), or QuickTime MOVie containers. JPEG 2000 is either a lossless or lossy file format for still images, standardized by ISO (ISO/IEC 15444-1). Motion JPEG 2000, which refers to the container, has been standardized in ISO/IEC 15444-3.

Acceptable versions:

JPEG 2000 part 1; [ISO_15444-1]; PRONOM: x-fmt/392

This file format can be included in the following containers:

Material Exchange Format [SMPTE_377]
Motion JPEG 2000 [MJ2]
Quicktime MOVie [MOV]

Vocabulary in the PREMIS formatName semantic unit:

video/jpeg2000

Metadata schema for the mandatory technical metadata:

VideoMD

2.3.4. MPEG-4 AVC – Advanced Video Coding (AVC)

MPEG-4 part 10 (AVC, H.264), developed by Moving Pictures Expert Group (MPEG), is a lossy compression method for video files. MPEG-4 is used in digital video, interactive graphic software, and multimedia. MPEG-4 part 10 has been standardized by ISO (ISO/IEC 14496-10).
Acceptable versions:

MPEG-4; [ISO_14496-10]; PRONOM: fmt/199

This file format may be included in the following containers:

Material Exchange Format [SMPTE_377]  
MPEG-2 part 1 Transport Stream [ISO_13818-1]  
MPEG-4 part 14 [ISO_14496-14]  
Quicktime MOVie [MOV]

Vocabulary in the PREMIS format

Name semantic unit:

video/mp4

Metadata schema for the mandatory technical metadata:

VideoMD

2.4. Still Image

2.4.1. Digital Negative (DNG)

DNG is a vendor-independent and royalty-free still image file format developed by Adobe Inc. for the lossless conversion of RAW images from digital cameras. DNG is a recommend file format if the SIP contains the same image or images also in either TIFF or JPEG 2000 format. Adobe has submitted DNG to ISO for incorporation into their revision of TIFF/EP.

Acceptable versions:

DNG 1.3.0.0; [DNG_13]; PRONOM: fmt/438  
DNG 1.4.0.0; [DNG_14]; PRONOM: fmt/730  
DNG 1.5.0.0; [DNG_15];

Vocabulary in the PREMIS format

Name semantic unit:

image/tiff

Metadata schema for the mandatory technical metadata:

MIX

2.4.2. Joint Photographic Experts Group (JPEG)

The JPEG standard (ISO/IEC 10918-1), developed by the Joint Photographic Experts Group, refers to both the file format and the compression method. The JPEG compression method is a lossy compression method for both true color (RGB) and black and white still images.

Acceptable versions:

JPEG part 1 version 1.00; [ISO_10918-1]; PRONOM: fmt/42  
JPEG part 1 version 1.01; [ISO_10918-1]; PRONOM: fmt/43  
JPEG part 1 version 1.02; [ISO_10918-1]; PRONOM: fmt/44  
JPEG/EXIF version 2.0; PRONOM; x-fmt/398  
JPEG/EXIF version 2.1; PRONOM; x-fmt/390  
JPEG/EXIF version 2.2; PRONOM; x-fmt/391  
JPEG/EXIF version 2.2.1; PRONOM; fmt/645  
JPEG/EXIF version 2.3  
JPEG/EXIF version 2.3.1  
JPEG/EXIF version 2.3.2
2.4.3. Joint Photographic Experts Group JPEG 2000 (JP2)

JPEG 2000, developed by the Joint Photographic Experts Group, provides both a lossless and a lossy compression method for still images. JPEG 2000 compression provides better image quality and smaller file size than JPEG. The ISO standardized (ISO/IEC 15444-1) JPEG 2000 supports the halftone, and the RGB and the CMYK color schemes.

Acceptable versions:

JPEG 2000 part 1; [ISO_15444-1]; PRONOM: x-fmt/392

2.4.4. Portable Network Graphics (PNG)

PNG is a still image file format specified by the W3C. Lossless PNG supports 16-bit halftone images and 48-bit color images. PNG is standardized in ISO/IEC 15948.

Acceptable versions:

PNG 1.2; [ISO_15948]; PRONOM: fmt/13

2.4.5. Scalable Vector Graphics (SVG)

Scalable Vector Graphics (SVG) is a file format for 2D vector graphics based on XML. The development of SVG started in 1999. The format is further developed and maintained by the World Wide Web Consortium (W3C). It is an open standard and its latest stable version is 1.1 [SVG].

Acceptable versions:

SVG 1.1; [SVG]; PRONOM: fmt/92

DPS does not require any technical metadata, but a producer may submit such information in a metadata format it prefers.

5 The file must not contain animations or interactive functionality.
2.4.6. Tagged Image File Format (TIFF)

TIFF is a raster image file format developed by Aldus (which is as of this writing part of Adobe). TIFF images may be compressed or uncompressed, black and white, halftone or color (RGB or CMYK). The DPS requires TIFF version 6.0, and only lossless images should be stored in TIFF containers.

Acceptable versions:

   TIFF 6.0; [TIFF]; PRONOM: fmt/353

Vocabulary in the PREMIS format
Name semantic unit:

   image/tiff

Metadata schema for the mandatory technical metadata:

   MIX

2.5. Web Archive

2.5.1. Web ARChive Format (WARC)

The WARC (Web ARChive) container format specifies a method for packing a large number of web pages into a single archive file. The format was originally developed by the Internet Archive, and it is now ISO standardized (ISO 28500). WARC is an advanced version of the ARC file format.

Acceptable versions:

   WARC version 1.0; [ISO_28500]; PRONOM: fmt/1355

Vocabulary in the PREMIS format
Name semantic unit:

   application/warc

Metadata schema for mandatory technical metadata

   DPS does not require any technical metadata for WARC files, but a producer may submit such information in the metadata format it prefers.

2.6. Geospatial data

2.6.1. Geographic Tagged Image File Format (GeoTIff)

GeoTIFF is an extension of TIFF 6.0. It specifies a method for incorporating geospatial metadata into TIFF images [GEOTIFF].

Acceptable versions:

   GeoTIFF 1.0 [GEOTIFF]; PRONOM: fmt/155

Vocabulary in the PREMIS format
Name semantic unit:

   image/tiff

Metadata schema for mandatory technical metadata:

   MIX

2.6.2. Geography Markup Language (GML)

Geography Markup Language (GML) is an XML-based markup language that enables the description of various features of geolocation data. GML has been standardized by ISO as ISO 19136 [ISO_19136].

Acceptable versions:

   GML 3.2.2 [ISO_19136] PRONOM: x-fmt/227
**Vocabulary in the PREMIS format**

Name semantic unit:
- application/gml+xml

**Permissible character sets**
- cf. section 6.1

**Metadata schema for mandatory technical metadata**

DPS does not require any technical metadata for GML files, but a producer may submit such information in the metadata format it prefers.

### 2.6.3. Keyhole Markup Language (KML)

Keyhole Markup Language (KML) is an XML-based markup language intended for the annotation and visualization of two- and three-dimensional maps. The file format was developed by Google, but it has also been endorsed by OGC [KML].

**Acceptable versions:**
- KML 2.3 [KML]; PRONOM: fmt/244

**Vocabulary in the PREMIS format**

Name semantic unit:
- application/vnd.google-earth.kml+xml

**Permissible character sets**
- cf. section 6.1

**Metadata schema for mandatory technical metadata**

DPS does not require any technical metadata for KML files, but a producer may submit such information in the metadata format it prefers.

### 2.7. Databases

#### 2.7.1. Software Independent Archiving of Relational Databases (SIARD)

SIARD is a file format designed for preserving relational databases in a vendor independent format. The file format is based on XML and SQL and it is developed by the Swiss Federal Archives.

**Acceptable versions:**
- SIARD 2.1.1 [SIARD2.1.1]
- SIARD 2.2 [SIARD2.2]

**Vocabulary in the PREMIS format**

Name semantic unit:
- application/x-siard

**Metadata schema for mandatory technical metadata**

DPS does not require any technical metadata for SIARD files, but a producer may submit such information in the metadata format it prefers.

### 2.8. Research Data

#### 2.8.1. SPSS Portable (POR)

SPSS (Statistical Package for the Social Sciences, IBM SPSS Statistics) Portable file format enables exchange of statistical files between different computers, operating systems applications, and versions of these applications. Any SPSS version and most statistical applications may be used to render the files. In POR format, data extracted from the SPSS program is stored as an ASCII flat file.
Acceptable versions:

- PRONOM: fmt/997

Vocabulary in the PREMIS formatName semantic unit:

- application/x-spss-por

Metadata schema for mandatory technical metadata:

DPS does not require any technical metadata for POR files, but a producer may submit such information in the metadata format it prefers.

2.9. Scientific Software

2.9.1. Matlab

Matlab is a software and programming language developed for numerical computing maintained by the MathWorks. Matlab material may contain various files, but only data files (.mat) and code files (.m) are recommended formats for preservation. Code files are essentially structured text files and are treated as such in the DPS.

Acceptable versions:

- Matlab Level 5, version 7; PRONOM: fmt/806
- Matlab version 7.3; PRONOM: fmt/828

Vocabulary in the PREMIS formatName semantic unit:

- application/matlab

Metadata schema for mandatory technical metadata:

DPS does not require any technical metadata for Matlab files, but a producer may submit such information in the metadata format it prefers.

2.9.2. Hierarchical Data Format (HDF5)

Hierarchical Data Format (HDF) is a general-purpose set of file formats especially designed for saving scientific data. The file format development is controlled by the not-for-profit HDF Group.

Acceptable versions:

- HDF5 1.10 [HDF5]; PRONOM\textsuperscript{6}: fmt/807; fmt/286; fmt/287

Vocabulary in the PREMIS formatName semantic unit:

- application/x-hdf5

Metadata schema for mandatory technical metadata:

DPS does not require any technical metadata for HDF5 files, but a producer may submit such information in the metadata format it prefers.

\textsuperscript{6} PRONOM PUIDs refer to different file format superblock versions. The version 3 does not yet have a PRONOM PUID.
3. FILE FORMATS ACCEPTABLE FOR TRANSFER

This section specifies the file formats acceptable for transfer to the national digital preservation service. A file format has been qualified if it is used by several partner organizations and if there is a lot of material to be preserved that is available in the format. The DPS may migrate these files to recommended file formats during ingest, according to the requirements and conditions specified by the preservation plan. If the partner organization normalizes its own digital resources prior to submission, the normalization shall always result in a recommended file format.

File formats are enumerated alphabetically by media type. If both container and file formats have been specified for a media type, the container format may contain only the recommended or acceptable for transfer file formats specified in this specification.

3.1. Text

3.1.1. Microsoft Office Suite

The Microsoft Office Suite is widely used commercial software for producing and presenting documents such as spreadsheets, slide show presentations, and word processing documents. Later versions of the Microsoft Office Suite support Office Open XML formats (Transitional and Strict), which have been standardized by ISO (ISO 29500-1).

The following file formats of the Microsoft Office Suite (version 97 or later) are acceptable for transfer.

- Word Document, Word processing
- Excel Spreadsheet, Spreadsheet
- Powerpoint, Slide show presentation

Acceptable versions:

- Microsoft Word (97-2003); PRONOM: fmt/40
- Microsoft Word (2007 onwards); [ISO_29500-1]; PRONOM: fmt/412
- Microsoft Excel (8); PRONOM: fmt/61
- Microsoft Excel (8X); PRONOM: fmt/62
- Microsoft Excel (2007 onwards); [ISO_29500-1]; PRONOM: fmt/214
- Microsoft Powerpoint (97-2003); PRONOM: fmt/126
- Microsoft Powerpoint (2007 onwards); [ISO_29500-1]; PRONOM: fmt/215

Vocabulary in the PREMIS formatName semantic unit:

- application/msword
- application/vnd.ms-excel
- application/vnd.ms-powerpoint
- application/vnd.openxmlformats-officedocument.wordprocessingml.document
- application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
- application/vnd.openxmlformats-officedocument.presentationml.presentation

Metadata schema for mandatory technical metadata:

DPS does not require any technical metadata for Microsoft office files, but a producer may submit such information in the metadata format it prefers.

3.1.2. Portable Document Format (PDF)

Adobe Inc.’s PDF is a platform-independent and open file format for text documents. The layout, fonts, graphics, and colors of the original document are preserved in the PDF file. PDF version 1.7 has been standardized by ISO (ISO 32000-1).
Acceptable versions:

- PDF 1.2; [PDF_1.2]; PRONOM: fmt/16
- PDF 1.3; [PDF_1.3]; PRONOM: fmt/17
- PDF 1.4; [PDF_1.4]; PRONOM: fmt/18
- PDF 1.5; [PDF_1.5]; PRONOM: fmt/19
- PDF 1.6; [PDF_1.6]; PRONOM: fmt/20
- PDF 1.7; [ISO_32000-1]; PRONOM: fmt/276

Vocabulary in the PREMIS formatName semantic unit:

- application/pdf

Metadata schema for mandatory technical metadata:

DPS does not require any technical metadata for PDF files, but a producer may submit such information in the metadata format it prefers.

3.2. Audio

3.2.1. Audio Interchange File Format (AIFF-C)

AIFF-C is a compressed version of an AIFF file (cf. section 2.2.1). Lossless LPCM-encoded audio in an AIFF-container is a recommended format (cs. section 2.2.1), but a compressed AIFF file is acceptable for transfer. For AIFF-C files, one does not have to make a distinction between container and soundtrack; both are treated as one in the DPS.

Acceptable versions:

- AIFF-C; [AIFF-C]; PRONOM: x-fmt/136

Vocabulary in the PREMIS formatName semantic unit:

- audio/x-aiff

Metadata schema for the mandatory technical metadata:

AudioMD

3.2.2. Moving Pictures Expert Group (MPEG) MPEG-1 layer-3, MPEG-2 layer-3 (MP3)

MP3 is a lossy audio format developed by Moving Pictures Expert Group (MPEG). It is widely used and supported by current audio players. The compression method of the MP3 deletes all audio frequencies which cannot be heard by humans, and compresses the remaining file. MPEG-1 layer-3 has been standardized by ISO as ISO/IEC 11172-3 and MPEG-2 layer-3 as ISO/IEC 13818-3.

Acceptable versions:

- MPEG-1 layer-3; [ISO_11172-3]; PRONOM: fmt/134
- MPEG-2 layer-3; [ISO_13818-3]; PRONOM: fmt/134

This file format can be included in the following containers:

- Audio Video Interleave [AVI]
- Material Exchange Format [SMPTE_377]
- MPEG-1 part 1 Program Stream [ISO_11172-1]
- MPEG-2 part 1 Program Stream [ISO_13818-1]
- MPEG-2 part 1 Transport Stream [ISO_13818-1]
- MPEG-4 part 14 [ISO_14496-14]
- Quicktime MOVie [MOV]
Vocabulary in the PREMIS formatName semantic unit:
  audio/mpeg

Metadata schema for the mandatory technical metadata:
  AudioMD

3.2.3. Windows Media Audio (WMA)
WMA is a lossy compression method for audio files developed by Microsoft. WMA is similar to MP3. In addition to the original WMA compression method, the current file format specification includes three other compression methods: WMA Pro, WMA Voice, and WMA Lossless. The specification is owned by Microsoft and it has not been published.

Acceptable versions:
  WMA 9; PRONOM: fmt/132

This file format can be included in the following containers:
  Advanced Systems Format [ASF]

Vocabulary in the PREMIS formatName semantic unit:
  audio/x-ms-wma

Metadata schema for the mandatory technical metadata:
  AudioMD

3.3. Video

3.3.1. Digital Video and its Variants (DV)
DV is a compression method for storing raw video. DV was developed in cooperation between several equipment manufacturers. It has been standardized in standards IEC 61834, SMPTE 314, and SMPTE 370, which are known as DV, DVCAM, DVCPRO-50, and DVCPRO-HD.

Acceptable versions:
  DV (and its variants); [IEC_61834, SMPTE_314, SMPTE_370]; PRONOM: x-fmt/152

This file format can be included in the following containers:
  Audio Video Interleave [AVI]
  DV (raw format) [IEC_61834, SMPTE_314, SMPTE_370]
  Material Exchange Format [SMPTE_377]
  Quicktime MOVie [MOV]

Vocabulary in the PREMIS formatName semantic unit:
  video/dv

Metadata schema for the mandatory technical metadata:
  VideoMD

3.3.2. Moving Pictures Expert Group (MPEG-1, MPEG-2)
MPEG-1 part 2 (MPG, MPEG) file format uses lossy compression for video. It was developed by Moving Pictures Expert Group (MPEG), and it has been standardized by ISO as ISO/IEC 11172-2.

MPEG-2 part 2 (H.262) is a lossy compression method for audio and video. It was developed by Moving Pictures Expert Group (MPEG), and ISO has standardized it as ISO/IEC 13818-2. MPEG-2 supports variable
bitrate via only storing changes in audio and video between frames. Of all the MPEG standards, MPEG-2 part 2 is the most widely used file format for HDTV quality video.

Acceptable versions:

- MPEG-1; [ISO_11172-2]; PRONOM: x-fmt/649
- MPEG-2; [ISO_13818-2]; PRONOM: x-fmt/640

This file format can be included in the following containers:

- Audio Video Interleave [AVI]
- Material Exchange Format [SMPTE_377]
- MPEG-1 part 1 Program Stream [ISO_11172-1]
- MPEG-2 part 1 Program Stream [ISO_13818-1]
- MPEG-2 part 1 Transport Stream [ISO_13818-1]
- MPEG-4 part 14 [ISO_14496-14]
- Quicktime MOVie [MOV]

Vocabulary in the PREMIS format
Name semantic unit: video/mpeg

Metadata schema for the mandatory technical metadata:

VideoMD

3.3.3. Windows Media Video (WMV)

WMV is a compression method for video. It was developed by Microsoft, and it has been standardized by the Society of Motion Picture and Television Engineers (SMPTE) as SMPTE 421M, better known as VC-1.

Acceptable versions:

- WMV 9; [SMPTE_421]; PRONOM: fmt/133

This file format can be included in the following containers:

- Advanced Systems Format [ASF]

Vocabulary in the PREMIS format
Name semantic unit: video/x-ms-wmv

Metadata schema for the mandatory technical metadata:

VideoMD

3.4. Still Image

3.4.1. Encapsulated Postscript (EPS)

EPS, developed by Adobe, is a PostScript file format that may contain both vector and raster images. The file format is used for exchanging PostScript images between applications.

Acceptable versions:

- EPS version 3.0; [EPS]; PRONOM: fmt/124

Vocabulary in the PREMIS format
Name semantic unit: application/postscript
3.4.2. Graphics Interchange Format (GIF)

GIF is a lossless compression method for images developed by CompuServe. GIF files can contain 8-bit grayscale or color images. The number of concurrent colors is limited to 256.

Acceptable versions:

- GIF 1987a; [GIF_87a]; PRONOM: fmt/3
- GIF 1989a; [GIF_89a]; PRONOM: fmt/4

Vocabulary in the PREMIS format

- Name semantic unit: image/gif

Metadata schema for the mandatory technical metadata:

- MIX
4. CONTAINER FORMATS

Containers are complex digital objects, which contain component parts. These components are not independent. For instance, a movie (as a container) may contain one or more soundtracks and film tracks. In such cases it may be necessary to have metadata both for the container object (movie) and the component parts. For instance, if there are multiple soundtracks, checksum is required only for the entire container, but there must be technical metadata for each soundtrack.

When a digital resource is a container, technical metadata shall be divided and described as shown in this section. Table 2 in section 6 gives the mandatory MIME type for each container. Containers do not have version numbers, except Matroska, for which version 4 is the only acceptable version. Currently each video format requires both container and videotrack (and possible soundtracks) in their own metadata blocks. For other digital resources, this kind of representation of technical metadata is not mandatory. An example of metadata as required here in XML format is available at https://digitalpreservation.fi/specifications/examples.

Each component part of the container shall have its own metadata record in PREMIS format. Each metadata record shall have its own METS techMD frame. The type of the record shall be “bitstream” and it shall contain the identifier of the component part, format designation, and the format version according to the following example:

```xml
<mets:techMD ID="tech-videopremis" CREATED="2011-03-15T12:13:14">
  <mets:mdWrap MDTYPE="PREMIS:OBJECT" MDTYPEVERSION="2.2">
    <mets:xmlData>
      <premis:object xsi:type="premis:bitstream">
        <premis:objectIdentifier>
          <premis:objectIdentifierType>UUID</premis:objectIdentifierType>
          <premis:objectIdentifierValue>387e9e6c-4596-4baf-b316-49cc6af1c4b2</premis:objectIdentifierValue>
        </premis:objectIdentifier>
        <premis:objectCharacteristics>
          <premis:compositionLevel>0</premis:compositionLevel>
          <premis:format>
            <premis:formatDesignation>
              <premis:formatName>video/mp4</premis:formatName>
            </premis:formatDesignation>
            <premis:formatRegistry>
              <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
              <premis:formatRegistryKey>fmt/199</premis:formatRegistryKey>
            </premis:formatRegistry>
          </premis:format>
        </premis:objectCharacteristics>
      </premis:object>
    </mets:xmlData>
  </mets:mdWrap>
</mets:techMD>
```

Other administrative and descriptive metadata of the component part shall be encapsulated in METS according to the requirements specified elsewhere in this document. For instance, technical metadata for audio and video shall be expressed in AudioMD- and VideoMD-formats as specified in Chapters 5.2 and 5.3.
Containers shall be described with PREMIS in the same way, but the host PREMIS record shall be linked to the previously described, "bitstream"-type PREMIS metadata records. An example of links to two component parts:

```xml
<premis:relationship>
  <premis:relationshipType>structural</premis:relationshipType>
  <premis:relationshipSubType>includes</premis:relationshipSubType>
  <premis:relatedObjectIdentification>
    <premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>
    <premis:relatedObjectIdentifierValue>387e9e6c-4596-4baf-b316-49cc6af1c6a2</premis:relatedObjectIdentifierValue>
  </premis:relatedObjectIdentification>
  <premis:relatedObjectIdentification>
    <premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>
    <premis:relatedObjectIdentifierValue>164780d4-bd77-487f-8619-dfb45ce91084</premis:relatedObjectIdentifierValue>
  </premis:relatedObjectIdentification>
</premis:relationship>
```

Component parts of the container shall be included in the METS file description using the `<stream>`-element. The ADMID attribute of the `<stream>` element shall contain links to the administrative metadata records of the component part. In the example below, the component part has links to two different administrative metadata records:

```xml
<mets:file ADMID="tech-container" ID="file-container">
  <mets:FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="file://movie.mp4" />
  <mets:stream ADMID="tech-videopremis tech-videomd" />
  <mets:stream ADMID="tech-audiopremis tech-audiomd" />
</mets:file>
```

Links between descriptive metadata and container objects and their component parts shall be made in a METS structmap using DMDID- and FILEID-attributes. References from structmap to component parts of the container can be made (FILEID-attribute may refer to the ID-attribute of the `<stream>`-element).

The container, its component parts and administrative metadata are linked together as expressed in Figure 1.
Figure 1: Connecting container, its component parts and administrative metadata
5. TECHNICAL METADATA

5.1. ADDML

The mandatory technical metadata for CSV files should be presented using the ADDML metadata schema [ADDML] in SIPs and DIPs. If a text contains structural information, the structure should be expressed using ADDML metadata schema.

5.1.1. Applying the Schema

The ADDML metadata schema in the SIP and DIP information packages should be used according to the original schema. For CSV files the field separator character (fieldSeparatingChar) and the record separator character (recordSeparator) are required.

5.1.2. Example

```
<addml:addml
  xmlns:addml="https://www.arkivverket.no/standarder/addml">
  <addml:dataset>
    <addml:reference name="Sample" />
    <addml:flatFiles>
      <addml:flatFile name="noname.csv" definitionReference="ref001">
        <addml:properties>
          <addml:property name="filesize">
            <addml:value>281109</addml:value>
          </addml:property>
        </addml:properties>
      </addml:flatFile>
    </addml:flatFiles>
    <addml:flatFileDefinitions>
      <addml:flatFileDefinition name="ref001" typeReference="rec001">
        <addml:recordDefinitions>
          <addml:recordDefinition name="rdef001">
            <addml:fieldDefinitions>
              <addml:fieldDefinition name="name" typeReference="str">
                <addml:description>Person name</addml:description>
              </addml:fieldDefinition>
              <addml:fieldDefinition name="email" typeReference="str">
                <addml:description>Email address</addml:description>
              </addml:fieldDefinition>
            </addml:fieldDefinitions>
          </addml:recordDefinition>
        </addml:recordDefinitions>
      </addml:flatFileDefinition>
    </addml:flatFileDefinitions>
    <addml:structureTypes>
      <addml:flatFileType name="rec001">
        <addml:charset>UTF-8</addml:charset>
        <addml:delimFileFormat>
          <addml:recordSeparator>CR+LF</addml:recordSeparator>
          <addml:fieldSeparatingChar>;</addml:fieldSeparatingChar>
        </addml:delimFileFormat>
      </addml:flatFileType>
    </addml:structureTypes>
  </addml:dataset>
</addml:addml>
```
</addml:fieldType>
</addml:fieldTypes>
</addml:structureTypes>

</addml:flatFiles>
</addml:dataset>
</addml:addml>
5.2. **AudioMD**

Mandatory technical metadata for audio files shall be expressed using the AudioMD [AUDIOMD] schema in the SIPs and DIPs.

5.2.1. **Applying the Schema**

The usage of the elements and attributes of the audioMD schema is specified below. The following abbreviations are used to express occurrences and obligations:

- **T** Repeateable
- **ET** Not repeatable
- **P** Mandatory
- **V** Optional (allowed in SIP and will be stored in the DPS)

Only extensions to the audioMD schema required by the national DPS are listed below. Any element or attribute of the audioMD schema may be used as specified in the schema. If unknown values are used in elements or attributes, then the `<note>` element should indicate why and what it means. For example, whether 0 (zero) means that the information is unavailable, or if the element is not suitable, etc.

<table>
<thead>
<tr>
<th>Element/attribute</th>
<th>Occurrence/obligation</th>
<th>Regulations and recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;audioMD&gt;</code></td>
<td>ET, P</td>
<td>The root element must contain <code>&lt;fileData /&gt;</code> and <code>&lt;audioInfo /&gt;</code> elements. In the audioMD schema these elements are optional.</td>
</tr>
<tr>
<td><code>&lt;fileData&gt;</code></td>
<td>ET, P</td>
<td>Mandatory element in the DPS.</td>
</tr>
<tr>
<td><code>&lt;audioDataEncoding&gt;</code></td>
<td>T, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS. A value (: unav) can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td><code>&lt;bitPerSample&gt;</code></td>
<td>T, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td><code>&lt;compression&gt;</code></td>
<td>T, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td><code>&lt;codecCreatorApp&gt;</code></td>
<td>ET, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS. Values (: unav) or (: unap) can be allowed as an unknown value if the information cannot be easily found out. (: unap) only for uncompressed audio.</td>
</tr>
<tr>
<td><code>&lt;codecCreatorAppVersion&gt;</code></td>
<td>ET, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS. Values (: unav) or (: unap) can be allowed as an unknown value if the information cannot be easily found out. (: unap) only for uncompressed audio.</td>
</tr>
<tr>
<td><code>&lt;codecName&gt;</code></td>
<td>ET, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS. Values (: unav) or (: unap) can be allowed as an unknown value if the information cannot be easily found out. (: unap) only for uncompressed audio.</td>
</tr>
<tr>
<td>Element/attribute</td>
<td>Occurrence/obligation</td>
<td>Regulations and recommendations</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>--&lt;codecQuality&gt;</td>
<td>ET, P</td>
<td>Optional in the audioMD schema, but mandatory in the DPS.</td>
</tr>
</tbody>
</table>
| --<dataRate>          | T, P                  | Optional in the audioMD schema, but mandatory in the DPS.  
A value 0 can be allowed as an unknown value if the information cannot be easily found out. |
| --<dataRateMode>      | T, P                  | Optional in the audioMD schema, but mandatory in the DPS.                                                                                            |
| --<samplingFrequency> | T, P                  | Optional in the audioMD schema, but mandatory in the DPS.  
A value 0 can be allowed as an unknown value if the information cannot be easily found out. |
| --<audioInfo>         | ET, P                 | Optional in the audioMD schema, but mandatory in the DPS.                                                                                            |
| --<duration>          | T, P                  | Optional in the audioMD schema, but mandatory in the DPS.  
A value (:unav) can be allowed as an unknown value if the information cannot be easily found out. |
| --<numChannels>       | T, P                  | Optional in the audioMD schema, but mandatory in the DPS.  
A value (:unav) can be allowed as an unknown value if the information cannot be easily found out. |

5.2.2. Example

```xml
<amd:AUDIOMD xmlns:amd="https://www.loc.gov/audioMD/" ANALOGDIGITALFLAG="FileDigital">
  <amd:fileData>
    <amd:audioDataEncoding>PCM</amd:audioDataEncoding>
    <amd:bitsPerSample>8</amd:bitsPerSample>
    <amd:compression>
      <amd:codecCreatorApp>SoundForge</amd:codecCreatorApp>
      <amd:codecCreatorAppVersion>10</amd:codecCreatorAppVersion>
      <amd:codecName>(:unap)</amd:codecName>
      <amd:codecQuality>lossy</amd:codecQuality>
    </amd:compression>
    <amd:dataRate>256</amd:dataRate>
    <amd:dataRateMode>Fixed</amd:dataRateMode>
    <amd:samplingFrequency>44.1</amd:samplingFrequency>
  </amd:fileData>

  <amd:audioInfo>
    <amd:duration>PT1H30M</amd:duration>
    <amd:numChannels>1</amd:numChannels>
  </amd:audioInfo>
</amd:AUDIOMD>
```
5.3. VideoMD

The usage of the elements and attributes of the videoMD schema is specified below. The following abbreviations are used to express occurrences and obligations:

- T Repeatable
- ET Not repeatable
- P Mandatory
- V Optional (allowed in SIP and will be stored in the DPS)

The table below lists the extensions to the videoMD schema required by the national DPS. Any element or attribute of the videoMD schema may be used as specified in the schema. If unknown values are used in elements or attributes, the <note> element should indicate the usage. For example, whether 0 (zero) means that the information is unavailable, or if the element is not suitable, etc.

<table>
<thead>
<tr>
<th>Element/attribute</th>
<th>Occurrence/obligation</th>
<th>Regulations and recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;videoMD&gt;</td>
<td>ET, P</td>
<td>The root element must contain &lt;fileData /&gt; and &lt;videoInfo /&gt; elements. In the videoMD schema these elements are optional.</td>
</tr>
<tr>
<td>--&lt;fileData&gt;</td>
<td>ET, P</td>
<td></td>
</tr>
<tr>
<td>--&lt;duration&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value (:unav) can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>--&lt;dataRate&gt;</td>
<td>ET,P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>--&lt;bitsPerSample&gt;</td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>--&lt;color&gt;</td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td>--&lt;compression&gt;</td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td>---&lt;codecCreatorApp&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (:unav) or (:unap) can be allowed as an unknown value if the information cannot be easily found out. (:unap) only for uncompressed video.</td>
</tr>
<tr>
<td>Element/attribute</td>
<td>Occurrence/obligation</td>
<td>Regulations and recommendations</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>---&lt;codecCreatorAppVersion&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (:unav) or (:unap) can be allowed as an unknown value if the information cannot be easily found out. (:unap) only for uncompressed video or for software that does not have versioning.</td>
</tr>
<tr>
<td>---&lt;codecName&gt;</td>
<td>EP, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (:unav) or (:unap) can be allowed as an unknown value if the information cannot be easily found out. (:unap) only for uncompressed video.</td>
</tr>
<tr>
<td>---&lt;codecQuality&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td>---&lt;dataRateMode&gt;</td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td>---&lt;frameRate&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>---&lt;frame&gt;</td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
<tr>
<td>---&lt;pixelsHorizontal&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>---&lt;pixelsVertical&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>---&lt;PAR&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. A value 0 can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>---&lt;DAR&gt;</td>
<td>ET, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (:unav) or (:etal) can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td>Element/attribute</td>
<td>Occurrence/obligation</td>
<td>Regulations and recommendations</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td><code>&lt;sampling&gt;</code></td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (: unav) or (:unap) can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td><code>&lt;signalFormat&gt;</code></td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS. Values (: unav) or (:unap) can be allowed as an unknown value if the information cannot be easily found out.</td>
</tr>
<tr>
<td><code>&lt;sound&gt;</code></td>
<td>T, P</td>
<td>Optional in the videoMD schema, but mandatory in the DPS.</td>
</tr>
</tbody>
</table>

### 5.3.1. Example

```xml
<vmd:VIDEOMD xmlns:vmd="https://www.loc.gov/videoMD/"
ANALOGDIGITALFLAG="FileDigital">
  <vmd:fileData>
    <vmd:duration>PT2H05M</vmd:duration>
    <vmd:dataRate>8</vmd:dataRate>
    <vmd:bitsPerSample>24</vmd:bitsPerSample>
    <vmd:color>Color</vmd:color>
    <vmd:compression>
      <vmd:codecCreatorApp>SoundForge</vmd:codecCreatorApp>
      <vmd:codecName>(:unav)</vmd:codecName>
      <vmd:codecQuality>lossy</vmd:codecQuality>
    </vmd:compression>
    <vmd:dataRateMode>Fixed</vmd:dataRateMode>
    <vmd:frame>
      <vmd:pixelsHorizontal>640</vmd:pixelsHorizontal>
      <vmd:pixelsVertical>480</vmd:pixelsVertical>
      <vmd:PAR>1.0</vmd:PAR>
      <vmd:DAR>4/3</vmd:DAR>
    </vmd:frame>
    <vmd:frameRate>24</vmd:frameRate>
  </vmd:fileData>
</vmd:VIDEOMD>
```
5.4. MIX

The mandatory technical metadata for still image files shall be expressed using the MIX [MIX] schema in SIPs and DIPs.

5.4.1. Applying the Schema

The usage of the elements and attributes of the MIX schema is given below. The following abbreviations are used to express occurrences and obligations:

- T Repeatable
- ET Not repeatable
- P Mandatory
- V Optional (allowed in SIP and will be stored in the DPS)
- ES Not recommended (allowed in SIP, but will not be stored in the DPS). Not used in DIP.

Only extensions to the MIX schema required by the national DPS are listed. Some BasicDigitalObjectInformation elements have been dropped due to their overlap with PREMIS. Any element or attribute of the MIX schema may be used as specified in the schema7.

<table>
<thead>
<tr>
<th>Element/attribute</th>
<th>Occurrence/obligation</th>
<th>Regulations and recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;BasicDigitalObjectInformation&gt;</td>
<td>P, ET</td>
<td>The element is optional in the MIX schema, but mandatory in the DPS</td>
</tr>
<tr>
<td>&lt;BasicImageInformation&gt;</td>
<td>P, ET</td>
<td>The element is optional in the MIX schema, but mandatory in the DPS</td>
</tr>
<tr>
<td>&lt;ImageAssessmentMetadata&gt;</td>
<td>P, ET</td>
<td>The element is optional in the MIX schema, but mandatory in the DPS</td>
</tr>
<tr>
<td>&lt;ObjectIdentifier&gt;</td>
<td>ES, ET</td>
<td>The element is mandatory in the MIX schema, but in the DPS this information is expressed using PREMIS. Therefore, using the element is not recommended in order to avoid repetition.</td>
</tr>
<tr>
<td>&lt;fileSize&gt;</td>
<td>ES, ET</td>
<td>The element is mandatory in the MIX schema, but in the DPS this information is expressed using PREMIS. Therefore, using the element is not recommended in order to avoid repetition.</td>
</tr>
<tr>
<td>&lt;FormatDesignation&gt;</td>
<td>ES, ET</td>
<td>The element is mandatory in the MIX schema, but in the DPS this information is expressed using PREMIS. Therefore, using the element is not recommended in order to avoid repetition.</td>
</tr>
<tr>
<td>&lt;FormatRegistry&gt;</td>
<td>ES, ET</td>
<td>The element is mandatory in the MIX schema, but in the DPS this information is expressed using PREMIS. Therefore, using the element is not recommended in order to avoid repetition.</td>
</tr>
<tr>
<td>&lt;Fixity&gt;</td>
<td>ES, T</td>
<td>The element is mandatory in the MIX schema, but in the DPS this information is expressed using PREMIS. Therefore, using the element is not recommended in order to avoid repetition.</td>
</tr>
</tbody>
</table>

7 The MIX specification defines some elements as mandatory, but in the MIX schema these elements are optional. In the DPS, MIX must be used according to specification with the exceptions defined in this document.
5.4.2. Example

```xml
<mix:mix xmlns:mix="https://www.loc.gov/mix/v20">
  <mix:BasicDigitalObjectInformation>
    <mix:Compression>
      <mix:compressionScheme>JPEG 2000 Lossless</mix:compressionScheme>
    </mix:Compression>
  </mix:BasicDigitalObjectInformation>
  <mix:BasicImageInformation>
    <mix:BasicImageCharacteristics>
      <mix:imageWidth>869</mix:imageWidth>
      <mix:imageHeight>1271</mix:imageHeight>
      <mix:PhotometricInterpretation>
        <mix:colorSpace>ICCBased</mix:colorSpace>
        <mix:ColorProfile>
          <mix:iccProfileName>Adobe RGB</mix:iccProfileName>
          <mix:iccProfileVersion>1998</mix:iccProfileVersion>
          <mix:iccProfileURI>
            https://www.adobe.com/digitalimag/adobergb.html
          </mix:iccProfileURI>
        </mix:ColorProfile>
      </mix:PhotometricInterpretation>
      <mix:SpecialFormatCharacteristics>
        <mix:JPEG2000>
          <mix:CodecCompliance>
            <mix:codec>Kakadu</mix:codec>
            <mix:codecVersion>5.2</mix:codecVersion>
            <mix:codestreamProfile>P1</mix:codestreamProfile>
            <mix:complianceClass>C1</mix:complianceClass>
          </mix:CodecCompliance>
          <mix:EncodingOptions>
            <mix:Tiles>
              <mix:tileWidth>256</mix:tileWidth>
              <mix:tileHeight>256</mix:tileHeight>
            </mix:Tiles>
            <mix:qualityLayers>12</mix:qualityLayers>
            <mix:resolutionLevels>6</mix:resolutionLevels>
          </mix:EncodingOptions>
        </mix:JPEG2000>
      </mix:SpecialFormatCharacteristics>
    </mix:BasicImageCharacteristics>
    <mix:ImageAssessmentMetadata>
      <mix:SpatialMetrics>
        <mix:samplingFrequencyPlane>
          camera/scanner focal plane
        </mix:samplingFrequencyPlane>
        <mix:samplingFrequencyUnit>cm</mix:samplingFrequencyUnit>
      </mix:SpatialMetrics>
      <mix:ImageColorEncoding>
        <mix:BitsPerSample>
          <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
          <mix:bitsPerSampleUnit>integer</mix:bitsPerSampleUnit>
        </mix:BitsPerSample>
        <mix:samplesPerPixel>3</mix:samplesPerPixel>
      </mix:ImageColorEncoding>
    </mix:ImageAssessmentMetadata>
  </mix:BasicImageInformation>
</mix:mix>
```
6. CONTROLLED VOCABULARY

The following table contains the controlled vocabulary to be used in SIPs and DIPs for expressing the file format and its version. The column labels refer to corresponding PREMIS elements [PREMIS]. This vocabulary should also be used, when applicable, in other parts of SIPs and DIPs (e.g., the MIMETYPE attribute in METS [METS]).

<table>
<thead>
<tr>
<th>File format</th>
<th>formatName</th>
<th>formatVersion</th>
<th>formatRegistryKey</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV</td>
<td>text/csv</td>
<td>-</td>
<td>x-fmt/18</td>
</tr>
<tr>
<td>EPUB</td>
<td>application/epub+zip</td>
<td>2.0.1</td>
<td>fmt/483</td>
</tr>
<tr>
<td>EPUB</td>
<td>application/epub+zip</td>
<td>3.0.0</td>
<td>fmt/483</td>
</tr>
<tr>
<td>EPUB</td>
<td>application/epub+zip</td>
<td>3.0.1</td>
<td>fmt/483</td>
</tr>
<tr>
<td>EPUB</td>
<td>application/epub+zip</td>
<td>3.2</td>
<td>fmt/483</td>
</tr>
<tr>
<td>XHTML</td>
<td>application/xhtml+xml</td>
<td>1.0</td>
<td>fmt/102</td>
</tr>
<tr>
<td>XHTML</td>
<td>application/xhtml+xml</td>
<td>1.1</td>
<td>fmt/103</td>
</tr>
<tr>
<td>XHTML</td>
<td>application/xhtml+xml</td>
<td>5.0</td>
<td>fmt/471</td>
</tr>
<tr>
<td>XHTML</td>
<td>application/xhtml+xml</td>
<td>5.1</td>
<td>-</td>
</tr>
<tr>
<td>XHTML</td>
<td>application/xhtml+xml</td>
<td>5.2</td>
<td>-</td>
</tr>
<tr>
<td>XML</td>
<td>text/xml</td>
<td>1.0</td>
<td>fmt/101</td>
</tr>
<tr>
<td>XML</td>
<td>text/xml</td>
<td>1.1</td>
<td>-</td>
</tr>
<tr>
<td>HTML</td>
<td>text/html</td>
<td>4.01</td>
<td>fmt/100</td>
</tr>
<tr>
<td>HTML</td>
<td>text/html</td>
<td>5.0</td>
<td>fmt/471</td>
</tr>
<tr>
<td>HTML</td>
<td>text/html</td>
<td>5.1</td>
<td>-</td>
</tr>
<tr>
<td>HTML</td>
<td>text/html</td>
<td>5.2</td>
<td>-</td>
</tr>
<tr>
<td>ODT</td>
<td>application/vnd.oasis.opendocument.text</td>
<td>1.0</td>
<td>fmt/136</td>
</tr>
<tr>
<td>ODT</td>
<td>application/vnd.oasis.opendocument.text</td>
<td>1.1</td>
<td>fmt/290</td>
</tr>
<tr>
<td>ODT</td>
<td>application/vnd.oasis.opendocument.text</td>
<td>1.2</td>
<td>fmt/291</td>
</tr>
<tr>
<td>ODT</td>
<td>application/vnd.oasis.opendocument.text</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>ODS</td>
<td>application/vnd.oasis.opendocument.spreadsheet</td>
<td>1.0</td>
<td>fmt/137</td>
</tr>
<tr>
<td>ODS</td>
<td>application/vnd.oasis.opendocument.spreadsheet</td>
<td>1.1</td>
<td>fmt/294</td>
</tr>
<tr>
<td>ODS</td>
<td>application/vnd.oasis.opendocument.spreadsheet</td>
<td>1.2</td>
<td>fmt/295</td>
</tr>
<tr>
<td>ODS</td>
<td>application/vnd.oasis.opendocument.spreadsheet</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>ODP</td>
<td>application/vnd.oasis.opendocument.presentation</td>
<td>1.0</td>
<td>fmt/138</td>
</tr>
<tr>
<td>ODP</td>
<td>application/vnd.oasis.opendocument.presentation</td>
<td>1.1</td>
<td>fmt/292</td>
</tr>
<tr>
<td>ODP</td>
<td>application/vnd.oasis.opendocument.presentation</td>
<td>1.2</td>
<td>fmt/293</td>
</tr>
<tr>
<td>ODP</td>
<td>application/vnd.oasis.opendocument.presentation</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>ODG</td>
<td>application/vnd.oasis.opendocument.graphics</td>
<td>1.0</td>
<td>fmt/139</td>
</tr>
<tr>
<td>ODG</td>
<td>application/vnd.oasis.opendocument.graphics</td>
<td>1.1</td>
<td>fmt/296</td>
</tr>
<tr>
<td>ODG</td>
<td>application/vnd.oasis.opendocument.graphics</td>
<td>1.2</td>
<td>fmt/297</td>
</tr>
<tr>
<td>ODG</td>
<td>application/vnd.oasis.opendocument.graphics</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>File format</td>
<td>formatName</td>
<td>formatVersion</td>
<td>formatRegistryKey</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------</td>
<td>---------------</td>
<td>------------------</td>
</tr>
<tr>
<td>ODF</td>
<td>application/vnd.oasis.opendocument.formula</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>ODF</td>
<td>application/vnd.oasis.opendocument.formula</td>
<td>1.2</td>
<td>-</td>
</tr>
<tr>
<td>ODF</td>
<td>application/vnd.oasis.opendocument.formula</td>
<td>1.3</td>
<td>-</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-1a</td>
<td>fmt/95</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-1b</td>
<td>fmt/354</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-2a</td>
<td>fmt/474</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-2b</td>
<td>fmt/477</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-2u</td>
<td>fmt/476</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-3a</td>
<td>fmt/479</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-3b</td>
<td>fmt/480</td>
</tr>
<tr>
<td>PDF/A</td>
<td>application/pdf</td>
<td>A-3u</td>
<td>fmt/481</td>
</tr>
<tr>
<td>Plain text</td>
<td>text/plain</td>
<td>-</td>
<td>x-fmt/111</td>
</tr>
<tr>
<td>AIFF</td>
<td>audio/x-aiff</td>
<td>1.3</td>
<td>x-fmt/135</td>
</tr>
<tr>
<td>BWF</td>
<td>audio/x-wav</td>
<td>2</td>
<td>fmt/527</td>
</tr>
<tr>
<td>FLAC</td>
<td>audio/flac</td>
<td>1.2.1</td>
<td>fmt/279</td>
</tr>
<tr>
<td>LPCM</td>
<td>audio/L16, audio/L8, audio/L20, audio/L24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AAC</td>
<td>audio/mp4</td>
<td>-</td>
<td>fmt/199</td>
</tr>
<tr>
<td>WAV</td>
<td>audio/x-wav</td>
<td>-</td>
<td>fmt/141</td>
</tr>
<tr>
<td>DPX</td>
<td>image/x-dpx</td>
<td>2.0</td>
<td>fmt/541</td>
</tr>
<tr>
<td>FFV1</td>
<td>video/x-ffv</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>JPEG2000</td>
<td>video/jpeg2000</td>
<td>-</td>
<td>x-fmt/392</td>
</tr>
<tr>
<td>AVC</td>
<td>video/mp4</td>
<td>-</td>
<td>fmt/199</td>
</tr>
<tr>
<td>DNG</td>
<td>image/tiff</td>
<td>1.3</td>
<td>fmt/438</td>
</tr>
<tr>
<td>DNG</td>
<td>image/tiff</td>
<td>1.4</td>
<td>fmt/730</td>
</tr>
<tr>
<td>DNG</td>
<td>image/tiff</td>
<td>1.5</td>
<td>-</td>
</tr>
<tr>
<td>JPEG</td>
<td>image/jpeg</td>
<td>1.00</td>
<td>fmt/42</td>
</tr>
<tr>
<td>JPEG</td>
<td>image/jpeg</td>
<td>1.01</td>
<td>fmt/43</td>
</tr>
<tr>
<td>JPEG</td>
<td>image/jpeg</td>
<td>1.02</td>
<td>fmt/44</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.0</td>
<td>x-fmt/398</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.1</td>
<td>x-fmt/390</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.2</td>
<td>x-fmt/391</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.2.1</td>
<td>fmt/645</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.3</td>
<td>-</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.3.1</td>
<td>-</td>
</tr>
<tr>
<td>JPEG/EXIF</td>
<td>image/jpeg</td>
<td>2.3.2</td>
<td>-</td>
</tr>
<tr>
<td>JP2</td>
<td>image/jp2</td>
<td>-</td>
<td>x-fmt/392</td>
</tr>
<tr>
<td>SVG</td>
<td>image/svg+xml</td>
<td>1.1</td>
<td>fmt/92</td>
</tr>
<tr>
<td>TIFF</td>
<td>image/tiff</td>
<td>6.0</td>
<td>fmt/353</td>
</tr>
<tr>
<td>PNG</td>
<td>image/png</td>
<td>1.2</td>
<td>fmt/13</td>
</tr>
<tr>
<td>WARC</td>
<td>application/warc</td>
<td>1.0</td>
<td>fmt/289</td>
</tr>
<tr>
<td>GEOTIFF</td>
<td>image/tiff</td>
<td>1.0</td>
<td>fmt/155</td>
</tr>
<tr>
<td>GML</td>
<td>application/gml+xml</td>
<td>3.2.1</td>
<td>fmt/1047</td>
</tr>
<tr>
<td>KML</td>
<td>application/vnd.google-earth.kml+xml</td>
<td>2.3</td>
<td>fmt/244</td>
</tr>
<tr>
<td>SIARD</td>
<td>application/x-siard</td>
<td>2.1.1</td>
<td>-</td>
</tr>
<tr>
<td>SIARD</td>
<td>application/x-siard</td>
<td>2.2</td>
<td>-</td>
</tr>
<tr>
<td>POR</td>
<td>application/x-spss-por</td>
<td>-</td>
<td>fmt/997</td>
</tr>
<tr>
<td>Matlab</td>
<td>application/matlab</td>
<td>7</td>
<td>fmt/806</td>
</tr>
<tr>
<td>Matlab</td>
<td>application/matlab</td>
<td>7.3</td>
<td>fmt/828</td>
</tr>
<tr>
<td>File format</td>
<td>formatName</td>
<td>formatVersion</td>
<td>formatRegistryKey</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>HDF5</td>
<td>application/x-hdf5</td>
<td>1.10</td>
<td>fmt/807, fmt/286, fmt/287</td>
</tr>
<tr>
<td>MS Word</td>
<td>application/msword</td>
<td>97-2003</td>
<td>fmt/40</td>
</tr>
<tr>
<td>MS Word</td>
<td>application/vnd.openxmlformats-officedocument.wordprocessingml.document</td>
<td>2007 onwards</td>
<td>fmt/412</td>
</tr>
<tr>
<td>MS Excel</td>
<td>application/vnd.ms-excel</td>
<td>8</td>
<td>fmt/61</td>
</tr>
<tr>
<td>MS Excel</td>
<td>application/vnd.ms-excel</td>
<td>8X</td>
<td>fmt/62</td>
</tr>
<tr>
<td>MS Excel</td>
<td>application/vnd.openxmlformats-officedocument.spreadsheetml.sheet</td>
<td>2007 onwards</td>
<td>fmt/214</td>
</tr>
<tr>
<td>MS Powerpoint</td>
<td>application/vnd.ms-powerpoint</td>
<td>97-2003</td>
<td>fmt/126</td>
</tr>
<tr>
<td>MS Powerpoint</td>
<td>application/vnd.openxmlformats-officedocument.presentationml.presentation</td>
<td>2007 onwards</td>
<td>fmt/215</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.2</td>
<td>fmt/16</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.3</td>
<td>fmt/17</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.4</td>
<td>fmt/18</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.5</td>
<td>fmt/19</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.6</td>
<td>fmt/20</td>
</tr>
<tr>
<td>PDF</td>
<td>application/pdf</td>
<td>1.7</td>
<td>fmt/276</td>
</tr>
<tr>
<td>AIFF-C</td>
<td>audio/x-aiff</td>
<td>-</td>
<td>x-fmt/136</td>
</tr>
<tr>
<td>MP3</td>
<td>audio/mpeg</td>
<td>1</td>
<td>fmt/134</td>
</tr>
<tr>
<td>MP3</td>
<td>audio/mpeg</td>
<td>2</td>
<td>fmt/134</td>
</tr>
<tr>
<td>WMA</td>
<td>audio/x-mpeg</td>
<td>9</td>
<td>fmt/132</td>
</tr>
<tr>
<td>DV</td>
<td>video/dv</td>
<td>-</td>
<td>x-fmt/152</td>
</tr>
<tr>
<td>MPEG</td>
<td>video/mpeg</td>
<td>1</td>
<td>fmt/649</td>
</tr>
<tr>
<td>MPEG</td>
<td>video/mpeg</td>
<td>2</td>
<td>fmt/640</td>
</tr>
<tr>
<td>MPEG</td>
<td>video/mpeg</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>WMV</td>
<td>video/x-ms-wmv</td>
<td>9</td>
<td>fmt/133</td>
</tr>
<tr>
<td>EPS</td>
<td>application/postscript</td>
<td>3.0</td>
<td>fmt/124</td>
</tr>
<tr>
<td>GIF</td>
<td>image/gif</td>
<td>1987a</td>
<td>fmt/3</td>
</tr>
<tr>
<td>GIF</td>
<td>image/gif</td>
<td>1989a</td>
<td>fmt/4</td>
</tr>
</tbody>
</table>

Tables 1 and 2 depicts acceptable containers respectively, and which audio and video streams can be included to a particular container. If the container uses both recommended and acceptable file formats, then the DPS handles the container as an acceptable file format.

**Table 1: Recommended container formats**

<table>
<thead>
<tr>
<th>Container</th>
<th>formatName</th>
<th>Audio</th>
<th>Video</th>
<th>formatRegistryKey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matroska v4#</td>
<td>video/x-matroska</td>
<td>FLAC, LPCM</td>
<td>FFV1</td>
<td>fmt/569</td>
</tr>
<tr>
<td>MPEG-2 (TS)</td>
<td>video/MP2T</td>
<td>AAC</td>
<td>AVC</td>
<td>fmt/585</td>
</tr>
<tr>
<td>MPEG-4</td>
<td>video/mp4</td>
<td>AAC</td>
<td>AVC</td>
<td>fmt/199</td>
</tr>
<tr>
<td>MXF</td>
<td>application/mxf</td>
<td>AAC, LPCM</td>
<td>AVC, JP2</td>
<td>-</td>
</tr>
<tr>
<td>MJPEG2000</td>
<td>video/mj2</td>
<td>LPCM</td>
<td>JP2</td>
<td>fmt/337</td>
</tr>
<tr>
<td>MOV</td>
<td>video/quicktime</td>
<td>AAC, LPCM</td>
<td>AVC, JP2</td>
<td>x-fmt/384</td>
</tr>
</tbody>
</table>

# Only Matroska version 4 is acceptable
### Table 2: Container formats acceptable for transfer

<table>
<thead>
<tr>
<th>Container</th>
<th>formatName</th>
<th>Audio</th>
<th>Video</th>
<th>format-RegistryKey</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASF</td>
<td>video/x-ms-asf</td>
<td>WMA</td>
<td>WMV</td>
<td>fmt/131</td>
</tr>
<tr>
<td>AVI</td>
<td>video/avi</td>
<td>MP3, LPCM</td>
<td>DV, MPG, H262</td>
<td>fmt/5</td>
</tr>
<tr>
<td>DV</td>
<td>video/dv</td>
<td>LPCM</td>
<td>DV</td>
<td>-</td>
</tr>
<tr>
<td>MPEG-1 (PS)</td>
<td>video/MP1S</td>
<td>MP3</td>
<td>MPG, H262</td>
<td>x-fmt/385</td>
</tr>
<tr>
<td>MPEG-2 (PS)</td>
<td>video/MP2P</td>
<td>MP3</td>
<td>MPG, H262</td>
<td>x-fmt/386</td>
</tr>
<tr>
<td>MPEG-2 (TS)</td>
<td>video/MP2T</td>
<td>MP3</td>
<td>MPG, H262</td>
<td>fmt/585</td>
</tr>
<tr>
<td>MPEG-4</td>
<td>video/mp4</td>
<td>MP3</td>
<td>MPG, H262</td>
<td>fmt/199</td>
</tr>
<tr>
<td>MXF</td>
<td>application/mxf</td>
<td>MP3</td>
<td>DV, MPG, H262</td>
<td>-</td>
</tr>
<tr>
<td>MOV</td>
<td>video/quicktime</td>
<td>MP3</td>
<td>DV, MPG, H262</td>
<td>x-fmt/384</td>
</tr>
</tbody>
</table>

### 6.1. Character Sets for Text Files

Character sets used in text files (CSV (2.1.1), XHTML (2.1.5), XML (2.1.4), HTML (2.1.5), plain text (2.1.8), GML (2.6.2), and KML (2.6.3)) must be expressed in the PREMIS formatName element using the charset parameter. Possible values for the charset parameter are ISO-8859-15 [ISO-8859-15], UTF-8, UTF-16, and UTF-32 [UNICODE].

**Example:**

```xml
<premis:format>
  <premis:formatDesignation>
    <premis:formatName>text/plain; charset=UTF-8</premis:formatName>
  </premis:formatDesignation>
  <premis:formatRegistry>
    <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
    <premis:formatRegistryKey>x-fmt/111</premis:formatRegistryKey>
  </premis:formatRegistry>
</premis:format>
```

A text file MUST NOT contain binary data. If any binary data should be included to the file, it MUST be Base64-encoded. Base64-encoded data is preserved only at bit level.

### 6.2. File Format Details

The national DPS supports only the file formats described in sections 2 and 3 of this specification. There are many unsupported file formats which have their own MIMes types. However, on closer scrutiny many of these file formats could in fact be supported by the DPS.

For example, there are text files like text/css, application/javascript, and application/json, which the DPS does accept as archiveable but only as text files. The DPS may validate them as text/plain files, but only if the file format in the SIP is text/plain and the real file format is expressed in the PREMIS formatName element using the alt-format parameter. Many XML-based file formats may be encoded in the same way.

The DPS does not utilize the value of the alt-format parameter during ingest.

**Examples:**

```xml
<premis:format>
  <premis:formatDesignation>
    <premis:formatName>text/plain; charset=ISO-8859-15; alt-format=text/css</premis:formatName>
  </premis:formatDesignation>
</premis:format>
```
If the given MIME type is text/plain, but the DPS identifies the file as having some other text format, the file will be accepted for preservation as a text/plain file. However, a note stating which file format was detected will be added to the validation report.
## ANNEX A. SUMMARY OF EVALUATION OF FILE FORMATS

The recommended file formats are evaluated using the criteria specified in [LAC]. The evaluation results are given in the table below.

<table>
<thead>
<tr>
<th>Content</th>
<th>File format</th>
<th>Openness/Transparency</th>
<th>Adoption as a Preservation Standard</th>
<th>Degree of Forward/Backward Compatibility</th>
<th>Degree of Protection Against File Corruption</th>
<th>Frequency of Version Releases</th>
<th>Dependencies/Interoperability</th>
<th>Standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEXT</td>
<td>Comma Separated Values (CSV)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Electronic Publications (EPUB)</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Extensible Hypertext Markup Language (XHTML)</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Extensible Markup Language (XML)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Hypertext Markup Language (HTML)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Open Document Format (ODF)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>PDF for long-term preservation (PDF/A)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Plain text</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>AUDIO</td>
<td>Audio Interchange File Format (AIFF), PCM encoded</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Broadcast Wave Format (BWF)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Free Lossless Audio Codec (FLAC)</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>MPEG-4 AAC – Advanced Audio Coding (AAC)</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Waveform Audio Format (WAV)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>VIDEO</td>
<td>Digital Picture Exchange (DPX)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>FF Video Codec 1 (FFV1)</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>JPEG 2000 sequence</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>MPEG-4</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>STILL IMAGE</td>
<td>Digital Negative (DN)</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>Joint photographic experts group (JPEG)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Joint photographic experts group JPEG 2000 (JP2)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Portable network graphics (PNG)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Scalable Vector Graphics (SVG)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Tagged Image File Format (TIFF)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>WEB ARCHIVE</td>
<td>Web Archive Format (WARC)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>DATABASES</td>
<td>Software Independent Archiving of Relational Databases (SIARD)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>GEODATA</td>
<td>Geographic Tagged Image File Format (GeoTiff)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

9 The specification is freely and publicly available. However, the file format development is controlled by a non-profit organisation (Swiss Federal Archives) that does not have an open membership policy.
<table>
<thead>
<tr>
<th>Content</th>
<th>File format</th>
<th>Openness/transparency</th>
<th>Adoption as a preservation standard</th>
<th>Degree of forward/backward compatibility</th>
<th>Degree of protection against file corruption</th>
<th>Frequency of version releases</th>
<th>Dependencies/Interoperability</th>
<th>Standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geography Markup Language (GML)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keyhole Markup Language (KML)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESEARCH DATA</td>
<td>SPSS Portable (POR)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>SCIENTIFIC SOFTWARE</td>
<td>Matlab</td>
<td>A/C</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hierarchical Data Format (HDF5)</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>
ANNEX B. HANDLING NON-RECOMMENDED AND NON-ACCEPTABLE FOR TRANSFER FORMATS

This annex specifies the possibilities and conditions for transferring files to the DPS which are not recommended or acceptable for transfer. Such cases can be roughly divided into three cases:

(1) The organization wants to preserve the original file format (not recommended or acceptable for transfer) in addition to the same file in recommended file format. This can be done in order to guarantee the authenticity of born digital material.

(2) The organization wants to preserve the original file format (not recommended or acceptable for transfer) in addition to the same file in a derived file format. This can be done because the derived file format is essential for usage and re-creation is not possible or very inconvenient.

(3) It is not possible or very inconvenient to migrate the file format to any recommended or acceptable for transfer format, but the file is valuable and worth preserving.

B.1. Requirements and Recommendations

Regardless of the case, the DPS offers bit-level preservation to other than recommended or acceptable for transfer format. Bit-level preservation only takes care of integrity and safe preservation. In such cases, responsibility for semantical/logical preservation is left totally to the partner organization.

Because semantical and logical preservation will be the responsibility of the partner organization, the DPS validates the content and ensures the partner organization’s capability for semantical and logical preservation by requiring a recommended file format (cases 1 and 2) or by applying strict criteria in order to guarantee content correctness (case 3). In cases 1 and 2, the DPS takes care of semantical and logical preservation of the file in a recommended file format as per normal.

Table 3 depicts the requirements and recommendations in the above-mentioned cases (+=recommended, *=required). If the file is not in a recommended or acceptable for transfer format, the file format must fulfill requirements in Table 3 and as many recommendations as possible. This ensures integrity and intelligibility in ingest and aims to preserve usefulness as long as possible. Requirements are checked beforehand for the recommended and acceptable for transfer file formats.

Table 3: Requirements and recommendations

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>DPS has accepted the file format for preservation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ The partner organization evaluates in cooperation with DPS whether the file is justified or if it could be migrated to some recommended format</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Acceptance is always tied to a contract</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>V2</td>
<td>The file format is supported by at least one generally available piece of software</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Software can be free or paid for. If a special software is needed to deal with the file, the name of the software and hyperlink to its home page must be provided.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ If software developed by the partner organization is the only software that supports the file format, the software and its source code must be included.</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
**FILE FORMATS – 1.10.0**

| V3 | The structure of the file format is documented  
|    | ▪ If the structure is unknown, for example in the case of vendor-specific closed formats, the documentation must include information on the file format that is as specific as possible. | + | + | * |
| V4 | If the interpretation of the file format requires metadata that is not inferential from the file, it must be saved in machine-interpretable format with files. | + | + | * |
| V5 | The file format can be identified by a machine. | + | + | * |
| V6 | The file format must have a MIME type. | + | + | * |
| V7 | The file format have a PRONOM PUID [PRONOM]  
|    | ▪ If there is no PUID already, it should be applied by following the PRONOM service guidelines. | + | + | + |
| V8 | The file format is widely used in the discipline. | + | + | + |
| V9 | The file format is standardized by and independent organization or community.  
|    | ▪ Standardized file format may be accepted as recommended or acceptable for transfer. This must be agreed with the DPS. | + | + | + |
| V10 | There is a validator for checking the file format’s correctness.  
|    | ▪ The validation is normally done in the DPS ingest. If the validator is not suitable for DPS, validation can be done in the partner organization’s own system before ingestion. | + | + | + |
| V11 | There is a script or guideline with which information can be read to some known software.  
|    | ▪ The script or guideline must be included to the preserved entirety. | + | + | + |

**B.2. Practical Guidelines**

If an organization wants to preserve the original file format (not recommended or acceptable for transfer) in addition to the same file in recommended file format, these files MUST be linked together in the METS/PREMIS level using the “migration” event. In this event, the file in a recommended format is the “outcome” of the event, and the original file is the “source”. Further, in the METS file element of the original file, the value of the USE attribute MUST be “fi-dpres-no-file-format-validation”. This is depicted in Figure 2.
If an organization wants to preserve the file in some converted format (not recommended or acceptable for transfer) in addition to the same file in recommended file format, these files MUST be linked together at the METS/PREMIS level using the “conversion” event. In this event, the file in a recommended format is the “source” of the event, and the converted file is the “outcome”. Further, in the METS file element of the converted file, the value of the USE attribute MUST be “fi-dpres-no-file-format-validation”. This is depicted in Figure 3.

If it is not possible or very inconvenient to migrate the file format to any recommended or acceptable for transfer format, but the file is valuable and worth preserving (case 3), in the METS file element of the file the value of the USE attribute MUST be “fi-dpres-file-format-identification”. This is depicted in Figure 4.
Figure 3: Preserving the original file format (not recommended or acceptable for transfer) in addition to the same file in a derived file format.

Figure 4: It is not possible to migrate the file format to any recommended format.
ANNEX C. LOCAL XML SCHEMA FILES

This Annex specifies what opportunities and conditions the partner organization may ingest XML schemas to DPS, which are relevant to the content to be preserved. If the content preserved uses XML schema to describe its structure, the schema file itself must be exploitable in the context of the content, both in validation and during preservation planning and preservation actions (such as migrations).

The DPS maintains a catalogue of general national and international XML schemas. Schemas in this catalogue are utilized when processing the content. If a content uses an XML schema that is not in the catalogue maintained by the DPS, the partner organization has two options:

1. The XML schema file is delivered separately to the DPS, which adds it to the schema catalog. This should always be agreed separately with the DPS, and in practice this can only be done for published schema and generally used by more organizations or entities. These schemas must also be published because the DPS’s schema catalogue is public.

2. XML schema files are provided as part of the submission information package. This approach is recommended for local, mainly schema files used internally by one organization. This option will be described in more detail below.

XML schema files that are delivered in the submission information package must be valid XML files and are validated and preserved in DPS among other content. They must be described in the metadata of the submission information package with technical metadata just like other content to be preserved (with the PREMIS metadata).

Since XML files refer to schemas using namespaces and URL paths that do not correspond to the internal relative file paths of the submission information package, the local XML schema that is delivered with the submission information package must be able to combine to XML files using the submission information package metadata. To do this, a PREMIS representation (premis:object xsi:type=”premis:representation”) containing links between schema files and XML files must be created to the submission information package metadata. This representation contains a single environmental block (premis:environment) containing links between schema and XML files. The environmental block MUST contain <premis:environmentPurpose> element with a value “xmlschemas”.

For each XML schema in the submission information package, a dedicated dependency block (premis:dependency) is created within the environmental block. The blocks describe how the XML schema sent along with the submission information package can be combined with schema paths used in XML files. In the dependency block <premis:dependencyName> element MUST provide the relative path of the local XML schema file from the root of the submission information package. The dependency block <premis:dependencyIdentifier> element describes the schema path used in XML files (xsi:schemaLocation or xsi:noNamespaceSchemaLocation).

Example

```xml
<mets:techMD ID="representation-001"
CREATED="2021-10-19T06:58:14.607036+00:00">
  <mets:mdWrap MDTYPE="PREMIS:OBJECT" MDTYPEVERSION="2.3">
    <mets:xmlData>
      <premis:object xsi:type="premis:representation">
        <premis:objectIdentifier>
          <premis:objectIdentifierType>local</premis:objectIdentifierType>
          <premis:objectIdentifierValue>
            my-representation-001
          </premis:objectIdentifierValue>
        </premis:objectIdentifier>
        <premis:environment>
          <premis:environmentPurpose>xml-schemas</premis:environmentPurpose>
```

In order to semantically differentiate the XML schema file(s) from the actual stored content in the submission information package, it is recommended that schema files be marked in the metadata of the submission information package as so-called accompanying content. This is done by adding XML schema files in the METS document <fileSec> element of the submission information package <fileGrp> element, which USE attribute value MAY be “fi-dpres-xml-schemas”.

**Example**

```
<fileGrp USE="fi-dpres-xml-schemas">
  <file ID="my-schema-file-001" ADMID="schema001-techmd">
    [...]
  </file>
</fileGrp>
```

The PREMIS representation with schema links MAY be put on its own structural map (<mets:structMap>), whose <div> attribute TYPE of the first element becomes “fi-dpres-supplementary”.

**Example**

```
<structMap TYPE="logical">
  <div TYPE="fi-dpres-supplementary">
    <div TYPE="fi-dpres-xml-schemas" ADMID="representation-001">
      <fptr FILEID="my-schema-file-001"/>
    </div>
  </div>
</structMap>
```

The XML schema file in the <fileSec> element and structural map are depicted in Figure 5.
Figure 5: XML schema file as part of the submission information package
REFERENCES


FILE FORMATS – 1.10.0


FILE FORMATS – 1.10.0


[SMPTE_314] Society of Motion Pictures & Television Engineers. Standard for Television - Data Structure for DV-Based Audio, Data and Compressed Video 625 and 50 Mb/s. SMPTE 314-2005

[SMPTE_370] Society of Motion Pictures & Television Engineers. Standard for Television - Data Structure for DV-Based Audio, Data and Compressed Video at 100 Mb/s 1080/60i, 1080/50i, 720/60p, 720/50p. SMPTE 370-2013.


