FILE FORMATS
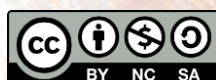
v1.10.0

This document forms part of the Ministry of Education and Culture's
Open science and digital cultural heritage entity.

CONTENT

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 Hardware, software, and file formats will become outdated, but the information must remain understandable.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 spefication 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[1]. 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[2]. 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[3] and the website hosted by the Library of Congress on file format sustainability[4]. 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

---

[2] https://digitalpreservation.fi/specifications/vocabularies/unknown-values
[3] https://www.nationalarchives.gov.uk/pronom/
[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 formatName 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 formatName 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 formatName 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 formatName 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 formatName 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 formatName 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]; <u>PRONOM: x-fmt/135</u>

Vocabulary in the PREMIS formatName 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]; <u>PRONOM: fmt/527</u>

Vocabulary in the PREMIS formatName 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]; <u>PRONOM: fmt/279</u>

This file format can be included in the following containers:

> Matroska [MATROSKA]

Vocabulary in the PREMIS formatName 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 formatName 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 formatName 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 formatName 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]; <u>PRONOM: fmt/199</u>

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 formatName 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]; <u>PRONOM: fmt/438</u>
> DNG 1.4.0.0; [DNG_14]; <u>PRONOM: fmt/730</u>
> DNG 1.5.0.0; [DNG_15];

Vocabulary in the PREMIS formatName 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]; <u>PRONOM: fmt/42</u>
> JPEG part 1 version 1.01; [ISO_10918-1]; <u>PRONOM: fmt/43</u>
> JPEG part 1 version 1.02; [ISO_10918-1]; <u>PRONOM: fmt/44</u>
> JPEG/EXIF version 2.0; <u>PRONOM; x-fmt/398</u>
> JPEG/EXIF version 2.1; <u>PRONOM; x-fmt/390</u>
> JPEG/EXIF version 2.2; <u>PRONOM; x-fmt/391</u>
> JPEG/EXIF version 2.2.1; <u>PRONOM; fmt/645</u>
> JPEG/EXIF version 2.3
> JPEG/EXIF version 2.3.1
> JPEG/EXIF version 2.3.2

Vocabulary in the PREMIS formatName semantic unit:

image/jpeg

Metadata schema for the mandatory technical metadata:

MIX

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

Vocabulary in the PREMIS formatName semantic unit:

image/jp2

Metadata schema for the mandatory technical metadata:

MIX

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

Vocabulary in the PREMIS formatName semantic unit:

image/png

Metadata schema for the mandatory technical metadata:

MIX

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[5]:

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

Vocabulary in the PREMIS formatName semantic unit:

image/svg+xml

Metadata schema for the mandatory technical metadata:

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 formatName 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 formatName 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 formatName 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 formatName 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 formatName 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 formatName 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, versio 7; PRONOM: fmt/806
> Matlab versio 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[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.

---

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

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

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]; <u>PRONOM: x-fmt/136</u>

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]; <u>PRONOM: fmt/134</u>
> MPEG-2 layer-3; [ISO_13818-3]; <u>PRONOM: fmt/134</u>

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 formatName 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 formatName 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 formatName 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 formatName 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:

```
<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-b316-49cc6af1cba2
          </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:

```
<premis:relationship>
  <premis:relationshipType>structural</premis:relationshipType>
  <premis:relationshipSubType>includes</premis:relationshipSubType>
  <premis:relatedObjectIdentification>
    <premis:relatedObjectIdentifierType>UUID
    </premis:relatedObjectIdentifierType>
    <premis:relatedObjectIdentifierValue>
       387e9e6c-4596-4baf-b316-49cc6af1cba2
    </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:

```
<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.
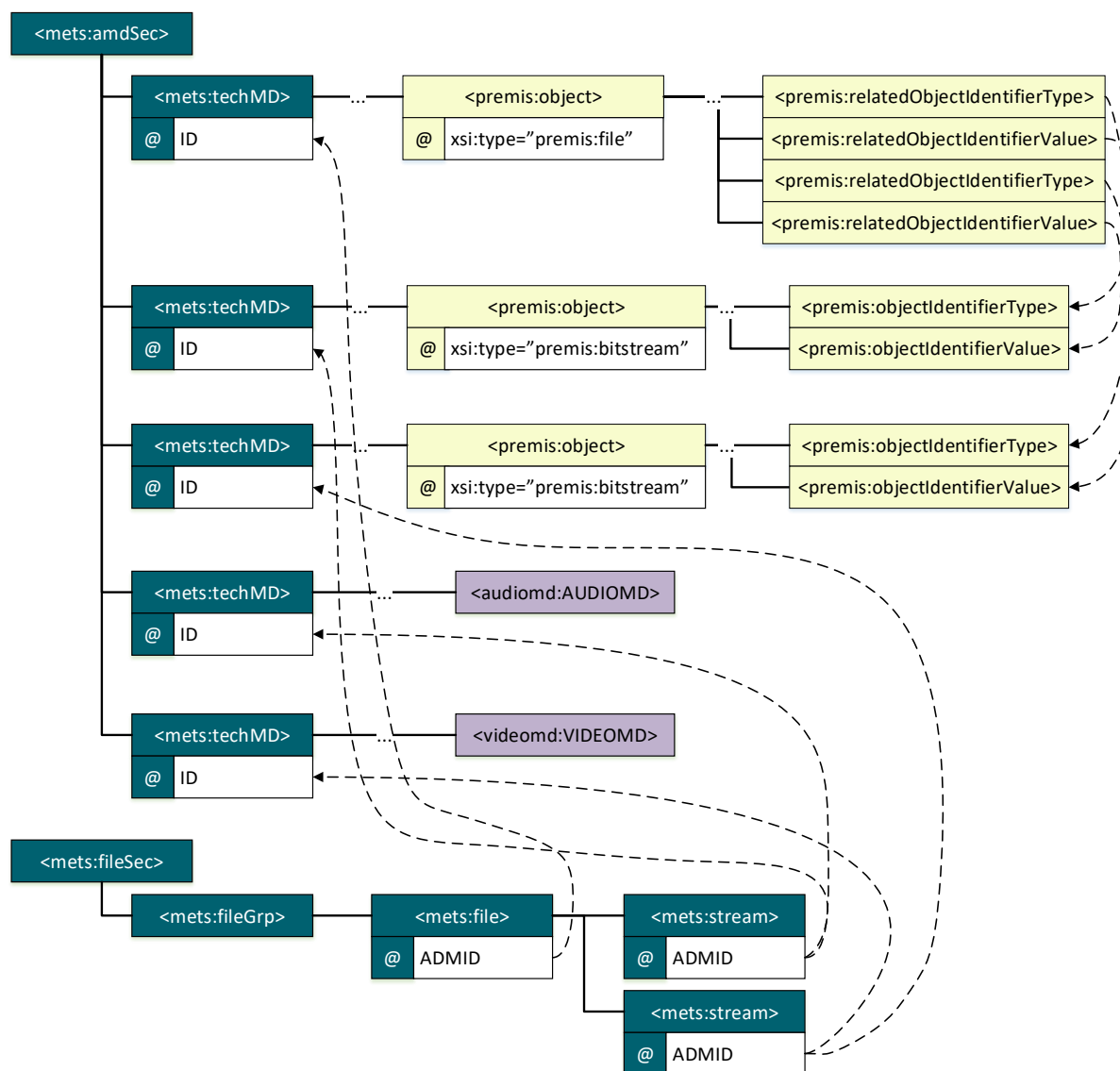
Figure 1: Connecting container, its component parts and administrative metadata

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: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:flatFileTypes>
               <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:flatFileTypes>
            <addml:fieldTypes>
               <addml:fieldType name="str">
                  <addml:dataType>string</addml:dataType>
```

```
            </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	Repeatable
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.

Element/attribute	Occurrence/ obligation	Regulations and recommendations
<audioMD>	ET, P	The root element must contain <fileData /> and <audioInfo /> elements. In the audioMD schema these elements are optional.
-<fileData>	ET, P	Mandatory element in the DPS.
--<audioDataEncoding>	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.
--<bitPerSample>	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.
--<compression>	T, P	Optional in the audioMD schema, but mandatory in the DPS.
---<codecCreatorApp>	ET, P	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.
---<codecCreatorAppVersion>	ET, P	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 or for software that does not have versioning.
---<codecName>	ET, P	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.

Element/attribute	Occurrence/ obligation	Regulations and recommendations
---<codecQuality>	ET, P	Optional in the audioMD schema, but mandatory in the DPS.
--<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

```
<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.

Element/attribute	Occurrence/obligation	Regulations and recommendations
<videoMD>	ET, P	The root element must contain <fileData /> and <videoInfo /> elements. In the videoMD schema these elements are optional.
-<fileData>	ET, P	
--<duration>	ET, P	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.
--<dataRate>	ET,P	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.
--<bitsPerSample>	T, P	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.
--<color>	T, P	Optional in the videoMD schema, but mandatory in the DPS.
--<compression>	T, P	Optional in the videoMD schema, but mandatory in the DPS.
---<codecCreatorApp>	ET, P	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.

Element/attribute	Occurrence/obligation	Regulations and recommendations
---<codecCreatorAppVersion>	ET, P	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.
---<codecName>	EP, P	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.
---<codecQuality>	ET, P	Optional in the videoMD schema, but mandatory in the DPS.
--<dataRateMode>	T, P	Optional in the videoMD schema, but mandatory in the DPS.
--<frameRate>	ET, P	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.
--<frame>	T, P	Optional in the videoMD schema, but mandatory in the DPS.
---<pixelsHorizontal>	ET, P	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.
---<pixelsVertical>	ET, P	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.
---<PAR>	ET, P	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.
---<DAR>	ET, P	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.

Element/attribute	Occurrence/obligation	Regulations and recommendations
--<sampling>	T, P	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.
--<signalFormat>	T, P	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.
--<sound>	T, P	Optional in the videoMD schema, but mandatory in the DPS.

5.3.1. Example

```
<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:codecCreatorAppVersion>10</vmd:codecCreatorAppVersion>
       <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:sampling>4:2:2</vmd:sampling>
    <vmd:signalFormat>PAL</vmd:signalFormat>
    <vmd:sound>No</vmd:sound>
  </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 schema[7].

Element/attribute	Occurrence/ obligation	Regulations and recommendations
\<BasicDigitalObjectInformation\>	P, ET	The element is optional in the MIX schema, but mandatory in the DPS
\<BasicImageInformation\>	P, ET	The element is optional in the MIX schema, but mandatory in the DPS
\<ImageAssessmentMetadata\>	P, ET	The element is optional in the MIX schema, but mandatory in the DPS
\<ObjectIdentifier\>	ES, ET	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.
\<fileSize\>	ES, ET	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.
\<FormatDesignation\>	ES, ET	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.
\<FormatRegistry\>	ES, ET	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.
\<Fixity\>	ES, T	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.

[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:IccProfile>
          <mix:iccProfileName>Adobe RGB</mix:iccProfileName>
          <mix:iccProfileVersion>1998</mix:iccProfileVersion>
          <mix:iccProfileURI>
            https://www.adobe.com/digitalimag/adobergb.html
          </mix:iccProfileURI>
        </mix:IccProfile>
      </mix:ColorProfile>
     </mix:PhotometricInterpretation>
    </mix:BasicImageCharacteristics>
    <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:BasicImageInformation>
  <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: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 1: File formats

File format	formatName	formatVersion	formatRegistryKey
CSV	text/csv	-	x-fmt/18
EPUB	application/epub+zip	2.0.1	fmt/483
EPUB	application/epub+zip	3.0.0	fmt/483
EPUB	application/epub+zip	3.0.1	fmt/483
EPUB	application/epub+zip	3.2	fmt/483
XHTML	application/xhtml+xml	1.0	fmt/102
XHTML	application/xhtml+xml	1.1	fmt/103
XHTML	application/xhtml+xml	5.0	fmt/471
XHTML	application/xhtml+xml	5.1	-
XHTML	application/xhtml+xml	5.2	-
XML	text/xml	1.0	fmt/101
XML	text/xml	1.1	-
HTML	text/html	4.01	fmt/100
HTML	text/html	5.0	fmt/471
HTML	text/html	5.1	-
HTML	text/html	5.2	-
ODT	application/vnd.oasis.opendocument.text	1.0	fmt/136
ODT	application/vnd.oasis.opendocument.text	1.1	fmt/290
ODT	application/vnd.oasis.opendocument.text	1.2	fmt/291
ODT	application/vnd.oasis.opendocument.text	1.3	-
ODS	application/vnd.oasis.opendocument.spreadsheet	1.0	fmt/137
ODS	application/vnd.oasis.opendocument.spreadsheet	1.1	fmt/294
ODS	application/vnd.oasis.opendocument.spreadsheet	1.2	fmt/295
ODS	application/vnd.oasis.opendocument.spreadsheet	1.3	-
ODP	application/vnd.oasis.opendocument.presentation	1.0 1.1 1.2	fmt/138 fmt/293
ODP	application/vnd.oasis.opendocument.presentation	1.1	fmt/292
ODP	application/vnd.oasis.opendocument.presentation	1.2	fmt/293
ODP	application/vnd.oasis.opendocument.presentation	1.3	-
ODG	application/vnd.oasis.opendocument.graphics	1.0	fmt/139
ODG	application/vnd.oasis.opendocument.graphics	1.1	fmt/296
ODG	application/vnd.oasis.opendocument.graphics	1.2	fmt/297
ODG	application/vnd.oasis.opendocument.graphics	1.3	-

File format	formatName	formatVersion	formatRegistryKey
ODF	application/vnd.oasis.opendocument.formula	1.0	-
ODF	application/vnd.oasis.opendocument.formula	1.2	-
ODF	application/vnd.oasis.opendocument.formula	1.3	-
PDF/A	application/pdf	A-1a	fmt/95
PDF/A	application/pdf	A-1b	fmt/354
PDF/A	application/pdf	A-2a	fmt/476
PDF/A	application/pdf	A-2b	fmt/477
PDF/A	application/pdf	A-2u	fmt/478
PDF/A	application/pdf	A-3a	fmt/479
PDF/A	application/pdf	A-3b	fmt/480
PDF/A	application/pdf	A-3u	fmt/481
Plain text	text/plain	-	x-fmt/111
AIFF	audio/x-aiff	1.3	x-fmt/135
BWF	audio/x-wav	2	fmt/527
FLAC	audio/flac	1.2.1	fmt/279
LPCM	audio/L16, audio/L8, audio/L20, audio/L24	-	-
AAC	audio/mp4	-	fmt/199
WAV	audio/x-wav	-	fmt/141
DPX	image/x-dpx	2.0	fmt/541
FFV1	video/x-ffv	3	-
JPEG2000	video/jpeg2000	-	x-fmt/392
AVC	video/mp4	-	fmt/199
DNG	image/tiff	1.3	fmt/438
DNG	image/tiff	1.4	fmt/730
DNG	image/tiff	1.5	-
JPEG	image/jpeg	1.00	fmt/42
JPEG	image/jpeg	1.01	fmt/43
JPEG	image/jpeg	1.02	fmt/44
JPEG/EXIF	image/jpeg	2.0	x-fmt/398
JPEG/EXIF	image/jpeg	2.1	x-fmt/390
JPEG/EXIF	image/jpeg	2.2	x-fmt/391
JPEG/EXIF	image/jpeg	2.2.1	fmt/645
JPEG/EXIF	image/jpeg	2.3	-
JPEG/EXIF	image/jpeg	2.3.1	-
JPEG/EXIF	image/jpeg	2.3.2	-
JP2	image/jp2	-	x-fmt/392
SVG	image/svg+xml	1.1	fmt/92
TIFF	image/tiff	6.0	fmt/353
PNG	image/png	1.2	fmt/13
WARC	application/warc	1.0	fmt/289
GEOTIFF	image/tiff	1.0	fmt/155
GML	application/gml+xml	3.2.1	fmt/1047
KML	application/vnd.google-earth.kml+xml	2.3	fmt/244
SIARD	application/x-siard	2.1.1	-
SIARD	application/x-siard	2.2	-
POR	application/x-spss-por	-	fmt/997
Matlab	application/matlab	7	fmt/806
Matlab	application/matlab	7.3	fmt/828

File format	formatName	formatVersion	formatRegistryKey
HDF5	application/x-hdf5	1.10	fmt/807, fmt/286, fmt/287
MS Word	application/msword	97-2003	fmt/40
MS Word	application/vnd.openxmlformats-officedocument.wordprocessingml.document	2007 onwards	fmt/412
MS Excel	application/vnd.ms-excel	8	fmt/61
MS Excel	application/vnd.ms-excel	8X	fmt/62
MS Excel	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	2007 onwards	fmt/214
MS Powerpoint	application/vnd.ms-powerpoint	97-2003	fmt/126
MS Powerpoint	application/vnd.openxmlformats-officedocument.presentationml.presentation	2007 onwards	fmt/215
PDF	application/pdf	1.2	fmt/16
PDF	application/pdf	1.3	fmt/17
PDF	application/pdf	1.4	fmt/18
PDF	application/pdf	1.5	fmt/19
PDF	application/pdf	1.6	fmt/20
PDF	application/pdf	1.7	fmt/276
AIFF-C	audio/x-aiff	-	x-fmt/136
MP3	audio/mpeg	1	fmt/134
MP3	audio/mpeg	2	fmt/134
WMA	audio/x-ms-wma	9	fmt/132
DV	video/dv	-	x-fmt/152
MPEG	video/mpeg	1	fmt/649
MPEG	video/mpeg	2	fmt/640
MPEG	video/mpeg	4	-
WMV	video/x-ms-wmv	9	fmt/133
EPS	application/postscript	3.0	fmt/124
GIF	image/gif	1987a	fmt/3
GIF	image/gif	1989a	fmt/4

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

Container	formatName	Audio	Video	format-RegistryKey
Matroska v4[8]	video/x-matroska	FLAC, LPCM	FFV1	fmt/569
MPEG-2 (TS)	video/MP2T	AAC	AVC	fmt/585
MPEG-4	video/mp4	AAC	AVC	fmt/199
MXF	application/mxf	AAC, LPCM	AVC, JP2	-
MJPEG2000	video/mj2	LPCM	JP2	fmt/337
MOV	video/quicktime	AAC, LPCM	AVC, JP2	x-fmt/384

---

[8] Only Matroska version 4 is acceptable

Table 2: Container formats acceptable for tranfer

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

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:

```
<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 archivable 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:

```
<premis:format>
  <premis:formatDesignation>
    <premis:formatName>
      text/plain; charset=ISO-8859-15; alt-format=text/css
    </premis:formatName>
  </premis:formatDesignation>
</premis:format>
```

```
<premis:format>
  <premis:formatDesignation>
    <premis:formatName>
      text/xml; charset=UTF-8; alt-format=application/mets+xml
    </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.

Content	File format	Openness/transparency	Adoption as a preservation standard	Degree of forward/backward compatibility	Degree of protection against file corruption	Frequency of version releases	Dependencies/Interoperability	Standardization
TEXT	Comma Separated Values (CSV)	A	A	A		A	A	A
	Electronic Publications (EPUB)	A	B	C		A	A	A
	Extensible Hypertext Markup Language (XHTML)	A	B	B		A	B	A
	Extensible Markup Language (XML)	A	A	A		A	A	A
	Hypertext Markup Language (HTML)	A	A	B		A	A	A
	Open Document Format (ODF)	A	A	B		B	A	A
	PDF for long-term preservation (PDF/A)	A€	A	B		A	A	A
	Plain text	A€	A	B		A	A	A
AUDIO	Audio Interchange File Format (AIFF), PCM encoded	A	A	A		A	A	A
	Broadcast Wave Format (BWF)	A	A	A		A	A	A
	Free Lossless Audio Codec (FLAC)	A	B	A	A	A	A	A
	MPEG-4 AAC – Advanced Audio Coding (AAC)	A	B	A			A	A
	Waveform Audio Format (WAV)	A	A	A		A	A	A
VIDEO	Digital Picture Exchange (DPX)	A	A/C		A	A	A	A
	FF Video Codec 1 (FFV1)	A	B	B	A	B	A	A
	JPEG 2000 sequence	A€	A	A	A	A	A	A
	MPEG-4	A	B	A		A	A	A
STILL IMAGE	Digital Negative (DNG)	A	B	B		A	A	C
	Joint photographic experts group (JPEG)	A€	A			A	A	A
	Joint photographic experts group JPEG 2000 (JP2)	A€	A			A	A	A
	Portable network graphics (PNG)	A€	A	A		A	A	A
	Scalable Vector Graphics (SVG)	A	A	B	B	B	A	A
	Tagged Image File Format (TIFF)	A	A			A	A	A
WEB ARCHIVE	Web ARchive Format (WARC)	A	A		B	A	A	A
DATABASES	Software Independent Archiving of Relational Databases (SIARD)	A[9]	B	B	C	B	B	A
GEODATA	Geographic Tagged Image File Format (GeoTiff)	A	A			A	A	A

---

[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.

Content	File format	Opennes/transparency	Adoption as a preservation standard	Degree of forward/backward compatibility	Degree of protection against file corruption	Frequency of version releases	Dependencies/Interoperability	Standardization
	Geography Markup Language (GML)							
	Keyhole Markup Language (KML)							
RESEARCH DATA	SPSS Portable (POR)	A€	A	A	A	A	A	C
SCIENTIFIC SOFTWARE	Matlab	A/C	B	B	C	B	B	C
	Hierarchical Data Format (HDF5)	A	B	B	A	A	B	A

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

		(1)	(2)	(3)
V1	DPS has accepted the file format for preservation ▪ The partner organization evaluates in cooperation with DPS whether the file is justified or if it could be migrated to some recommended format ▪ Acceptance is always tied to a contract	*	*	*
V2	The file format is supported by at least one generally available piece of software ▪ Softwarecan 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. ▪ 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.	+	+	*

		(1)	(2)	(3)
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.
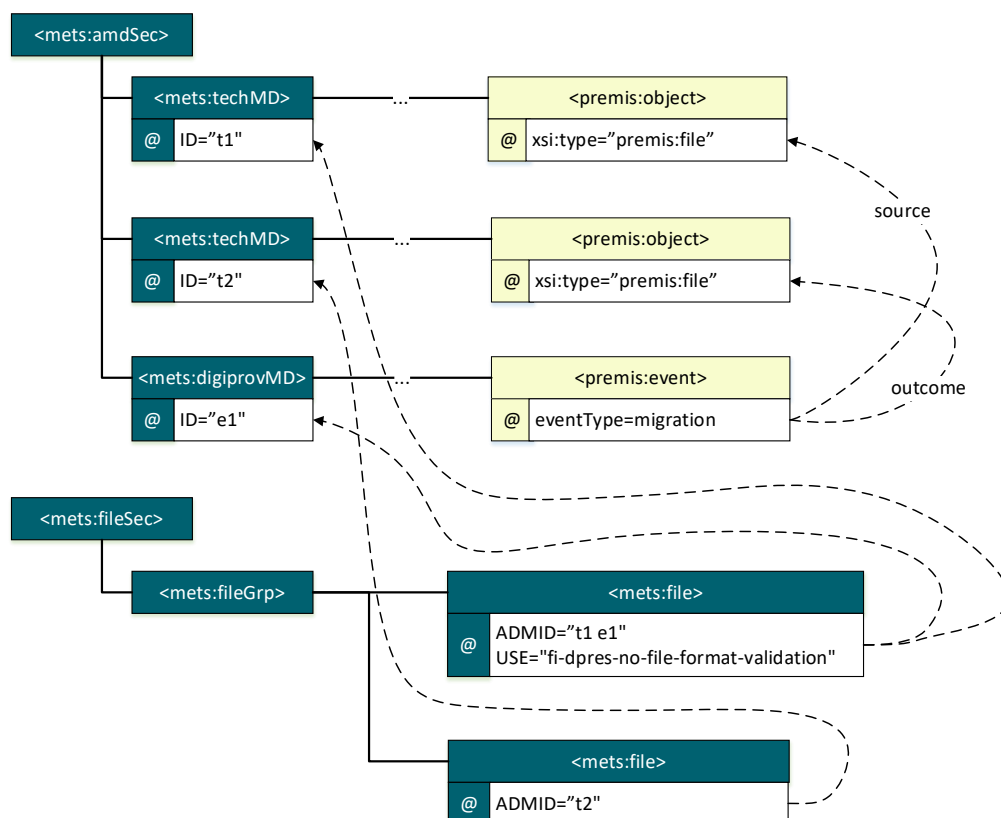
Figure 2: Preserving the original file format (not recommended or acceptable for transfer) in addition to the same file in recommended file format

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.
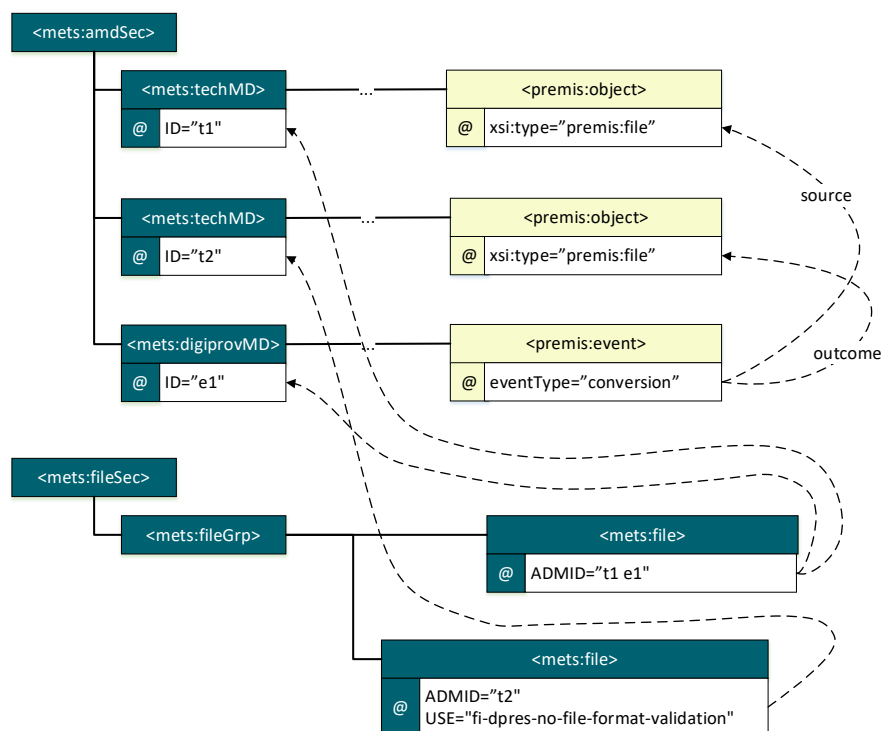
Figure 3: Preserving the original file format (not recommended or acceptable for transfer) in addition to the same file in a derived file format.
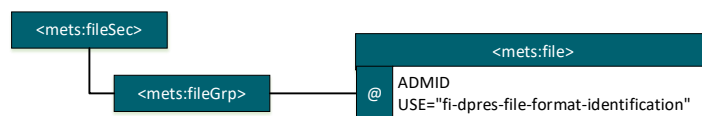

Figure 4: It is not possible to migrate the file format to any recommended format

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 "xml-schemas".

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

```
<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>
```

```
      <premis:dependency>
       <premis:dependencyName>
        file:///schemas/local_schema1.xsd
       </premis:dependencyName>
       <premis:dependencyIdentifier>
        <premis:dependencyIdentifierType>
         URI
        </premis:dependencyIdentifierType>
        <premis:dependencyIdentifierValue>
         http://myschemas.fi/local_schema.xsd
        </premis:dependencyIdentifierValue>
       </premis:dependencyIdentifier>
      </premis:dependency>
       <premis:dependency>
       <premis:dependencyName>
        file:///schemas/local_schema2.xsd
       </premis:dependencyName>
       <premis:dependencyIdentifier>
        <premis:dependencyIdentifierType>
         URI
        </premis:dependencyIdentifierType>
        <premis:dependencyIdentifierValue>
         file:///c/my-local-harddrive/my-schemas/local_schema.xsd
        </premis:dependencyIdentifierValue>
       </premis:dependencyIdentifier>
      </premis:dependency>
     </premis:environment>
    </premis:object>
   </mets:xmlData>
  </mets:mdWrap>
 </mets:techMD>
```

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

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

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

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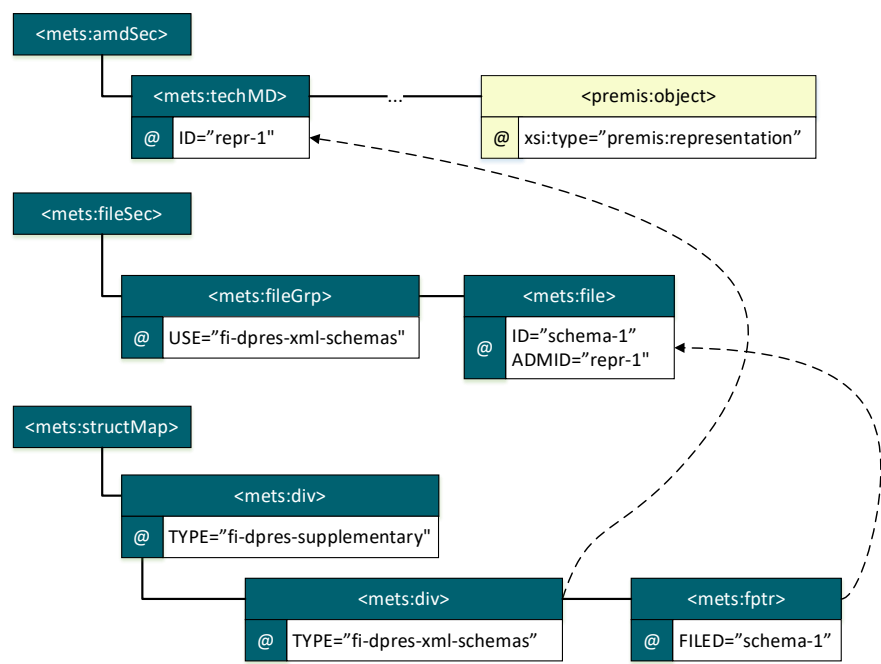
Figure 5: XML schema file as part of the submission information package

REFERENCES

[ADDML]   Archival Data Description Markup Language. https://www.arkivverket.no/forvaltning-og-
                 utvikling/regelverk-og-standarder/andre-arkivstandarder/addml-archival-data-description-markup-
                 language

[AIFF]   Audio Interchange File Format: "AIFF". A Standard for Sampled Sound Files. Version 1.3. Apple
                 Computer, Inc. http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf

[AIFF-C]   Audio Interchange File Format AIFF-C. A revision to include compressed audio data. Apple
                 Computer, Inc. Draft 26.8.1991. http://www-
                 mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-C.9.26.91.pdf

[ASF]   Advanced Systems Format (ASF) Specification. Revision 01.20.03. Microsoft Corporation, December
                 2004. https://download.microsoft.com/download/7/9/0/790fecaa-f64a-4a5e-a430-
                 0bccdab3f1b4/ASF_Specification.doc

[AUDIOMD]   Audio Technical Metadata Extension Schema
                 https://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html

[AVI]   AVI (Audio Video Interleave). Library of Congress. https://docs.microsoft.com/en-
                 us/windows/win32/directshow/avi-riff-file-reference

[BWF]   EBU Tech 3285 - Specification of the Broadcast Wave Format (BWF) - A format for audio data files in
                 broadcasting. Version 2.0. Geneva May 2011. https://tech.ebu.ch/docs/tech/tech3285.pdf

[DNG_13]   Digital Negative (DNG) Specification, Version 1.3.0.0. Adobe Systems Incorporated. June 2009.
                 https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec.pdf
                 [DNG_14]   Digital Negative (DNG) Specification, Version 1.4.0.0. Adobe Systems Incorporated.
                 June 2012.
                 https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf

[DNG_15]   Digital Negative (DNG) Specification, Version 1.5.0.0. Adobe Systems Incorporated. May 2019.
                 https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.5.0.0.pdf

[EPS]   Encapsulated PostScript File Format Specification. Adobe Developer Support. Version 3.0. Adobe
                 Systems Incorporated. 1.5.1992.
                 https://www.adobe.com/content/dam/acom/en/devnet/actionscript/articles/5002.EPSF_Spec.pdf
                 pdf

[EPUB_2.0.1]   EPUB version 2.0.1. International Digital Publishing Forum. http://idpf.org/epub/201

[EPUB_3.0.0]   EPUB version 3.0.0. International Digital Publishing Forum. http://idpf.org/epub/30. In November
                 2014, EPUB 3.0 was published by the International Standards Organization as ISO/IEC TS 30135
                 (parts 1-7).

[EPUB_3.0.1]   EPUB version 3.0.1. International Digital Publishing Forum. http://idpf.org/epub/301. In January
                 2020, EPUB 3.0.1 was published by the International Standards Organization as ISO/IEC 23736 (parts
                 1-6).

[EPUB_3.2]   EPUB 3.2. Final Community Group Specification 08 May 2019.
                 https://www.w3.org/publishing/epub/epub-spec.html

[FI_API]   Nation al Digital Preservation Services of Finland. Interfaces for Digital Preservation Services.
                 urn:nbn:fi-fe2020100578098

[FI_META]   National Digital Preservation Services of Finland. Metadata Requirements and Preparing Content for
                 Digital Preservation. urn:nbn:fi-fe2020100578094

[FLAC]   FLAC Format Specification. https://xiph.org/flac/format.html

[GEOTIFF]   GeoTIFF Format Specification, GeoTIFF Revision 1.0, Specification Version: 1.8.2, November 1995.
                 http://geotiff.maptools.org/spec/geotiffhome.html

[GIF_87a]   GIF(tm). GRAPHICS INTERCHANGE FORMAT(tm). A standard defining a mechanism for the storage and transmission of raster-based graphics information. CompuServe Incorporated. 15.6.1987. https://www.w3.org/Graphics/GIF/spec-gif87.txt

[GIF_89a]   GRAPHICS INTERCHANGE FORMAT(tm), Version 89a. CompuServe Incorporated, Columbus, Ohio. 31.7.1990. https://www.w3.org/Graphics/GIF/spec-gif89a.txt

[HDF5]   The HDF Group: HDF5 file format home page. https://docs.hdfgroup.org/hdf5/develop/

[HTML5]   HTML5. A vocabulary and associated APIs for HTML and XHTML. W3C Recommendation 28 October 2014. https://www.w3.org/TR/html5/

[HTML5.1]   HTML 5.1 2nd Edition. W3C Recommendation, October 2017. https://www.w3.org/TR/html51/

[HTML5.2]   HTML 5.2. W3C Recommendation, December 2017. https://www.w3.org/TR/html52/

[IEC_61834]   International Electrotechnical Commission. Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems). IEC 61834:1998-2008

[ISO_10918-1]   International Organization for Standardization. Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines. ISO/IEC 10918-1:1994

[ISO_11172-1]   International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 1: Systems. ISO/IEC 11172-1:1993

[ISO_11172-2]   International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 2: Video. ISO/IEC 11172-2:1993

[ISO_11172-3]   International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio. ISO/IEC 11172-3:1993

[ISO_13818-1]   International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information: Systems. ISO/IEC 13818-1:2019

[ISO_13818-2]   International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information: Video. ISO/IEC 13818-2:2013

[ISO_13818-3]   International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio. ISO/IEC 13818-3:1998

[ISO_14496-10]   International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding. ISO/IEC 14496-10:2020

[ISO_14496-14]   International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 14: MP4 file format. ISO/IEC 14496-14:2020

[ISO_14496-3]   International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 3: Audio. ISO/IEC 14496-3:2019

[ISO_15444-1]   International Organization for Standardization. Information technology — JPEG 2000 image coding system: Core coding system. ISO/IEC 15444-1:2019

[ISO_15444-3]   International Organization for Standardization. Information technology — JPEG 2000 image coding system: Motion JPEG 2000. ISO/IEC 15444-3:2007

[ISO_15445]   International Organization for Standardization. Information technology — Document description and processing languages — HyperText Markup Language (HTML). ISO/IEC 15445:2000

[ISO_15948]   International Organization for Standardization. Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2004

[ISO_19005]   International Organization for Standardization. Document management —Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1). ISO 19005-1:2005

[ISO_19005-2]   International Organization for Standardization. Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2). ISO 19005-2:2011

[ISO_19005-3]   International Organization for Standardization. Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3). ISO 19005-3:2012

[ISO_19136]   International Organization for Standardization. Geographic information -- Geography Markup Language (GML). ISO 19136-1:2020

[ISO_26300:2006]   International Organization for Standardization. Information technology — Open Document Format for Office Applications (OpenDocument) v1.0. ISO/IEC 26300:2006

[ISO_26300:2015]   International Organization for Standardization. Information technology — Open Document Format for Office Applications (OpenDocument) v1.2. ISO/IEC 26300:2015

[ISO_28500]   International Organization for Standardization. Information and documentation — WARC file format. ISO 28500:2009

[ISO_29500-1]   International Organization for Standardization. Information technology — Document description and processing languages — Office Open XML File Formats — Part 1: Fundamentals and Markup Language Reference. ISO/IEC 29500-1:2016

[ISO_32000-1]   International Organization for Standardization. Document management — Portable document format — Part 1: PDF 1.7. ISO 32000-1:2008

[ISO_8859-15]   International Organization for Standardization. Information technology — 8-bit single-byte coded graphic character sets — Part 15: Latin alphabet No. 9. ISO/IEC 8859-15:1999

[KML]   Open Geospatial Consortium. OGC KML 2.3. http://www.opengis.net/doc/IS/kml/2.3

[LAC]   Library and Archives Canada. Trusted Digital Repository, Local Digital Format Registry (LDFR), File Format Guidelines for Preservation and Long-term Access, Version 1.0

[LPCM]   Linear Pulse Code Modulated Audio (LPCM). Library of Congress. February 19, 2008. https://www.loc.gov/preservation/digital/formats/fdd/fdd000011.shtml

[MATROSKA]   Matroska Specifications. IETF Draft. October 2019. https://datatracker.ietf.org/doc/draft-ietf-cellar-matroska/

[METS]   Metadata Encoding and Transfer Standard. https://www.loc.gov/standards/mets/

[MIX]   NISO Data Dictionary Technical Metadata for Digital Still Images. https://www.loc.gov/standards/mix/

[MOV]   QuickTime File Format. Classic Version of the QuickTime File Format Specification. Apple Computer, Inc. 1.3.2001. https://developer.apple.com/standards/qtff-2001.pdf

[ODF_1.3]   Open Document Format for Office Applications (OpenDocument) Version 1.3 Committee Specification 01. December 2019. https://www.oasis-open.org/news/announcements/open-document-format-for-office-applications-opendocument-v1-3-from-the-opendocum

[PDF_1.2]   Portable Document Format Reference Manual. Version 1.2. Adobe Systems Incorporated. November 12, 1996. https://web.archive.org/web/19961219053721/http://www.adobe.com:80/supportservice/devrelations/PDFS/TN/PDFSPEC.PDF

[PDF_1.3]   PDF reference, second edition. Adobe Portable Document Format. Version 1.3. Adobe Systems Incorporated. July, 2000. https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdfs/PDFReference13.pdf

[PDF_1.4]   PDF reference, third edition. Adobe Portable Document Format. Version 1.4. Adobe Systems Incorporated. December, 2001. https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference.pdf

[PDF_1.5]	PDF reference, fourth edition. Adobe Portable Document Format. Version 1.5. Adobe Systems Incorporated. Aug, 2003. https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdf_reference_archive/PDFReference15_v6.pdf
[PDF_1.6]	PDF reference, fifth edition. Adobe Portable Document Format. Version 1.6. Adobe Systems Incorporated. Nov, 2004. https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdf_reference_archive/PDFReference16.pdf
[PREMIS]	Data Dictionary for Preservation Metadata: PREMIS version 2.2. July 2012. https://www.loc.gov/standards/premis/v2/premis-2-2.pdf
[PRONOM]	The National Archives, UK. The technical registry PRONOM. https://www.nationalarchives.gov.uk/aboutapps/pronom/
[RFC_2046]	Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types. Request for Comments: 2046. Request for Comments: 2046 November 1996. https://datatracker.ietf.org/doc/rfc2046/
[RFC_2119]	Key words for use in RFCs to Indicate Requirement Levels. https://datatracker.ietf.org/doc/rfc2119/
[RFC_4180]	Common Format and MIME Type for Comma-Separated Values (CSV) Files. Request for Comments: 4180. October 2005. https://datatracker.ietf.org/doc/rfc4180/
[RFC_9043]	FFV1 Video Coding Format Version 0, 1, and 3. https://datatracker.ietf.org/doc/rfc9043/
[SIARD2.1.1]	SIARD Format Specification 2.1.1, 2020. https://www.bar.admin.ch/dam/bar/en/dokumente/kundeninformation/siard_formatbeschreibung.pdf.download.pdf/siard_format_description.pdf
[SIARD2.2]	SIARD Format Specification 2.2. 2021. https://siard.dilcis.eu/SIARD%202.2/SIARD%202.2.pdf
[SMPTE_268]	Society of Motion Picture and Television Engineers. For File Format for Digital Moving-Picture Exchange (DPX), Version 2.0. SMPTE 268:2014
[SMPTE_314]	Society of Motion Pictures & Television Engineers. Standard for Television - Data Structure for DV-Based Audio, Data and Compressed Video ó25 and 50 Mb/s. SMPTE 314-2005
[SMPTE_370]	Society of Motion Pictures & Television Engineers. Standard for Television - ata Structure for DV-Based Audio, Data and Compressed Video at 100 Mb/s 1080/60i, 1080/50i, 720/60p, 720/50p. SMPTE 370-2013.
[SMPTE_377]	Society of Motion Pictures & Television Engineers. Material Exchange Format (MXF) - File Format Specification (Revision of SMPTE 377-1-2019).
[SMPTE_421]	Society of Motion Pictures & Television Engineers. Standard for Television - VC-1 Compressed Video Bitstream Format and Decoding Process. SMPTE 0421-2013.
[SVG]	W3C: Scalable Vector Graphics (SVG) 1.1 specification (Second Edition). https://www.w3.org/TR/SVG11/
[TIFF]	Adobe Developers Association. TIFF Revision 6.0 Final — June 3, 1992. https://www.adobe.io/content/dam/udp/en/open/standards/tiff/TIFF6.pdf
[UNICODE]	The Unicode Consortium. The Unicode Standard, Version 13.0.0, 2020 https://www.unicode.org/versions/Unicode13.0.0/
[VIDEOMD]	Video Technical Metadata Extension Schema. https://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html
[WAV_IBM_MS]	Multimedia Programming Interface and Data Specifications 1.0. Issued as a joint design by IBM Corporation and Microsoft Corporation. August 1991. https://www.aelius.com/njh/wavemetatools/doc/riffmci.pdf
[WAV_LOC]	WAVE Audio File Format. Library of Congress. https://www.loc.gov/preservation/digital/formats/fdd/fdd000001.shtml

[XHTML_1.0]   XHTML 1.0 — The Extensible HyperText Markup Language (Second Edition). A Reformulation of HTML 4 in XML 1.0. W3C Recommendation, January 2000, revised August 2002. https://www.w3.org/TR/xhtml1/

[XHTML_1.1]   XHTML 1.1 — Module-based XHTML (Second Edition). W3C Recommendation, November 2010. https://www.w3.org/TR/xhtml11/

[XML_1.0]   Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation, November 2008. https://www.w3.org/TR/xml/

[XML_1.1]   Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation, September 2006. https://www.w3.org/TR/xml11/