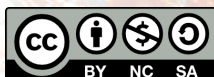


SÄILYTYS- JA SIIRTOKELPOISET TIEDOSTOMUODOT

v1.13.0

Tämä määrittely on osa opetus- ja kulttuuriministeriön
Avoimen tieteen ja digitaalisen kulttuuriperinnön kokonaisuutta



Lisenssi

Creative Commons Suomi CC-BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.fi>)

Tämän määrittelyn käyttäjälle annetaan oikeus jakaa eli kopioida, levittää, näyttää ja esittää teosta sekä valmistaa muutettuja teoksia seuraavilla ehdoilla:

- *Opetus- ja kulttuuriministeriö nimetään teoksen tekijäksi (ei kuitenkaan siten, että ilmoitus viittaisi lisenssinantajan tukevan lisenssinsaajaa tai teoksen käyttötapaa).*
- *Epäkaupallinen osapuoli eli käyttäjä ei saa käyttää teosta kaupallisesti.*
- *Mikäli käyttäjä tekee muutoksia tai käyttää teosta omien teostensa pohjana, tulee johdannaisteos jakaa samalla tai samankaltaisella lisenssillä.*

SISÄLTÖ

1	JOHDANTO	5
1.1	PAS-palvelut	5
1.2	Säilytys- ja siirtokelpoiset tiedostomuodot PAS-palveluissa	5
1.3	Tiedostomuotojen tekniset piirteet	6
1.4	Tiedostomuotokirjasto	7
1.5	Määrittelyn asema ja päivittäminen	8
2	SÄILYTYSKELOPOISET TIEDOSTOMUODOT	9
2.1	Teksti	9
2.1.1	Comma Separated Values (CSV)	9
2.1.2	Electronic Publications (EPUB)	9
2.1.3	Extensible Hypertext Markup Language (XHTML)	9
2.1.4	Extensible Markup Language (XML)	10
2.1.5	Hypertext Markup Language (HTML)	10
2.1.6	Industry Foundation Classes (IFC)	11
2.1.7	Open Document Format (ODF)	11
2.1.8	PDF for long-term preservation: PDF-Archive (PDF/A)	12
2.1.9	Tekstitiedosto (Plain text)	12
2.2	Ääni	13
2.2.1	Audio Interchange File Format (AIFF), LPCM-koodattu	13
2.2.2	Broadcast Wave Format (BWF)	13
2.2.3	Free Lossless Audio Codec (FLAC)	13
2.2.4	Linear Pulse-Code Modulation (LPCM)	14
2.2.5	MPEG-4 AAC – Advanced Audio Coding (AAC)	14
2.2.6	Waveform Audio Format (WAV)	15
2.3	Elävä kuva	15
2.3.1	Digital Picture Exchange (DPX)	15
2.3.2	FF Video Codec 1 (FFV1)	15
2.3.3	JPEG 2000 -sekvenssi	15
2.3.4	MPEG-4 AVC – Advanced Video Coding (AVC)	16
2.3.5	MPEG-H HEVC – High Efficiency Video Coding (HEVC)	16
2.4	Kuva	17
2.4.1	Digital Negative (DNG)	17
2.4.2	Joint Photographic Experts Group (JPEG)	17
2.4.3	Joint Photographic Experts Group JPEG 2000 (JP2)	17
2.4.4	Portable network graphics (PNG)	18
2.4.5	Scalable Vector Graphics (SVG)	18
2.4.6	Tagged Image File Format (TIFF)	18
2.4.7	WebP Image Format (WEBP)	19
2.5	Verkkoarkisto	19
2.5.1	Web ARChive Format (WARC)	19
2.6	Paikkatietoaineistot	19
2.6.1	Geographic Tagged Image File Format (GeoTiff)	19
2.6.2	Geography Markup Language (GML)	20
2.6.3	Geopackage	20
2.6.4	Keyhole Markup Language (KML)	20
2.7	Tietokannat	21
2.7.1	Software Independent Archiving of Relational Databases (SIARD)	21
2.8	Tilastoaineistot	21
2.8.1	SPSS Portable (POR)	21
2.9	Tieteelliset ohjelmistot	21
2.9.1	Matlab	21
2.9.2	Hierarchical Data Format (HDF)	22

3	SIIRTOKELPOISET TIEDOSTOMUODOT.....	23
3.1	Teksti.....	23
3.1.1	Microsoft Office Suite.....	23
3.1.2	Portable Document Format (PDF)	24
3.2	Ääni	24
3.2.1	Audio Interchange File Format (AIFF-C)	24
3.2.2	Moving Pictures Expert Group (MPEG) MPEG-1 layer-3, MPEG-2 layer-3 (MP3)	24
3.2.3	Windows Media Audio (WMA).....	25
3.3	Elävä kuva.....	25
3.3.1	Digital Video ja sen variantit (DV).....	25
3.3.2	Moving Pictures Expert Group (MPEG-1, MPEG-2)	26
3.3.3	Windows Media Video (WMV)	26
3.4	Kuva.....	26
3.4.1	Encapsulated postscript (EPS)	26
3.4.2	Graphics Interchange Format (GIF)	27
4	SÄILIÖMUODOT	28
5	TEKNISET METATIEDOT.....	31
5.1	ADDML	31
5.1.1	Skeeman soveltaminen	31
5.1.2	Esimerkki	31
5.2	AudioMD	32
5.2.1	Skeeman soveltaminen	32
5.2.2	Esimerkki	33
5.3	VideoMD	34
5.3.1	Skeeman soveltaminen	34
5.3.2	Esimerkki	36
5.4	MIX	37
5.4.1	Skeeman soveltaminen	37
5.4.2	Esimerkki	38
6	KONTROLLOITU SANASTO	40
6.1	Tekstitiedoston sallitut merkistöt	43
6.2	Tiedostomuodon tarkentaminen	43
LIITE A.	TIEDOSTOMUODOISTA.....	45
LIITE B.	TIEDOSTOMUOTOJEN PYSYVYYDEN ARVIOINTI.....	46
B.1.	Arviointikriteerit.....	46
B.2.	Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista.....	49
LIITE C.	MUIDEN KUIN SÄILYTYS- JA SIIRTOKELPOISTEN TIEDOSTOMUOTOJEN KÄSITTELY PAS-PALVELUSSA...	51
C.1.	Vaatimukset ja suositukset	51
C.2.	Metatietovaatimukset.....	52
LIITE D.	PAIKALLISET XML-SKEEMATIEDOSTOT	55
LÄHTEET		58

1 JOHDANTO

Tämä dokumentti määrittelee tiedostomuodot, ja niiltä vaadittavat tekniset metatiedot, joissa kansalliset pitkäaikaissäilytyspalvelut (PAS-palvelut) säilyttävät ja vastaanottavat aineistoja. Dokumentissa käsitellään ainoastaan tallennusvälineestä riippumattomia sisältöjä eli digitaalista sisältöä, jota hallinnoidaan tiedostoina.

1.1 PAS-palvelut

PAS-palveluilla tarkoitetaan kulttuuriperintöaineistojen ja tutkimusaineistojen pitkäaikaissäilyttämiseen tuotettuja palveluita yhdessä. Tässä dokumentissa hyödyntävällä organisaatiolla tarkoitetaan PAS-palvelua digitaalisten aineistojen pitkäaikaissäilyttämiseen käyttävää organisaatiota, laitosta tai muuta toimijaa.

Pitkäaikaissäilytys tarkoittaa digitaalisen informaation säilyttämistä ymmärrettävänä ja käytettävänä useiden kymmenien ja jopa satojen vuosien ajan. Laitteet, ohjelmistot ja tiedostomuodot vanhenevat ajan myötä, mutta informaation täytyy säilyä. Luotettava pitkäaikaissäilyttäminen edellyttää sisällön eheyden aktiivista valvontaa ja monenlaisiin riskeihin varautumista. Tässä ovat keskeisessä asemassa metatiedot, jotka kuvailevat mm. aineiston sisällön, historian ja alkuperän sekä tiedot siitä, miten informaatiota voidaan käyttää.

Kulttuuriperintöaineistojen pitkäaikaissäilyttämiseen tuotettu PAS-palvelu (Kulttuuriperintö-PAS-palvelu) takaa kirjastojen, arkistojen ja museoiden keskeisten kansallisten digitaalisten tietovarantojen pitkäaikaissäilyttämisen. Digitaalisilla kulttuuriperintöaineistoilla tarkoitetaan sekä digitoituja että digitaaliseen muotoon tuotettuja kulttuuriperintöaineistoja. Näitä ovat esimerkiksi lakisääteisen säilyttämisen piiriin kuuluvat digitaaliset kulttuuriaineistot, kansalliseen kulttuuriperintöön kuuluvat digitaaliset asiakirjalliset aineistot sekä aineellisen ja henkisen kulttuuriperinnön säilyttämisestä vastaavien, opetus- ja kulttuuriministeriön hallinnonalalla toimivien organisaatioiden muut pitkäaikaissäilytyksen piiriin kuuluvat digitaaliset tietovarannot.

Tutkimusaineistojen pitkäaikaissäilyttämiseen tuotettu PAS-palvelu (Fairdata PAS-palvelu) varmistaa tutkimuksen digitaalisten aineistojen saatavuuden ja pitkäaikaisen säilyvyyden. Tämä PAS-palvelu on osa pysyvää ja koordinoitua tutkimusaineistojen hallinnan toimintamallia. Pyrkimyksenä on, että tutkimuksen todennettavuus ja toistettavuus elinkaaren eri vaiheissa onnistuu ja tulosten hyödyntäminen on helppoa. Tällöin tutkimustuloksia voidaan käyttää yhä uudelleen, arvioida ja hyödyntää päätöksenteossa, ja samalla digitalisoitumisen myötä yhä nopeammin kasvavat tietomäärät on turvattu tulevien tutkijasukupolvien käyttöön.

1.2 Säilytys- ja siirtokelpoiset tiedostomuodot PAS-palveluissa

PAS-palveluiden tukemia tiedostomuotoja on kahdenlaisia: *säilytyskelpoisia* ja *siirtokelpoisia*. Säilytyskelpoisten tiedostomuotojen arvioidaan olevan käyttökelpoisia vielä pitkään, kun taas siirtokelpoiset tiedostomuodot ovat arvion mukaan niitä, joissa pitkäaikaissäilytettävää aineistoa on runsaasti tallennettu. Mikäli säilytettävän aineiston tiedostomuoto ei ole säilytys- eikä siirtokelpoinen, tiedostot tulee normalisoida suoraan säilytyskelpoiseen tiedostomuotoon, jos mahdollista. Tällaisista normalisoinneista vastaa ensisijaisesti hyödyntävä organisaatio, jolle keskitetty PAS-palvelu voi tarjota muunnoksiin liittyvää tukea (esim. tietoa soveltuvista työkaluista). Ennen normalisointien tekemistä on suositeltavaa olla yhteydessä PAS-palveluun, jotta tiedostot normalisoidaan mahdollisimman tarkoituksenmukaiseen tiedostomuotoon. On tärkeää, että tiedostoa ei muunneta ensin vain siirtokelpoiseen muotoon, koska tällöin jouduttaisiin tekemään vielä uusi, ylimääräinen migraatio säilytyskelpoiseen muotoon. Jokainen tiedostomuodon muunnos on riski datan säilyvyydelle, joten muunnoksien määrä on minimoitava. Uudet sekä alkujaan digitaaliset että digitoidut aineistot tulee tuottaa säilytyskelpoiseen tiedostomuotoon. Jos tiedostomuodosta on säilytyskelpoiseksi hyväksytty useampia versioita, tulee suosia uusinta hyväksyttyä versiota.

PAS-palvelun ja sitä käyttävän organisaation tulee sopia tiedostomuunnoksien (migraatioiden) yksityiskohdista siirtopaketin osana toimitettavassa säilytysuunnitelmassa; mitä tapahtuu siirtokelpoiselle

tiedostomuodolle silloin, jos se muunnetaan säilytyskelpoiseksi tiedostomuodoksi PAS-palvelussa. Samoin tulee pohtia, esimerkiksi mihin säilytyskelpoiseen tiedostomuotoon siirtokelpoinen tiedostomuoto olisi tarkoituksenmukaista muuntaa.

Tiedostomuotojen valinta on tärkeää pitkäaikaissäilytyksen onnistumiseksi. Tämän dokumentin liitteessä B esitellään kriteeristö, jolla tiedostomuotojen valinta on suoritettu. Tämän kriteeristön ja sen pohjalta tehtyjen arvioiden ylläpito on olennainen osa PAS-työtä. Valinnassa etusija on annettu vapaasti saatavilla oleville virallisille standardeille. Jos niitä ei ole ollut tarjolla, on hyväksytty tiedostomuotoja, jotka ovat ns. de facto- eli teollisuusstandardeja. Liitteessä B on käsitelty valittujen tiedostomuotojen pysyvyyden arviointia tarkemmin.

Tässä dokumentissa on määritelty säilytys- ja siirtokelpoisia tiedostomuotoja tekstin, äänen, elävän kuvan, kuvan, verkkoarkiston, paikkatietoaineistojen ja tilastoaineistojen osalta. On kuitenkin huomattava, että jotkin tiedostomuodot voivat sisältää erityyppistä informaatiota kuin tiedostomuodon luokittelu tässä dokumentissa antaa ymmärtää. Esimerkiksi PDF, joka on luokiteltu tekstitiedostomuodoksi, voi sisältää kuvan. Tiedoston todellinen sisältö tulee huomioida säilytys suunnitelmassa. Määrittelyssä on hyödynnetty Kanadan kansalliskirjaston ja -arkiston selvityksiä [LAC] ja muita kansainvälisiä selvityksiä.

Muut PAS-palveluiden määrittelyt määrittelevät muun muassa säilytettävien aineistojen metatiedot sekä siirto- ja jakelupaketin rakenteen [FI_META] ja rajapinnat miten aineistoja siirretään PAS-palveluun tai noudetaan sieltä käyttöön [FI_API].

Tässä määrittelyssä käytetään seuraavia termejä niin kuin ne on määritelty IETF:n (Internet Engineering Task Force) toimesta [RFC_2119].

- PITÄÄ (MUST)
- EI SAA (MUST NOT)
- PAKOLLINEN (REQUIRED)
- PITÄISI (SHOULD)
- EI PITÄISI (SHOULD NOT)
- SAA (MAY)
- VAPAAEHTOINEN (OPTIONAL)

1.3 Tiedostomuotojen tekniset piirteet

PAS-palvelu asettaa kaikille säilytettävälle tiedostoille yleisiä rajoituksia, jotka koskevat kaikkia tiedostomuotoja, vaikka kaikki rajoitukset eivät aina ole edes sovellettavissa¹. Erityisesti:

- 1) Tiedostoissa EI SAA käyttää salasanasuojauksia eikä mitään muita salaustekniikoita. Näiden käyttäminen hankaloittaa aineistojen pitkäaikaista säilyttämistä, ja joissakin tapauksissa jopa estää sen.
- 2) Tiedostoissa EI SAA käyttää DRM (Digital Rights Management) -tekniikoita. Nämä voivat aiheuttaa ongelmia jopa bittitason säilyttämiseksi.
- 3) Tiedostossa oleva digitaalinen allekirjoitus EI SAA estää tiedoston käsittelyä. Digitaalisen allekirjoituksen käyttäminen saattaa hankaloittaa aineistojen pitkäaikaista säilyttämistä, ja joissakin tapauksissa jopa estää sen.
- 4) Tiedostoja EI SAA pakata. Monet tiedostomuodot ovat pakattuja ja tällainen pakkaaminen on sallittua, mutta esimerkiksi erillinen pakkaaminen (zip jne.) on kiellettyä.
- 5) Tiedostosta EI SAA puuttua sen esittämiseen tarvittavia ulkoisia komponentteja. Tällaisia ovat esimerkiksi fontit ja muut linkitetty aineistot, jotka vaikuttavat tiedoston esittämiseen. Myös XML-skeemat tulkitaan ulkoisiksi komponenteiksi, jotka PITÄÄ lisätä mukaan siirtopakettiin liitteessä D kuvatulla tavalla.

¹ Näistä rajoituksista on mahdollista poiketa erityisistä syistä, kuitenkin niin että poikkeuksista sovitaan kirjallisesti PAS-palvelusopimuksessa. Poikkeukset käsitellään aina tapauskohtaisesti neuvottelemalla hyödyntävän organisaation kanssa ennen PAS-palvelusopimuksen allekirjoittamista.

- 6) Jos säilytettävä digitaalinen objekti sisältää upotettuja metatietoja (esim. XMP, Exif), niin näissä on käytettävä vain ASCII tai UTF-8 merkistöjä, ellei ko. metatiedon määrittely määrittele merkistöä tarkemmin.

Yllä olevien yleisten rajoitusten lisäksi siirtopaketin teknisissä metatiedoissa on kerrottava tarkemmat tiedot säilytettävistä digitaalisista objekteista.

Pakollisten teknisten piirteiden lisäksi on olemassa aineistokohtaisia teknisiä piirteitä, joita ei esiinny muilla aineistoilla, mutta joiden kuvaaminen helpottaa tämänkaltaisten tiedostojen tulkintaa tai on sen kannalta oleellista (esim. audiovisuaaliseen aineistoon liittyvät tekniset piirteet kuten näytteenottotaajuus). Näillä teknisillä piirteillä tarkoitetaan yleisesti sellaista teknistä tietoa, joka parantaa tiedon käytettävyyttä ja säilytettävyyttä.

Siirtopaketissa PITÄÄ olla vähintään seuraavat tekniset metatiedot jokaiselle tiedostolle:

- 1) Metatiedot, joita ei voida päätellä tiedostosta. Esimerkiksi tekstitiedoston merkistöä ei voi päätellä tekstitiedostosta, joten se on ilmoitettava aineiston teknisissä metatiedoissa.
- 2) Metatiedot, joita käytetään säilytys suunnitelmassa. Esimerkiksi, jos kuvasuhde halutaan säilyttää migraatiossa, on siirtopaketissa ilmoitettava riittävät tiedot alkuperäisen aineiston kuvasuhteesta.
- 3) Metatiedot, joiden halutaan säilyvän ja päivittyvän migraatiossa. Esimerkiksi, jos kuvatiedostossa on metatietoja (esim. EXIF), joiden säilymistä ei voida taata migraatiossa, on nämä metatiedot ilmoitettava aineiston teknisissä metatiedoissa.

Kunkin tiedostomuodon osalta on tiedostomuodon kuvauksen yhteydessä kerrottu millä metatietoskeemalla tekniset metatiedot on esitettävä siirtopaketissa. PAS-palvelu määrittelee tarkasti ne metatietoskeemojen elementit ja attribuutit, jotka ovat pakollisia PAS-palvelussa. Joissakin tapauksissa pakolliseen elementtiin tai attribuuttiin voidaan hyväksyä myös tuntematon arvo². Tuntemattomien arvojen käytöstä hyödyntävän organisaation on sovittava PAS-palvelun kanssa käyttöönottovaiheessa.

1.4 Tiedostomuotokirjasto

Tiedostomuotokirjasto (Format registry) on PAS-palvelun osa, joka sisältää tietoa tiedostomuodoista. Yleisimmin tunnettuja tiedostomuotokirjastoja ovat muun muassa PRONOM³ ja Yhdysvaltain kongressin kirjaston ylläpitämä digitaalisten tiedostomuotojen pysyvyyttä käsittelevä sivusto⁴. Tällaiset rekisterit ja kirjastot ovat merkittäviä tiedonlähteitä tiedostomuotojen kirjon hallitsemiseksi, vaikka sekä niiden kattavuudessa että ajantasaisuudessa on puutteita.

Osa kuvailutiedoista on yhteistä eri tiedostomuotokirjastoille, kuten tiedostomuodon nimi, versio, tunniste, kuvaus sekä edeltäjä ja seuraaja. Esimerkiksi nimi voi olla JPEG versio 1.00, tunniste image/jpeg ja seuraaja JPEG versio 1.01. Toisaalta on tietoja, jotka esiintyvät vain joissakin tiedostomuotokirjastoissa, mikä vaikeuttaa kuvailujen siirtoa kirjastosta toiseen. Tähän tarkoitukseen ei myöskään ole kehitetty järjestelmäriippumatonta vaihtomuotoa.

Tiedostomuotokirjaston sisältämän tiedon sekä hyödyntävän organisaation keräämien tietojen (mitä sovelluksia ne voivat käyttää ja miten pitkään) avulla PAS-palvelu pystyy tunnistamaan säilytettyjen digitaalisten objektien tiedostomuotojen vanhentumisen ja käynnistämään yhteistyössä aineiston omistavan organisaation kanssa säilytys suunnitelman uudelleenarvioinnin. Tämän tuloksena päätetään, onko tiedostot muunnettava ajantasaisempaan tiedostomuotoon, ja tarvittaessa valitaan uusi tiedostomuoto, johon vanhentumassa oleva digitaalinen objekti muunnetaan. Valitun tiedostomuodon yhteydessä viitataan sen

² <https://digitalpreservation.fi/support/vocabularies>

³ <https://www.nationalarchives.gov.uk/pronom/>

⁴ https://www.digitalpreservation.gov/formats/content/content_categories.shtml

tarkempaan kuvaukseen ja käytetään tiedostomuodolle PRONOMin pysyvää tunnistetta (PUID, PRONOM Persistent Unique Identifier) silloin kun se on mahdollista.

1.5 Määrittelyn asema ja päivittäminen

Tämä määrittely ohjaa niin hyödyntäviä organisaatioita kuin PAS-palveluitakin suunnittelemaan ja toteuttamaan omaa toimintaansa. Määrittely kuvaa tavoitetilaa. Se ei siten välttämättä ole kaikilta osin kuvaus PAS-palveluiden tämänhetkisestä käytännön toiminnasta, vaan PAS-palvelut toteuttavat tiedostomuototukea tämän määrittelyn perusteella tarvelähtöisesti. PAS-palveluiden verkkosivuilla on lisätietoja tiedostomuotojen käytännön tuesta ja toteutusperiaatteista⁵.

Määrittelyn ajanmukaisuus arvioidaan vuosittain, ja sitä kehitetään eteenpäin hyödyntävien organisaatioiden tarpeiden mukaisesti yhteistyössä, erityisesti uusien tiedostomuotojen levitessä käyttöön ja nykyisten tiedostomuotojen vanhentuessa.

Määrittelyn kolmitasoinen versionumerointi kuvaa, minkä tasoisia muutoksia määrittelyyn on tehty. Jos ensimmäinen numero päivittyy, se tarkoittaa merkittäviä muutoksia myös hyödyntävien organisaatioiden järjestelmiin, jotka on integroitu PAS-palveluun. Kesimmäisen numeron kasvu kuvaa tilanteita, joissa muutoksilla voi olla vaikutuksia organisaatioiden järjestelmiin. Kolmannen numeron muutos kuvaa tilanteita, joissa määrittelyyn on tehty vähäisempiä korjauksia tai täsmennyksiä. Tällaisia voivat olla vaikkapa määrittelyssä olevien esimerkkien täydentäminen.

Tämän määrittelyn uusimman version pysyvä tunniste on [urn:nbn:fi-fe2020100578095](https://nbn-resolving.org/urn:nbn:fi-fe2020100578095).

⁵ https://digitalpreservation.fi/support/format_support

2 SÄILYTYSKELPOISET TIEDOSTOMUODOT

Tässä luvussa määritellään pitkäaikaissäilytyspalveluissa säilytettävien digitaalisten objektien tiedostomuodot. Säilytyskelpoiseksi hyväksytään sellaiset tiedostomuodot, joissa tietosisällön säilyminen ja ymmärrettävyys voidaan taata pidemmällä aikavälillä.

Tiedostomuodot on lueteltu aineistotyypeittäin ja aakkosjärjestyksessä. Jos aineistotyyppille on määritelty sekä säiliö- että tiedostomuotoja, säiliöformaatit voivat sisältää vain tässä määrittelyssä lueteltuja säilytyskelpoisia tiedostomuotoja.

2.1 Teksti

2.1.1 Comma Separated Values (CSV)

CSV (comma separated values) on tiedostomuoto, jolla tallennetaan taulukkomuotoista tietoa tekstitiedostoon. CSV-tiedostoissa tietokokonaisuus on esitetty yhdellä rivillä, joka koostuu erotinmerkeillä erotetuista kentistä.

Hyväksyttävät versiot:

CSV; [RFC_4180]; PRONOM: x-fmt/18

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

text/csv

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

ADDML

2.1.2 Electronic Publications (EPUB)

EPUB on laitteisto- ja ohjelmistoriippumaton elektronisten julkaisujen tallennusmuoto, jonka aiemmat versiot (2.0.1 – 3.0.1) on standardoitu International Digital Publishing Forum (IDPF) toimesta ja uusimmat versiot (3.2 – 3.3) W3C:n toimesta. Standardissa julkaisun sisällön rakenne paketoitua säiliöön ja tekstin muotoilu sekä ulkoasu muodostetaan lukulaitteessa. Tiedostomuoto on ns. rakenteinen nippuformaatti. EPUB-tiedostomuodon määrittelystä on myös olemassa kaksi ISO-standardia [ISO_23736, ISO_30135].

Hyväksyttävät versiot:

EPUB versio 3; [EPUB_3]; PRONOM: fmt/483

EPUB versio 2.0.1; [EPUB_2.0.1]; PRONOM: fmt/483

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/epub+zip

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.3 Extensible Hypertext Markup Language (XHTML)

XHTML on HTML 4.0:sta kehitetty XML-sovitin, jonka viimeisin W3C:n suosittelema versio on XHTML 5.

Hyväksyttävät versiot:

XHTML versio 5; [HTML5]; PRONOM: fmt/471

XHTML versio 1.1; [XHTML_1.1]; PRONOM: fmt/103

XHTML versio 1.0; [XHTML_1.0]; PRONOM: fmt/102

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/xhtml+xml

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.4 Extensible Markup Language (XML)

XML on alustariippumaton SGML:stä johdettu merkintäkieli. W3C:n suosittelema XML kuvaa dokumentin loogisen rakenteen, mutta ei sen ulkoasua.

Jos säilytettävä XML-tiedosto käyttää skeemoja, hyödyntävän organisaation pitää sopia PAS-palvelun kanssa skeemojen saatavuudesta aineiston vastaanoton yhteydessä.

Hyväksyttävät versiot:

XML versio 1.1; [XML_1.1]; PRONOM: fmt/1776

XML versio 1.0; [XML_1.0]; PRONOM: fmt/101

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

text/xml

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.5 Hypertext Markup Language (HTML)

HTML on yksinkertainen SGML:stä johdettu hypertekstin merkintäkieli, joka on tullut standardikieleksi www-sivujen tuottamiseksi. HTML kuvaa dokumentin sisällön. PAS-palvelu suosittelee HTML-tiedostomuodon tuottamiseen joko versiota 4.01, joka on standardoitu ISO-standardina ISO/IEC 15445, tai versiota 5 [HTML5].

Hyväksyttävät versiot:

HTML versio 5; [HTML5]; PRONOM: fmt/471

HTML versio 4.01; [ISO_15445, HTML_4.01]; PRONOM: fmt/100

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

text/html

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.6 Industry Foundation Classes (IFC)

IFC (Industry Foundation Classes) on standardi, joka määrittelee, miten fyysisiä rakennusosia vastaavat tietokomponentit tulee kuvata. Tätä avoimen standardin mukaista tietomuotoa hyödynnetään laajasti ohjelmistojen välisessä tiedonvaihdoissa suunnittelusta rakentamiseen ja ylläpitoon. IFC on standardoitu ISO-standardina ISO 16739.

Hyväksyttävät versiot:

IFC versio 4.3.2.0; [ISO_16739:2024, IFC_4.3.2.0]; PRONOM: fmt/700

IFC versio 4.0.2.1; [ISO_16739:2018, IFC_4.0.2.1]; PRONOM: fmt/700

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

model/step

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.7 Open Document Format (ODF)

ODF on OASIS:n kehittämä avoin tiedostomuoto, jonka avulla voidaan esittää toimisto-ohjelmistoilla tuotettuja aineistoja kuten taulukkolaskenta-, kaavio-, esitys- ja tekstinkäsittelydokumentteja. ODF 1.0-tiedostomuoto on standardoitu ISO-standardina ISO/IEC 26300 ja versio 1.2 ISO-standardina ISO/IEC 26300. Teknisesti Open document on ZIP-pakattu arkisto eli ns. yksinkertainen nippuformaatti, joka sisältää useita tiedostoja ja hakemistoja. Dokumentin sisältö, esitystyyli, metatieto ja asetukset on eroteltu omiksi XML-tiedostoihin.

Yleisimmät ODF-standardia noudattavat tiedostotyyppit ovat mm.:

- .odt, tekstinkäsittely (ODT)
- .ods, taulukkolaskenta (ODS)
- .odp, esitys (ODP)
- .odg, piirto (ODG)
- .odf, kaava (ODF)

Hyväksyttävät versiot:

ODT versio 1.3; [ODF_1.3]; PRONOM: fmt/1756

ODT versio 1.2; [ISO_26300:2015, ODF_1.2]; PRONOM: fmt/291

ODT versio 1.1; [ODF_1.1]; PRONOM: fmt/290

ODT versio 1.0; [ISO_26300:2006, ODF_1.0]; PRONOM: fmt/136

ODS versio 1.3; [ODF_1.3]; PRONOM: fmt/1755

ODS versio 1.2; ISO_26300:2015, ODF_1.2]; PRONOM: fmt/295

ODS versio 1.1; [ODF_1.1]; PRONOM: fmt/294

ODS versio 1.0; [ISO_26300:2006, ODF_1.0]; PRONOM: fmt/137

ODP versio 1.3; [ODF_1.3]; PRONOM: fmt/1754

ODP versio 1.2; ISO_26300:2015, ODF_1.2]; PRONOM: fmt/293

ODP versio 1.1; [ODF_1.1]; PRONOM: fmt/292

ODP versio 1.0; [ISO_26300:2006, ODF_1.0]; PRONOM: fmt/138

ODG versio 1.3; [ODF_1.3]; PRONOM: fmt/1753

ODG versio 1.2; ISO_26300:2015, ODF_1.2]; PRONOM: fmt/297

ODG versio 1.1; [ODF_1.1]; PRONOM: fmt/296

ODG versio 1.0; [ISO_26300:2006, ODF_1.0]; PRONOM: fmt/139

ODF versio 1.3; [ODF_1.3]

ODF versio 1.2; [ISO_26300:2015, ODF_1.2]

ODF versio 1.0; [ISO_26300:2006, ODF_1.0]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

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

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.8 PDF for long-term preservation: PDF-Archive (PDF/A)

Tekstimuotoisen dokumentin esittämiseen tarkoitettu PDF/A on PDF-tiedostomuodon versio, joka soveltuu pitkäaikaiseen säilyttämiseen. PDF/A-tiedosto sisältää kaikki tarvittavat tiedot dokumentin visuaalisen ulkoasun säilyttämiseksi. PDF/A-tiedostosta on riisuttu ne ominaisuudet, joita ei voida luotettavasti säilyttää, kuten muun muassa skriptit, äänet, elävä kuva tai salaus. PAS-palvelu suosittelee versioita PDF/A-1a ja PDF/A-1b, jotka on standardoitu ISO-standardina ISO 19005-1, versioita PDF/A-2a, PDF/A-2b ja PDF/A-2u, jotka on standardoitu ISO-standardina ISO 19005-2, ja versioita PDF/A-3a, PDF/A-3b ja PDF/A-3u, jotka on standardoitu ISO-standardina ISO 19005-3. PDF/A-3-standardi on PDF/A-2-standardia vastaava, mutta PDF/A-3-standardin mukaiseen tiedostoon voidaan lisäksi upottaa liitteeksi mitä tahansa tiedostoja. PAS-palvelu takaa ymmärrettävyyden säilyttämisen varsinaiselle PDF-dokumentille (PDF/A-2-standardia vastaavat osuudet), mutta vain bittitason säilyttämisen tiedostoon upotetuille liitetiedostoille.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/pdf

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.1.9 Tekstitiedosto (Plain text)

Tekstitiedosto sisältää pelkkää tekstiä ilman muotoiluja tai muita ulkoasumäärittäjiä. Tekstitiedoston sisältämien merkkien valikoiman määrittelee merkkistö, josta on olemassa useita eri versioita esimerkiksi maakohtaisiin tarpeisiin.

Hyväksyttävät versiot:

Plain text; [RFC_2046]; PRONOM: [x-fmt/111](#)

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

text/plain

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

Jos tekstitiedosto sisältää taulukkomuotoista informaatiota, rakenteen kuvaaminen käyttäen ADDML [ADDML] metatietoskeemaa on suositeltavaa.

2.2 Ääni

2.2.1 Audio Interchange File Format (AIFF), LPCM-koodattu

AIFF-tiedostomuoto on Apple Inc:n kehittämä äänitiedostomuoto. Säilytyskelppoinen AIFF-tiedosto voi sisältää vain häviötöntä LPCM-koodattua ääntä. AIFF-äänitiedostoille ei tarvitse tehdä erottelua säiliön ja ääniraidan välillä, vaan ne käsitellään yhtenä PAS-palveluissa.

Hyväksyttävät versiot:

AIFF versio 1.3; [AIFF]; PRONOM: x-fmt/135

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/x-aiff

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

2.2.2 Broadcast Wave Format (BWF)

BWF on EBU:n (European Broadcasting Union) kehittämä äänitiedostojen vaihtoon sopiva tiedostomuoto. Se pohjautuu WAV-tiedostomuotoon. BWF on WAVin laajennus, mikä takaa yhteensopivuuden WAV-tiedostojen kanssa. Säilytyskelppoinen BWF-tiedosto voi sisältää vain LPCM koodauksen. BWF:n keskeinen ominaisuus on Broadcast extension -chunk (Bext), johon voidaan tallentaa metadataa. The International Association of Sound and Audiovisual Archives (IASA) suosittelee BWF-äänitiedostomuotoa arkistointiin. BWF-äänitiedostoille ei tarvitse tehdä erottelua säiliön ja ääniraidan välillä, vaan ne käsitellään yhtenä PAS-palveluissa.

Hyväksyttävät versiot:

BWF versio 2; [BWF]; PRONOM: fmt/527

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/x-wav

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

2.2.3 Free Lossless Audio Codec (FLAC)

FLAC on Xiph.org Foundationin kehittämä häviötön pakkausmenetelmä äänitiedostolle. FLAC:n pakkausmenetelmä on avoin ja lisenssitön. FLAC-muoto standardoidaan RFC-standardina RFC 9639 [RFC_9639].

Hyväksyttävät versiot:

FLAC; [RFC_9639]; PRONOM: fmt/279

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Matroska [RFC_9559]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/flac

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

2.2.4 Linear Pulse-Code Modulation (LPCM)

PCM (Pulse-Code Modulation) on analogisen äänisignaalin digitaalinen esitys, jossa signaalin suuruudesta otetaan näytteitä säännöllisesti tasaisin väliajoin. Linear Pulse-Code Modulation (LPCM) [RFC_3551, RFC_3190] on tietyn tyyppinen PCM, jossa kvantisointitasot ovat lineaarisesti yhdenmukaiset ja jota käytetään tyypillisesti säiliömuotojen ääniraidoissa.

Hyväksyttävät versiot:

LPCM; [RFC_3551, RFC_3190]

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Audio Video Interleave [AVI]
DV (raakamuoto) [IEC_61834, SMPTE_314, SMPTE_370]
Material Exchange Format [SMPTE_377]
Motion JPEG 2000 [ISO_15444-3]
Matroska [RFC_9559]
Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/L8, audio/L16, audio/L20, audio/L24

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

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

AAC-tiedostomuoto on Fraunhofer-Gesellschaftin kehittämä häviöllistä tiedonpakkausta käyttävä äänitiedostomuoto, joka suunniteltiin korvaamaan MP3-tiedostomuoto. MPEG-4 AAC -tiedostomuoto (MPEG-4 part 3) on standardoitu ISO-standardina (ISO/IEC 14496-3), osana MPEG-2- ja MPEG-4-määrittelyitä. MPEG-4 AAC on uudempi versio MPEG-2 AAC-standardista.

Hyväksyttävät versiot:

MPEG-4 part 3; [ISO_14496-3]

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/aac

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

2.2.6 Waveform Audio Format (WAV)

WAV-tiedostomuoto on IBM/Microsoftin kehittämä äänitiedostomuoto. WAV on säiliö, johon voi tallentaa ääntä erilaisilla koodausmenetelmillä. Säilytyskelpoiseen WAV-tiedostoon voi tallentaa vain häviötöntä LPCM-koodattua ääntä. WAV-äänitiedostoille ei tarvitse tehdä erottelua säiliön ja ääniraidan välillä, vaan ne käsitellään yhtenä PAS-palveluissa.

Hyväksyttävät versiot:

WAV; [WAV]; PRONOM: fmt/141

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/x-wav

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

2.3 Elävä kuva

2.3.1 Digital Picture Exchange (DPX)

Digital Picture Exchange (DPX) tiedostomuodossa elävä kuva on tallennettu peräkkäisinä DPX-kuvina ilman paketoitua. Tiedostomuotoa käytetään tyypillisesti elokuvien digitoinnissa. Tiedostomuoto on standardoitu SMPTE-standardina [SMPTE_268].

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/x-dpx

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.3.2 FF Video Codec 1 (FFV1)

FFV1 on avoimen lähdekoodin FFmpeg-projektin vuodesta 2003 alkaen kehittämä häviötön pakkausmenetelmä elävälle kuvalle. Vuonna 2015 muoto otettiin mukaan Internet Engineering Task Forcen (IETF) standardointiprosessiin osana Cellar-työryhmää. Uusin vakaa versio on vuonna 2013 julkaistu FFV1.3, jossa on vain pienehköjä muutoksia aiempiin FFV1.0- ja FFV1.1-versioihin verrattuna.

Hyväksyttävät versiot:

FFV1 versio 3; [RFC_9043]

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Matroska [RFC_9559]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/x-ffv

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

2.3.3 JPEG 2000 -sekvenssi

Tiedostomuodossa elävä kuva on tallennettu peräkkäisinä JPEG 2000 kuvina, jotka on paketoitu Material Exchange Format- (MXF), Motion JPEG 2000- (MJPEG2000) tai Quicktime MOVie (MOV) -säiliöön. JPEG 2000

on häviötön tai häviöllinen kuvatiedostomuoto. JPEG 2000 on standardoitu ISO-standardina ISO/IEC 15444-1. Motion JPEG 2000, jolla tarkoitetaan säiliömuotoa, on standardoitu ISO-standardina ISO/IEC 15444-3.

Hyväksyttävät versiot:

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

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Material Exchange Format [SMPTE_377]

Motion JPEG 2000 [ISO_15444-3]

Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/jpeg2000

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

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

MPEG-4 part 10 (AVC, H.264) on Moving Pictures Expert Groupin (MPEG) kehittämä häviöllinen pakkausmenetelmä elävän kuvan tiedostoille. MPEG-4:ää käytetään mm. digitaalisessa elävässä kuvassa sekä interaktiivisissa grafiikkasovelluksissa ja multimediassa. MPEG-4 Part 10 on standardoitu ISO-standardina ISO/IEC 14496-10.

Hyväksyttävät versiot:

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

MPEG-tiedostomuodot voivat sisältyä säiliömuotoihin:

Material Exchange Format [SMPTE_377]

MPEG-2 part 1 Transport Stream [ISO_13818-1]

MPEG-4 part 14 [ISO_14496-14]

Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/H264

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

2.3.5 MPEG-H HEVC – High Efficiency Video Coding (HEVC)

MPEG-H part 2 (HEVC, H.265) on Moving Pictures Expert Groupin (MPEG) kehittämä häviöllinen pakkausmenetelmä elävän kuvan tiedostoille. MEPEG-H part 2 on tarkoitettu MPEG-4 part 10 (AVC, H.264) seuraajaksi. MPEG-4 part 10 on standardoitu ISO-standardina ISO/IEC 23008-2.

Hyväksyttävät versiot:

MPEG-H part 2; [ISO_23008-2]

MPEG-tiedostomuodot voivat sisältyä säiliömuotoihin:

Matroska [RFC_9559]

MPEG-2 part 1 Transport Stream [ISO_13818-1]

MPEG-4 part 14 [ISO_14496-14]

Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/H265

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

2.4 Kuva

2.4.1 Digital Negative (DNG)

DNG on Adoben kehittämä valmistajariippumaton ja rojaltiton kuvatiedostomuoto, johon esimerkiksi digitaalikameran RAW-raakakuvatiedostot saa muunnettua häviöttömästi.

Hyväksyttävät versiot:

DNG versio 1.5.0.0; [DNG_15]; PRONOM: fmt/1841

DNG versio 1.4.0.0; [DNG_14]; PRONOM: fmt/730

DNG versio 1.3.0.0; [DNG_13]; PRONOM: fmt/438

DNG versio 1.2.0.0; [DNG_12]; PRONOM: fmt/437

DNG versio 1.1.0.0; [DNG_11]; PRONOM: fmt/152

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/x-adobe-dng

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.4.2 Joint Photographic Experts Group (JPEG)

Joint Photographic Experts Groupin kehittämä JPEG-standardi (ISO/IEC 10918-1) viittaa sekä tiedostomuotoon että pakkausmenetelmään. JPEG-pakkausmenetelmä on standardoitu häviöllinen kuvien pakkausmenetelmä täysväri- (RGB) ja mustavalkokuville.

Hyväksyttävät versiot:

JPEG part 1; versio 1.02; [ISO_10918-1]; PRONOM: fmt/44

JPEG part 1; versio 1.01; [ISO_10918-1]; PRONOM: fmt/43

JPEG part 1; versio 1.00; [ISO_10918-1]; PRONOM: fmt/42

JPEG/EXIF versio 2.3.2; PRONOM: fmt/1507

JPEG/EXIF versio 2.3.1; PRONOM: fmt/1507

JPEG/EXIF versio 2.3; PRONOM: fmt/1507

JPEG/EXIF versio 2.2.1; PRONOM: fmt/645

JPEG/EXIF versio 2.2; PRONOM: x-fmt/391

JPEG/EXIF versio 2.1; PRONOM: x-fmt/390

JPEG/EXIF versio 2.0; PRONOM: x-fmt/398

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/jpeg

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.4.3 Joint Photographic Experts Group JPEG 2000 (JP2)

Joint Photographic Experts Groupin kehittämä JPEG 2000 tarjoaa sekä häviöttömän että häviöllisen pakkausmenetelmän JPEGiä paremmalla kuvanlaadulla ja pienemmällä tiedostokoolla. ISO-standardoitu (ISO/IEC 15444-1) JPEG 2000 tukee harmaasävykuvia sekä RGB- ja CMYK-väritiloja.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/jp2

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.4.4 Portable network graphics (PNG)

PNG on W3C:n määrittelemä tiedostomuoto kuvatiedostolle. Häviötön PNG tukee 16-bittisiä harmaasävykuvia ja 48-bittisiä värikuvia. PNG on standardoitu ISO-standardina ISO/IEC 15948.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/png

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.4.5 Scalable Vector Graphics (SVG)

Scalable Vector Graphics (SVG) on XML-pohjainen tiedostomuoto kaksiulotteisten vektorikuvien esittämistä varten. SVG:n kehitys alkoi 1999 ja sitä kehittää sekä hallinnoi edelleen World Wide Web Consortium (W3C). Se on avoin standardi, jonka uusin vakaa versio on 1.1 [SVG].

Hyväksyttävät versiot⁶:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/svg+xml

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.4.6 Tagged Image File Format (TIFF)

TIFF on bittikarttakuva, jonka on kehittänyt Aldus (nykyisin osa Adobea). TIFF voi olla pakattu tai pakkaamaton mustavalko-, harmaasävy- tai värikuva (RGB tai CMYK). TIFF-säiliöön on suositeltavaa tallentaa vain pakkaamattomia kuvia.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/tiff

Pakollisten teknisten metatietojen metatietoskeema:

⁶ Tiedosto ei saa sisältää animaatioita eikä interaktiivista toiminnallisuutta.

MIX

2.4.7 WebP Image Format (WEBP)

WebP on Googlen kehittämä kuvatiedostomuoto, joka tukee häviötöntä ja häviöllistä pakkausta. Sen käyttötarkoitukset ovat vastaavat kuin JPEG, PNG ja GIF-muodoilla. WebP on standardoitu RFC-standardina RFC 9649.

Hyväksyttävät versiot:

WebP; [RFC_9649]; PRONOM⁷: [fmt/566](#), [fmt/567](#), [fmt/568](#)

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/webp

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.5 Verkkoarkisto

2.5.1 Web ARChive Format (WARC)

WARC-tiedostomuoto (Web ARChive) määrittelee menetelmän useiden digitaalisten resurssien yhdistämiseksi yhdeksi arkistointitiedostoksi. ISO-standardoitu (ISO 28500) WARC on The Internet Archiven kehittämän ARC-tiedostomuodon kehittyneempi versio.

Hyväksyttävät versiot:

WARC versio 1.1; [ISO_28500:2017, WARC_1.1]; PRONOM: [fmt/1281](#)

WARC versio 1.0; [ISO_28500:2009, WARC_1.0]; PRONOM: [fmt/1355](#)

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/warc

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.6 Paikkatietoaineistot

2.6.1 Geographic Tagged Image File Format (GeoTiff)

GeoTIFF on TIFF 6.0 -tiedostomuodon laajennus, joka määrittelee paikkatietoaiheisten metatietokenttien tallennuksen TIFF-kuvien yhteyteen [GEOTIFF].

Hyväksyttävät versiot:

GeoTIFF versio 1.0 [GEOTIFF]; PRONOM: [fmt/155](#)

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

image/tiff

Pakollisten teknisten metatietojen metatietoskeema:

MIX

⁷ Eri PRONOM-tunnisteet viittaavat tiedostomuodon eri profiileihin (lossy, lossless, extended).

2.6.2 Geography Markup Language (GML)

Geography Markup Language (GML) on XML-pohjainen merkitäkieli erilaisten paikkatietoa sisältävien piirteiden esittämiseen. GML on standardoitu ISO-standardina ISO 19136 [ISO_19136].

Hyväksyttävät versiot:

GML versio 3.2.2 [ISO_19136] PRONOM: [fmt/1047](#)

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/gml+xml

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot voi esittää sopivalla metatietoskeemalla.

2.6.3 Geopackage

GeoPackage on avoin, alustasta riippumaton ja standardeihin perustuva tiedostomuoto paikkatiedoille, jossa paikkatiedot tallennetaan SQLite-tietokantaan. Geopackage on standardoitu Open Geospatial Consortiumin (OGC) toimesta [OGC_1.3.1, OGC_1.3.0].

Hyväksyttävät versiot:

Geopackage versio 1.3.1 [OGC_1.3.1]; PRONOM: [fmt/1700](#)

Geopackage versio 1.3.0 [OGC_1.3.0]; PRONOM: [fmt/1700](#)

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/geopackage+sqlite3

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.6.4 Keyhole Markup Language (KML)

Keyhole Markup Language (KML) on Googlen kehittämä XML-pohjainen merkitäkieli, joka on tarkoitettu erityisesti kaksi- ja kolmiulotteisten karttojen annotointiin ja visualisointiin. Tiedostomuoto on myös OGC:n hyväksymä standardi [KML].

Hyväksyttävät versiot:

KML versio 2.3 [KML]; PRONOM: [fmt/244](#)

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/vnd.google-earth.kml+xml

Sallitut merkistöt:

ks. luku 6.1

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot voi esittää sopivalla metatietoskeemalla.

2.7 Tietokannat

2.7.1 Software Independent Archiving of Relational Databases (SIARD)

SIARD on tiedostomuoto, joka on suunniteltu relaatiotietokantojen pitkäaikaissäilyttämiseen toimittajariippumattomassa muodossa. Tiedostomuoto perustuu XML:ään ja SQL-kieleen, ja sen on kehittänyt Sveitsin liittovaltion arkisto (Swiss Federal Archives), ja nykyisin sen ylläpito on organisoitu DILCIS-boardin⁸ alle.

Hyväksyttävät versiot:

SIARD versio 2.2 [SIARD_2.2]; PRONOM: fmt/1777

SIARD versio 2.1.1 [SIARD_2.1.1]

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/x-siard

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.8 Tilastoaineistot

2.8.1 SPSS Portable (POR)

SPSS (Statistical Package for the Social Sciences, IBM SPSS Statistics) Portable file format on tiedostomuoto, joka on luotu mahdollistamaan tilastollisten matriisimuotoisten tiedostojen siirrot eri tietokoneiden, käyttöjärjestelmien ja ohjelmaversioiden välillä. Tiedostot on mahdollista avata SPSS-ohjelman eri versioilla ja useimmilla muilla tilastomatemaattisilla ohjelmilla. POR-tekstitiedostossa SPSS-ohjelmasta viety data tallennetaan ASCII-muodossa.

Hyväksyttävät versiot:

SPSS Portable; [SPSS]; PRONOM: fmt/997

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/x-spss-por

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.9 Tieteelliset ohjelmistot

2.9.1 Matlab

Matlab (matrix laboratory) on the MathWorks -yhtiön ylläpitämä numeeriseen laskentaan tarkoitettu tietokoneohjelmisto sekä siinä käytettävä ohjelmointikieli. Matlab-aineistoihin voi liittyä hyvin monenlaisia tiedostoja, mutta vain Matlabin data- (.mat) ja kooditiedostot (.m) hyväksytään säilytyskelpoisiksi. Kooditiedostot ovat rakenteisia tekstitiedostoja ja niitä käsitellään PAS-palvelussa kuten tekstitiedostoja.

Hyväksyttävät versiot:

Matlab versio 7.3; PRONOM: fmt/828

Matlab Level 5 versio 7; PRONOM: fmt/806

⁸ <https://dilcis.eu/>

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/matlab

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

2.9.2 Hierarchical Data Format (HDF)

Hierarchical Data Format (HDF) on yleiskäyttöinen, erityisesti tieteellisen datan tallentamiseen suunniteltu tiedostomuotojen joukko [HDF5]. Tiedostomuodon kehitystä kontrolloi voittoa tuottamaton yritys HDF Group.

Hyväksyttävät versiot:

HDF5 versio 1.10; [HDF5]; PRONOM⁹: [fmt/807](#); [fmt/286](#); [fmt/287](#)

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

application/x-hdf5

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

⁹ Eri PRONOM tunnisteet viittaavat tiedostomuodon Superblocking eri versioihin. Versiolle 3 ei vielä ole PRONOM tunnistetta.

3 SIIRTOKELPOISET TIEDOSTOMUODOT

Tässä luvussa määritellään säilytyskelpoisten tiedostomuotojen lisäksi PAS-palveluun siirrettävissä tietopaketeissa hyväksyttävät tiedostomuodot. Siirtokelpoiksi hyväksytään vain sellaisia tiedostomuotoja, joita käytetään useassa PAS-palvelua hyödyntävässä organisaatiossa ja joissa pitkäaikaissäilytettävää aineistoa on runsaasti tallennettu. PAS-palvelu voi muuntaa säilytysuunnitelman asettamien vaatimusten ja ehtojen mukaisesti siirtokelpoisessa muodossa vastaanotetut tiedostot säilytyskelpoiseen tiedostomuotoon ennen säilyttämisen aloittamista. Jos hyödyntävä organisaatio normalisoi aineistonsa tiedostomuotoja, tulee normalisointi aina suorittaa säilytyskelpisiin tiedostomuotoihin. PAS-palveluun sisältyy myös migraatiotarpeen säännöllinen arviointi muun muassa tiedostomuotokirjastoon tallennettujen tietojen nojalla.

Tiedostomuodot on lueteltu aineistotyypeittäin ja aakkosjärjestyksessä. Jos aineistotyyppille on määritelty sekä säiliö- että tiedostomuotoja, säiliöformaattit voivat sisältää vain tässä määrittelyssä lueteltuja säilytys- tai siirtokelpoisia tiedostomuotoja.

3.1 Teksti

3.1.1 Microsoft Office Suite

Microsoft Office Suite on laajasti käytössä oleva kaupallinen ohjelmisto, jonka avulla voidaan esittää toimisto-ohjelmistoilla tuotettuja aineistoja kuten taulukkolaskenta-, kaavio-, esitys- ja tekstinkäsittelydokumentteja. Microsoft Office Suite -ohjelmiston uudemmat versiot tukevat Office Open XML -muotoja (Transitional ja Strict), jotka on ISO-standardoitu (ISO 29500-1). Microsoft Office Suite -ohjelmistosta (versiosta 97 ylöspäin) siirtokelpoisia tiedostomuotoja ovat:

- Word Document, tekstinkäsittely
- Excel Spreadsheet, taulukkolaskenta
- Powerpoint, esitys

Hyväksyttävät versiot:

Microsoft Word versio 2007 onwards; [ISO_29500-1, ECMA_376]; PRONOM: fmt/412
 Microsoft Word versio 97-2003; [MS_DOC]; PRONOM: fmt/40
 Microsoft Excel versio 2007 onwards; [ISO_29500-1, ECMA_376]; PRONOM: fmt/214
 Microsoft Excel versio 8X; [MS_XLS]; PRONOM: fmt/62
 Microsoft Excel versio 8; [MS_XLS]; PRONOM: fmt/61
 Microsoft Powerpoint versio 2007 onwards; [ISO_29500-1, ECMA_376]; PRONOM: fmt/215
 Microsoft Powerpoint versio 97-2003; [MS_PPT]; PRONOM: fmt/126

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

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

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

3.1.2 Portable Document Format (PDF)

PDF on Adoben kehittämä tekstimuotoisen dokumentin esittämiseen tarkoitettu alustariippumaton ja avoin tiedostomuoto. Alkuperäisen dokumentin taitto, kirjasintyyli, grafiikka ja värit siirtyvät muuttumattomana PDF-tiedoston mukana. PDF-tiedostomuoto on standardoitu ISO-standardina ISO 32000-2. Standardi kattaa kaikki PDF-tiedostomuotoversiot 1.0 – 2.0.

Hyväksyttävät versiot:

PDF versio 1.7; [ISO_32000-2]; PRONOM: [fmt/276](#)

PDF versio 1.6; [ISO_32000-2]; PRONOM: [fmt/20](#)

PDF versio 1.5; [ISO_32000-2]; PRONOM: [fmt/19](#)

PDF versio 1.4; [ISO_32000-2]; PRONOM: [fmt/18](#)

PDF versio 1.3; [ISO_32000-2]; PRONOM: [fmt/17](#)

PDF versio 1.2; [ISO_32000-2]; PRONOM: [fmt/16](#)

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/pdf

Pakollisten teknisten metatietojen metatietoskeema:

PAS-palvelu ei vaadi tiedostoilta teknistä metatietoa, mutta tarvittaessa tekniset metatiedot saa esittää sopivalla metatietoskeemalla.

3.2 Ääni

3.2.1 Audio Interchange File Format (AIFF-C)

AIFF-C on pakattu versio AIFF-tiedostosta (ks. luku 2.2.1). Pakkaamaton LPCM-koodattu ääni AIFF-säiliössä on säilytyskelppoinen muoto (ks. luku 2.2.1), mutta PAS-palvelu hyväksyy pakatun AIFF-tiedoston siirtokelpoisena tiedostomuotona. AIFF-C-äänitiedostoille ei tarvitse tehdä erottelua säiliön ja ääniraidan välillä, vaan ne käsitellään yhtenä PAS-palveluissa.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/x-aiff

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

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

MP3 on Moving Pictures Expert Groupin (MPEG) kehittämä häviöllinen äänitiedostomuoto, joka on laajasti käytössä ja tuettu useimmissa nykyisissä mediasoittimissa. MP3:n pakkausmenetelmä hävittää äänestä ihmisen kuuloalueen ulkopuoliset taajuudet ja pakkaa tiedoston. MPEG-1 layer-3 on standardoitu ISO-standardina ISO/IEC 11172-3. MPEG-2 layer-3 on standardoitu ISO-standardina ISO/IEC 13818-3.

Hyväksyttävät versiot:

MPEG-2 layer-3; [ISO_13818-3]; PRONOM: [fmt/134](#)

MPEG-1 layer-3; [ISO_11172-3]; PRONOM: [fmt/134](#)

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

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]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/mpeg

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

3.2.3 Windows Media Audio (WMA)

WMA on Microsoftin kehittämä häviöllinen pakkausmenetelmä äänitiedostolle. WMA on samankaltainen tiedostomuoto kuin MP3. Tiedostomuodolla on alkuperäisen WMA-pakkausmenetelmän lisäksi pakkausmenetelmät WMA Pro, WMA Voice ja häviötön WMA Lossless. WMA on Microsoftin yksityisomistuksellinen tiedostomuoto, jonka määrittelyä ei ole julkaistu.

Hyväksyttävät versiot:

WMA versio 9; PRONOM: fmt/132

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Advanced Systems Format [ASF]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/x-ms-wma

Pakollisten teknisten metatietojen metatietoskeema:

AudioMD

3.3 Elävä kuva

3.3.1 Digital Video ja sen variantit (DV)

DV on usean laitevalmistajan yhteistyön tuloksena syntynyt pakkausmenetelmä elävän kuvan raakataltointiin. DV on standardoitu standardeina IEC 61834, SMPTE 314 ja SMPTE 370, jotka tunnetaan nimillä DV, DVCAM, DVCPRO-50 ja DVCPRO-HD.

Hyväksyttävät versiot:

DV (ja sen variantit); [IEC_61834, SMPTE_314, SMPTE_370]; PRONOM: x-fmt/152

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Audio Video Interleave [AVI]
DV (raakamuoto) [IEC_61834, SMPTE_314, SMPTE_370]
Material Exchange Format [SMPTE_377]
Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/dv

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

3.3.2 Moving Pictures Expert Group (MPEG-1, MPEG-2)

MPEG-1 part 2 (MPG, MPEG) on Moving Pictures Expert Groupin (MPEG) kehittämä häviöllistä pakkausmenetelmää käyttävä tiedostomuoto elävälle kувalle. MPEG-1 part 2 on standardoitu ISO-standardina ISO/IEC 11172-2.

MPEG-2 part 2 (H.262) on Moving Pictures Expert Groupin (MPEG) kehittämä häviöllinen pakkausmenetelmä äänelle ja kувalle. ISO-standardoitu (ISO/IEC 13818-2) MPEG-2 tukee vaihtelevaa bittinopeutta, joka tallentaa kuvan ja äänen muutokset ns. ruutujen (eng. frame) välillä. MPEG-standardeista MPEG-2 part 2 on laajimmin käytössä oleva HDTV-tasoinen tiedostomuoto.

Hyväksyttävät versiot:

MPEG-2 part 2; versio 2; [ISO_13818-2]; PRONOM: fmt/640

MPEG-1 part 2; versio 1; [ISO_11172-2]; PRONOM: fmt/649

MPEG-tiedostomuodot voivat sisältyä säiliömuotoihin:

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]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/mpeg

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

3.3.3 Windows Media Video (WMV)

Microsoftin kehittämä tiedostomuoto WMV on pakkausmenetelmä elävälle kувalle. WMV 9 on standardoitu Society of Motion Picture and Television Engineers (SMPTE)-standardina SMPTE 421M, joka tunnetaan paremmin tunnisteenä VC-1.

Hyväksyttävät versiot:

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

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Advanced Systems Format [ASF]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

video/x-ms-wmv

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

3.4 Kuva

3.4.1 Encapsulated postscript (EPS)

EPS on Adobe'n kehittämä PostScript tiedostomuoto, joka voi sisältää sekä vektori- että bittikarttakuvia. Tiedostomuotoa käytetään PostScript-kielisten kuvien siirtoon sovellusten välillä.

Hyväksyttävät versiot:

EPS versio 3.0; [EPS]; PRONOM: fmt/124

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/postscript

3.4.2 Graphics Interchange Format (GIF)

GIF on CompuServen kehittämä häviötön pakkausmenetelmä kuvatiedostoille. GIF-tiedostot voivat sisältää 8-bittisiä harmaasävy- tai värikuvia. Yhtäaikaisten värien määrä rajoittuu 256:een.

Hyväksyttävät versiot:

GIF versio 1989a; [GIF_89a]; PRONOM: fmt/4

GIF versio 1987a; [GIF_87a]; PRONOM: fmt/3

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/gif

Pakollisten teknisten metatietojen metatietoskeema:

MIX

4 SÄILIÖMUODOT

Säiliömuotoiset digitaaliset objektit sisältävät osia, jotka ovat digitaalisia objekteja, mutta jotka eivät ole erillisiä tiedostoja. Esimerkiksi videotiedosto (säiliömuoto) voi sisältää yhden tai useamman kuva- tai ääniraidan. Tällöin voidaan tarvita sekä digitaalisen objektin metatietoja että sen sisältämien osien metatietoja. Esimerkiksi jos elokuvatiedosto digitaalisena objektina sisältää kuva- ja ääniraitoja, tarvitaan eheystieto digitaalisesta objektista, mutta kunkin raidan ominaisuudet digitaalisen objektin osista.

Digitaalisten objektien tekniset metatiedot PITÄÄ jakaa osiin ja kuvata tässä luvussa esitetyllä tavalla, kun digitaalisen objektin tiedostomuotona on luvussa 6 lueteltu säiliömuoto. Taulukossa 2 ilmenee säiliömuotojen pakollinen MIME-tyyppi. Versionumeroa ei käytetä säiliömuodoille, paitsi Matroska-säiliöstä hyväksytään vain versio 4. Tällä hetkellä jokaisen elävän kuvan tiedostomuoto vaatii sekä säiliömuodon että kuvaraidan (sekä mahdolliset ääniraidat) tiedot omina teknisen metatietojen lohkoina. Muille tiedostomuodoille osiin jakaminen on vapaaehtoista. Kokonainen XML-muotoinen esimerkki tämän luvun esityksestä on saatavilla osoitteesta <https://digitalpreservation.fi/support/examples>.

Jokaiselle digitaalisen objektin osalle esitetään oma PREMIS-muotoinen metatieto-objekti omassa METS:n techMD-lohkossaan. Sen tyyppin PITÄÄ olla "bitstream", ja sen PITÄÄ sisältää digitaalisen objektin osan tunniste, muoto ja sen versio alla olevan esimerkin mukaisesti.

```
<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/H264</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>
```

Digitaalisen objektin osan muut hallinnolliset ja kuvailevat metatiedot esitetään määrittelyiden mukaisesti omissa lohkoissaan ja omilla metatietomuodoillaan. Esimerkiksi äänen ja elävän kuvan osalta tekniset piirteet PITÄÄ esittää AudioMD- ja VideoMD-metatietomuodoilla lukujen 5.2 ja 5.3 mukaisesti.

Säiliömuotoiset digitaaliset objektit kuvataan PREMIS:llä tavallisesti, mutta lisättynä relaatiolla edellä kuvattuihin "bitstream"-tyyppisiin PREMIS-metatietoihin. Alla on esimerkki relaatiosta kahteen digitaalisen objektin osaan:

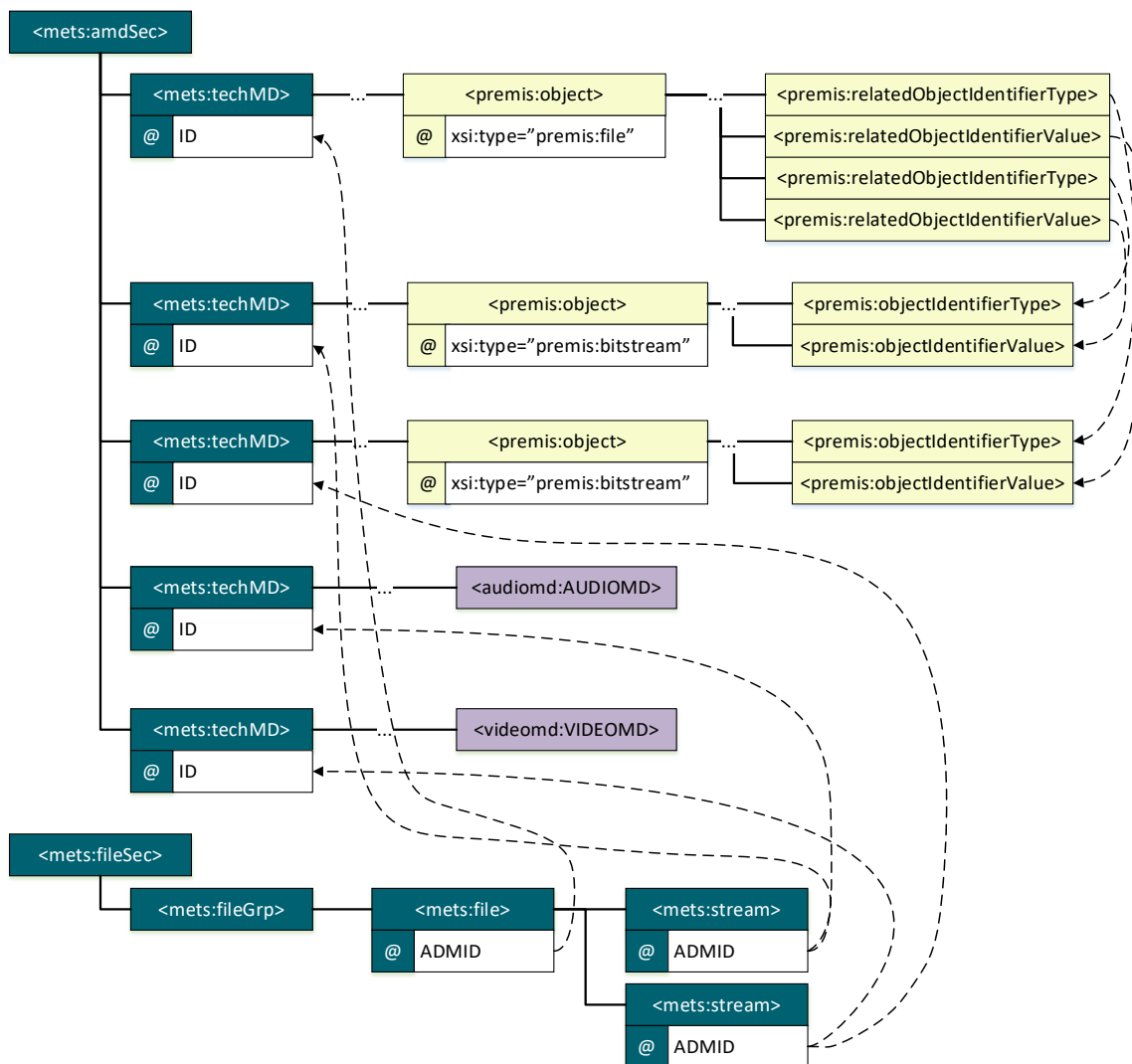
```
<premis:relationship>
  <premis:relationshipType>structural</premis:relationshipType>
  <premis:relationshipSubType>includes</premis:relationshipSubType>
  <premis:relatedObjectIdentification>
    <premis:relatedObjectIdentifierType>UUID
    </premis:relatedObjectIdentifierType>
    <premis:relatedObjectIdentifierValue>
      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>
```

METS:n tiedosto-metatietoon lisätään digitaalisen objektin osat <stream>-elementeillä. Tämän elementin ADMID-attribuutin PITÄÄ sisältää linkitykset kyseisen osan hallinnollisiin metatietoihin. Alla olevassa esimerkissä on kaksi digitaalisen objektin osaa, joissa molemmissa on viittaukset kahteen eri hallinnolliseen metatietoon.

```
<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>
```

Kuvailevien metatietojen linkitys digitaalisiin objekteihin ja niiden osiin tehdään kokonaan METS-rakennekartassa DMDID- ja FILEID-attribuuttien avulla. Rakennekartasta saa tällöin viitata digitaalisen objektin osaan (FILEID-attribuutti saa viitata <stream>-elementin ID-attribuuttiin).

Edellä esitetty tapa kytkee digitaalisen objektin, sen osat ja hallinnolliset metatiedot toisiinsa kuvan 1 mukaisesti.



Kuva 1: Digitaalisen objektin, sen osien ja hallinnollisten metatietojen kytkeminen toisiinsa

5 TEKNISET METATIEDOT

5.1 ADDML

CSV-tiedostojen pakolliset tekniset metatiedot esitetään siirto- ja jakelupaketeissa käyttäen ADDML-metatietoskeemaa [ADDML]. Jos tekstitiedosto sisältää rakenteista informaatiota, rakenteen kuvaaminen käyttäen ADDML-metatietoskeemaa on suositeltavaa.

5.1.1 Skeeman soveltaminen

Siirto- ja jakelupaketeissa ADDML-metatietomuotoa käytetään ADDML-skeeman mukaisesti. CSV-tiedostojen osalta metatiedoissa on esitettävä vähintään tiedostossa käytetyt kentän ja tietueen erotinmerkit (fieldSeparatingChar ja recordSeparator). Muita skeeman rakenteita voi siirto- ja jakelupaketeissa käyttää skeeman mukaisesti.

5.1.2 Esimerkki

```
<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

Äänitiedostojen pakolliset tekniset metatiedot esitetään siirto- ja jakelupaketeissa käyttäen audioMD-metatietoskeemaa [AUDIOMD].

5.2.1 Skeeman soveltaminen

Seuraavassa on esitetty audioMD-skeeman elementtien ja attribuuttien käyttö PAS-palvelussa. Taulukossa on ilmoitettu elementin tai attribuutin esiintymät ja velvoitteet seuraavin lyhentein:

T	Toistettavissa
ET	Ei toistettavissa
P	Pakollinen
V	Vapaaehtoinen (saa olla siirtopaketissa ja tallennetaan PAS-palveluun)

Taulukossa 1 on lueteltu vain PAS-palvelun vaatimat muutokset audioMD-skeemaan. Muita audioMD-skeeman elementtejä ja niihin liittyviä attribuutteja voi PAS-palvelussa käyttää audioMD-skeeman mukaisesti. Jos tuntemattomia arvoja käytetään elementeissä tai attribuuteissa, niin <note>-elementtiin tulee merkitä miksi ja mitä se merkitsee. Esimerkiksi, tarkoittaako 0 (nolla), että tietoa ei ole saatavilla, elementti ei sovellu käytettäväksi, jne.

Taulukko 1: AudioMD-metatietoskeeman muutokset

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
<audioMD>	ET, P	Juurielementin pitää sisältää <fileData>- ja <audioInfo>-elementit. audioMD-skeemassa nämä ovat vapaaehtoisia.
-<fileData>	ET, P	Pakollinen elementti PAS-palvelussa
--<audioDataEncoding>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav), jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<bitsPerSample>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<compression>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa
---<codecCreatorApp>	ET, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav), jos tietoa ei voida kohtuullisella vaivalla selvittää, tai (:unap) pakkaamattomalle äänelle.

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
---<codecCreatorAppVersion>	ET, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman äänen osalta tai niille ohjelmistoille, joilla ei ole versiointia.
---<codecName>	ET, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman äänen osalta.
---<codecQuality>	ET, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa
--<dataRate>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<dataRateMode>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa.
--<samplingFrequency>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
-<audioInfo>	ET, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa
--<duration>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav), jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<numChannels>	T, P	Vapaaehtoinen audioMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav), jos tietoa ei voida kohtuullisella vaivalla selvittää.

5.2.2 Esimerkki

```

<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

Elävien kuvien pakolliset tekniset metatiedot esitetään siirto- ja jakelupaketeissa käyttäen videoMD-metatietoskeemaa [VIDEOMD].

5.3.1 Skeeman soveltaminen

Seuraavassa on esitetty videoMD-skeeman elementtien ja attribuuttien käyttö PAS-palvelussa. Taulukossa on ilmoitettu elementin tai attribuutin esiintymät ja velvoitteet seuraavin lyhentein:

T	Toistettavissa
ET	Ei toistettavissa
P	Pakollinen
V	Vapaaehtoinen (saa olla siirtopaketissa ja tallennetaan PAS-palveluun)

Taulukossa 2 on lueteltu vain PAS-palvelun vaatimat muutokset videoMD-skeemaan. Muita videoMD-skeeman elementtejä ja niihin liittyviä attribuutteja voi PAS-palvelussa käyttää videoMD-skeeman mukaisesti. Jos tuntemattomia arvoja käytetään elementeissä tai attribuuteissa, niin <note>-elementtiin tulee merkitä miksi ja mitä se merkitsee. Esimerkiksi, tarkoittaako 0 (nolla), että tietoa ei ole saatavilla, elementti ei sovellu käytettäväksi, jne.

Taulukko 2: VideoMD-metatietoskeeman muutokset

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
<videoMD>	ET, P	Juurielementin pitää sisältää <fileData>- ja <videoInfo>-elementit. VideoMD-skeemassa nämä ovat vapaaehtoisia.
-<fileData>	ET, P	
--<duration>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav), jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<dataRate>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<bitsPerSample>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
--<color>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa
--<compression>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa
---<codecCreatorApp>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman kuvan osalta.
---<codecCreatorAppVersion>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman kuvan osalta tai niille ohjelmistoille, joilla ei ole versiointia.
---<codecName>	EP, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman kuvan osalta.
---<codecQuality>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa
--<dataRateMode>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa
--<frameRate>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<frame>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa
---<pixelsHorizontal>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
---<pixelsVertical>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.
---<PAR>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia 0, jos tietoa ei voida kohtuullisella vaivalla selvittää.

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
---<DAR>	ET, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:etal), jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<sampling>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää.
--<signalFormat>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa. Tuntemattomana arvona voidaan sallia (:unav) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää, mutta (:unap) vain jos kuva ei ole digitaalinen tv/video.
--<sound>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa

5.3.2 Esimerkki

```
<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

Kuvatiedostojen pakolliset tekniset metatiedot esitetään siirto- ja jakelupaketeissa käyttäen MIX-metatietoskeemaa [MIX].

5.4.1 Skeeman soveltaminen

Seuraavassa on esitetty MIX-skeeman elementtien ja attribuuttien käyttö PAS-palvelussa. Taulukossa on ilmoitettu elementin tai attribuutin esiintymät ja velvoitteet seuraavin lyhentein:

T	Toistettavissa
ET	Ei toistettavissa
P	Pakollinen
V	Vapaaehtoinen (saa olla siirtopaketissa ja tallennetaan PAS-palveluun)
ES	Ei suositeltava (saa olla siirtopaketissa, mutta PAS-palvelu ei hyödynnä eikä tallenna). Jakelupaketissa ei ole käytössä.

Taulukossa 3 on lueteltu vain PAS-palvelun vaatimat muutokset MIX-skeemaan. MIX:n osalta on jätetty elementtejä pois BasicDigitalObjectInformation-elementistä, koska nämä ovat päällekkäisiä PREMIS:in kanssa. Muita MIX-skeeman elementtejä ja niihin liittyviä attribuutteja käytetään PAS-palvelussa MIX-skeeman mukaisesti¹⁰.

Taulukko 3: MIX-metatietoskeeman muutokset

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
<BasicDigitalObjectInformation>	P, ET	Elementti on MIX-skeemassa vapaaehtoinen, mutta pakollinen PAS-palvelussa.
<BasicImageInformation>	P, ET	Elementti on MIX-skeemassa vapaaehtoinen, mutta pakollinen PAS-palvelussa.
<ImageAssessmentMetadata>	P, ET	Elementti on MIX-skeemassa vapaaehtoinen, mutta pakollinen PAS-palvelussa.
<ObjectIdentifier>	ES, ET	Elementti on MIX-skeemassa pakollinen, mutta PAS-palvelussa tämä esitetään PREMIS:llä, joten elementin käyttö MIX-skeemassa ei ole suositeltavaa toiston välttämiseksi.
<fileSize>	ES, ET	Elementti on MIX-skeemassa pakollinen, mutta tiedoston koko esitetään PAS-palvelussa PREMIS:llä, joten elementin käyttö MIX-skeemassa ei ole suositeltavaa toiston välttämiseksi.
<FormatDesignation>	ES, ET	Elementti on MIX-skeemassa pakollinen, mutta tiedostomuoto esitetään PAS-palvelussa PREMIS:llä, joten elementin käyttö MIX-skeemassa ei ole suositeltavaa toiston välttämiseksi.

¹⁰ MIX-määrittely [MIX] määrittelee jotkin elementit pakollisiksi, mutta MIX-skeemassa nämä määritellyt vapaaehtoisiksi. PAS-palvelussa pitää käyttää MIX:iä määrittelyn mukaisesti pois lukien tässä dokumentissa määritellyt poikkeukset.

Elementti / attribuutti	Esiintymä / velvoite	Säännöt ja suositukset
<FormatRegistry>	ES, ET	Elementti on MIX-skeemassa vapaaehtoinen, mutta lisätiedot tiedostomuodosta esitetään PAS-palvelussa PREMIS:llä, joten elementin käyttö MIX-skeemassa ei ole suositeltavaa toiston välttämiseksi.
<Fixity>	ES, T	Elementti on MIX-skeemassa pakollinen, mutta PAS-palvelussa tämä esitetään PREMIS:llä, joten elementin käyttö MIX-skeemassa ei ole suositeltavaa toiston välttämiseksi.

5.4.2 Esimerkki

```

<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>
              http://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 KONTROLLOITU SANASTO

Taulukossa 4 on esitetty kontrolloitu sanasto, jota pitää käyttää siirto- ja jakelupaketeissa ilmaistaessa tiedostomuoto ja sen versio. Taulukon sarakkeiden nimet viittaavat vastaaviin PREMIS-elementteihin [PREMIS]. Sanastoa tulee käyttää soveltuvin osin myös muissa siirto- tai jakelupaketin osissa (esim. METS:in MIMETYPE-attribuutti [METS]).

Taulukko 4: Tiedostomuodot

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
CSV	text/csv	-	x-fmt/18
EPUB	application/epub+zip	3	fmt/483
EPUB	application/epub+zip	2.0.1	fmt/483
XHTML	application/xhtml+xml	5	fmt/471
XHTML	application/xhtml+xml	1.1	fmt/103
XHTML	application/xhtml+xml	1.0	fmt/102
XML	text/xml	1.1	fmt/1776
XML	text/xml	1.0	fmt/101
HTML	text/html	5	fmt/471
HTML	text/html	4.01	fmt/100
IFC	model/step	4.3.2.0	fmt/700
IFC	model/step	4.0.2.1	fmt/700
ODT	application/vnd.oasis.opendocument.text	1.3	fmt/1756
ODT	application/vnd.oasis.opendocument.text	1.2	fmt/291
ODT	application/vnd.oasis.opendocument.text	1.1	fmt/290
ODT	application/vnd.oasis.opendocument.text	1.0	fmt/136
ODS	application/vnd.oasis.opendocument.spreadsheet	1.3	fmt/1755
ODS	application/vnd.oasis.opendocument.spreadsheet	1.2	fmt/295
ODS	application/vnd.oasis.opendocument.spreadsheet	1.1	fmt/294
ODS	application/vnd.oasis.opendocument.spreadsheet	1.0	fmt/137
ODP	application/vnd.oasis.opendocument.presentation	1.3	fmt/1754
ODP	application/vnd.oasis.opendocument.presentation	1.2	fmt/293
ODP	application/vnd.oasis.opendocument.presentation	1.1	fmt/292
ODP	application/vnd.oasis.opendocument.presentation	1.0	fmt/138
ODG	application/vnd.oasis.opendocument.graphics	1.3	fmt/1753
ODG	application/vnd.oasis.opendocument.graphics	1.2	fmt/297
ODG	application/vnd.oasis.opendocument.graphics	1.1	fmt/296
ODG	application/vnd.oasis.opendocument.graphics	1.0	fmt/139
ODF	application/vnd.oasis.opendocument.formula	1.3	-
ODF	application/vnd.oasis.opendocument.formula	1.2	-
ODF	application/vnd.oasis.opendocument.formula	1.0	-
PDF/A	application/pdf	A-3a	fmt/479
PDF/A	application/pdf	A-3b	fmt/480

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
PDF/A	application/pdf	A-3u	fmt/481
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-1a	fmt/95
PDF/A	application/pdf	A-1b	fmt/354
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	-	fmt/279
LPCM	audio/L8, audio/L16, audio/L20, audio/L24	-	-
AAC	audio/aac	-	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/H264	-	fmt/199
HEVC	video/H265	-	-
DNG	image/x-adobe-dng	1.5	fmt/1841
DNG	image/x-adobe-dng	1.4	fmt/730
DNG	image/x-adobe-dng	1.3	fmt/438
DNG	image/x-adobe-dng	1.2	fmt/437
DNG	image/x-adobe-dng	1.1	fmt/152
JPEG	image/jpeg	1.02	fmt/44
JPEG	image/jpeg	1.01	fmt/43
JPEG	image/jpeg	1.00	fmt/42
JPEG/EXIF	image/jpeg	2.3.2	fmt/1507
JPEG/EXIF	image/jpeg	2.3.1	fmt/1507
JPEG/EXIF	image/jpeg	2.3	fmt/1507
JPEG/EXIF	image/jpeg	2.2.1	fmt/645
JPEG/EXIF	image/jpeg	2.2	x-fmt/391
JPEG/EXIF	image/jpeg	2.1	x-fmt/390
JPEG/EXIF	image/jpeg	2.0	x-fmt/398
JP2	image/jp2	-	x-fmt/392
PNG	image/png	1.2	fmt/13
SVG	image/svg+xml	1.1	fmt/92
TIFF	image/tiff	6.0	fmt/353
WEBP	image/webp	-	fmt/566
WARC	application/warc	1.1	fmt/1281
WARC	application/warc	1.0	fmt/1355
GEOTIFF	image/tiff	1.0	fmt/155
GML	application/gml+xml	3.2.2	fmt/1047
GEOPACKAGE	application/geopackage+sqlite3	1.3.1	fmt/1700
GEOPACKAGE	application/geopackage+sqlite3	1.3.0	fmt/1700
KML	application/vnd.google-earth.kml+xml	2.3	fmt/244
SIARD	application/x-siard	2.2	fmt/1777
SIARD	application/x-siard	2.1.1	-
POR	application/x-spss-por	-	fmt/997

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
Matlab	application/matlab	7.3	fmt/828
Matlab	application/matlab	7	fmt/806
HDF5	application/x-hdf5	1.10	fmt/807, fmt/286, fmt/287
MS Word	application/vnd.openxmlformats-officedocument.wordprocessingml.document	2007 onwards	fmt/412
MS Word	application/msword	97-2003	fmt/40
MS Excel	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	2007 onwards	fmt/214
MS Excel	application/vnd.ms-excel	8X	fmt/62
MS Excel	application/vnd.ms-excel	8	fmt/61
MS Powerpoint	application/vnd.openxmlformats-officedocument.presentationml.presentation	2007 onwards	fmt/215
MS Powerpoint	application/vnd.ms-powerpoint	97-2003	fmt/126
PDF	application/pdf	1.7	fmt/276
PDF	application/pdf	1.6	fmt/20
PDF	application/pdf	1.5	fmt/19
PDF	application/pdf	1.4	fmt/18
PDF	application/pdf	1.3	fmt/17
PDF	application/pdf	1.2	fmt/16
AIFF-C	audio/x-aiff	-	x-fmt/136
MP3	audio/mpeg	2	fmt/134
MP3	audio/mpeg	1	fmt/134
WMA	audio/x-ms-wma	9	fmt/132
DV	video/dv	-	x-fmt/152
MPEG	video/mpeg	2	fmt/640
MPEG	video/mpeg	1	fmt/649
WMV	video/x-ms-wmv	9	fmt/133
EPS	application/postscript	3.0	fmt/124
GIF	image/gif	1989a	fmt/4
GIF	image/gif	1987a	fmt/3

Taulukoissa 5 ja 6 on kuvattu PAS-palvelun hyväksymät säilytys- ja siirtokelpoiset säiliömuodot ja mitä ääni- ja kuvaraitojen muotoja saa sisällyttää säiliöihin. Jos säiliömuodossa käytetään sekä säilytys- että siirtokelpoisia tiedostomuotoja, niin PAS-palvelu käsittelee säiliömuotoa kuten siirtokelpoisia tiedostomuotoja.

Taulukko 5: Säilytyskelppoiset säiliömuodot

Säiliömuoto	formatName	Ääni	Kuva	formatRegistryKey
Matroska v4 ¹¹	video/x-matroska	FLAC, LPCM	FFV1, HEVC	fmt/569
MPEG-2 (TS)	video/MP2T	AAC	AVC, HEVC	fmt/585
MPEG-4	video/mp4	AAC	AVC, HEVC	fmt/199
MXF	application/mxf	AAC, LPCM	AVC, JP2	-
MJPEG2000	video/mj2	LPCM	JP2	fmt/337
MOV	video/quicktime	AAC, LPCM	AVC, HEVC, JP2	x-fmt/384

¹¹ Matroska säiliömuodosta hyväksytään vain versio 4

Taulukko 6: Siirtokelpoiset säiliömuodot

Säiliömuoto	formatName	Ääni	Kuva	formatRegistryKey
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 Tekstitiedoston sallitut merkistöt

Tekstimuotoisessa tiedostossa (CSV (2.1.1), XHTML (2.1.3), XML (2.1.4), HTML (2.1.5), plain text (2.1.9), GML (2.6.2) ja KML (2.6.3)) käytettävä merkistö pitää ilmaista PREMIS formatName -elementissä käyttäen charset-parametria. Charset-parametrin mahdolliset arvot ovat ISO-8859-15 [ISO-8859-15], UTF-8, UTF-16 ja UTF-32 [UNICODE].

Esimerkki:

```
<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>
```

Tekstimuotoinen tiedosto ei saa sisältää binääridataa. Jos tiedostoon on tarvetta sisällyttää binääridataa, se pitää Base64-koodata. Base64-koodatulle datalle PAS-palvelu tarjoaa vain bittitason säilytystä.

6.2 Tiedostomuodon tarkentaminen

PAS-palvelu tukee luvuissa 2 ja 3 kuvattuja tiedostomuotoja. On kuitenkin olemassa suuri joukko erilaisia tiedostomuotoja, joilla on oma MIME-tyyppi, jota PAS-palvelu ei tue, mutta jotka kuitenkin ovat pohjimmiltaan jotakin PAS-palvelun tukemaa tiedostomuotoa.

Esimerkiksi tekstitiedostojen tapauksessa tällaisia tiedostomuotoja ovat muun muassa text/css, application/javascript ja application/json. PAS-palvelu hyväksyy tällaiset tiedostomuodot säilytykseen tekstitiedostoina validoiden ne kuten text/plain -tiedostot. Tällaisissa tapauksissa tiedostomuodoksi ilmoitetaan text/plain, ja todellinen tiedostomuoto voidaan ilmaista PREMIS formatName -elementissä käyttäen alt-format -parametria. Vastaavalla tavalla voidaan tarkentaa lukuisa joukko erilaisia XML-pohjaisia tiedostomuotoja.

PAS-palvelu ei hyödynnä alt-format -parametrin arvoa.

Esimerkkejä:

```
<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>
```

Jos tekstitiedoston MIME-tyypiksi on asetettu text/plain, mutta PAS-palvelun vastaanotto tunnistaa sen joksikin toiseksi tekstitiedostotyyppiksi, tiedosto hyväksytään säilytykseen text/plain -tyyppinä, mutta vastaanoton raporttiin kirjataan huomio siitä, että toinen tiedostomuoto tunnistettiin ensiksi, mutta tiedostoa säilytetään kuitenkin text/plain-muodossa.

LIITE A. TIEDOSTOMUODOISTA

Tiedostomuodot ovat määrämuotoja ja rakenteita, jotka järjestävät ja määrittelevät dataa. Jotkut tiedostomuodot sisältävät vain yhden, pakkaamattoman datavirran; jotkut voivat sisältää *koodekkeja*, joilla dataa voidaan koodata ja pakata; ja jotkut voivat tukea useampia datavirtoja.

Eräs tiedostomuotojen osajoukko on *säiliöformaattit*. Nämä formaatit voivat sisältää ja tukea monen tyyppistä datavirtaa ja täten käytännössä sisältää ääntä, kuvaa ja elävää kuvaa ja näiden tasoja (layer) sekä niihin liittyvää metadataa. Kutakin näistä tiedostomuodoista voidaan käsitellä erilaisin ohjelmistoin, laittein ja prosessein, mutta jotta multimediallinen datavirta voitaisiin tulkita oikein, informaation tulee olla kapseloitu yhteen. PAS-palvelussa säilytettävät säiliöformaattit voivat sisältää vain tässä määrittelyssä lueteltuja säilytys- ja siirtokelpoisia tiedostomuotoja. Yhdysvaltain kongressin kirjasto (Library of Congress) jakaa säiliöformaattit kolmeen ryhmään:

- **”Kääreformaattit” (wrappers).** Digitaalisen pitkäaikaissäilytyksen piirissä kääreiksi kutsutaan tiedostomuotoja, jotka kapseloivat bittivirtoja sekä sisältävät metadataa, joka kuvailee kääreen sisällön. TIFF ja WAVE ovat tavallisia kääreformaatteja. Kääretiedoston sisältämät datavirrat voivat olla hyvin erityyppisiä. Esimerkiksi WAVE-tiedosto voi sisältää (a) pakkaamattomia LPCM-audiotiedostoja, (b) äärimmäisen pakattuja äänitiedostoja tai (c) muita äänitiedostoja. Kääreformaatin itsensäselittävä sisältönsä selostava ominaisuus taas ilmenee esimerkiksi TIFF-otsikoissa (TIFF header). Verraten monimutkaiset ja monikäyttöiset kääreformaattit kuten QuickTime (QTFF) voivat sisältää useita objekteja, esim. yhden tai useampia leikkeitä elävästä kuvasta sekä erillisiä äänileikkeitä.
- **”Yksinkertaiset nippuformaattit” (simple bundling formats).** Nämä tiedostomuodot paketoivat sisältämänsä tiedostot mutta eivät – tiedostonimien listan sisältävän hakemiston lisäksi – kuvaile sisältöjä tai tiedostojen välisiä suhteita. Esimerkkejä tällaisista formaateista ovat ZIP, Stuffit ja TAR. Yksinkertaiset nippuformaattit ovat yleensä geneerisiä, eli niillä voidaan paketoida kaikenlaisia tiedostomuotoja.
- **”Rakenteiset nippuformaattit” (self-describing bundling formats).** Näitä tiedostomuotoja käytetään esittämään sellainen tiedostojen joukko, joka muodostaa kompleksisen digitaalisen objektin – esimerkiksi kirjan kuvituksineen; tai elokuvan, joka koostuu monista elävän kuvan segmenteistä sekä ääniraidasta useilla eri kielillä. Rakenteiset nippuformaattit sisältävät listan osistaan ja niiden suhteista (rakenteellinen metadata) ja voivat ilmaista, miten kokonaisuus voidaan esittää tai miten sitä muuten voi käyttää. Nämä tiedostomuodot sisältävät usein teknisiä tietoja kustakin osastaan erikseen, koska yksi digitaalinen objekti voi sisältää monenlaista tekstiä, ääntä, kuvaa jne. Ne voivat kapseloida (mutta eivät välttämättä kapseloi) sisältämänsä tiedostot. Niihin sisältyy metadataa, joka kuvailee niiden sisällön sekä tiedostojen väliset suhteet. Esimerkkejä tällaisista formaateista ovat METS ja MPEG-21.

LIITE B. TIEDOSTOMUOTOJEN PYSYVYYDEN ARVIOINTI

Tiedostomuotojen huolellinen valinta on tärkeää pitkäaikaissäilytyksen onnistumiseksi. PAS-palveluissa valinnassa on pyritty tasapainottamaan eri vaatimuksia, joita tiedostomuodoille on asetettava: laatu, vakaus, todennäköinen pitkäikäisyys ja levinneisyys. Etusija on pyritty antamaan kansallisille ja kansainvälisille vapaille standardeille. Mikäli vapaita standardeja ei ole ollut tarjolla, on hyväksytty tiedostomuotoja, jotka ovat de facto -standardeja. De facto -standardilla tarkoitetaan laajasti käytettyjä ja tunnettuja tiedostomuotoja, joista on käytännössä muodostunut alansa standardeja niiden yleisyyden ja tuettuuden takia, vaikka ne eivät ole yhdenkään standardiorganisaation muodollisesti hyväksymiä. Tähän dokumenttiin on myös valittu tiedostomuotoja, joiden uskotaan yleistävän digitaalisen säilyttämisen piirissä lähitulevaisuudessa. Seuraavassa esitetään tarkemmin, miten tiedostomuotojen pysyvyyttä on arvioitu.

B.1. Arviointikriteerit

Alla on esitetty kriteerit tiedostomuotojen arvioimiseksi pitkäaikaissäilytyksen ja käytettävyyden turvaamisen kannalta. Ne perustuvat Yhdysvaltain kongressin kirjaston (Library of Congress), Yhdistyneen kuningaskunnan kansallisarkiston (National Archives), Alankomaiden kansalliskirjaston (Koninklijke Bibliotheek) sekä Kanadan kansalliskirjaston ja -arkiston (Library and Archives Canada) selvityksiin.

Avoimuus	Kuinka helppoa tiedostomuodosta on saada tietoja?
Käyttö PAS-standardina	Missä määrin tiedostomuoto on muodollisesti hyväksytty pitkäaikaissäilytyksen välineeksi kansalliskirjastoissa, kansallisarkistoissa ja muissa alan laitoksissa?
Vakaus / yhteensopivuus	(a) Missä määrin tiedostomuoto on eteen- ja taaksepäin yhteensopiva? (b) Missä määrin tiedostomuoto on suojattu tiedoston korruptoitumista vastaan? (c) Kuinka usein tiedostomuodosta julkaistaan korvaavia versioita?
Riippuvuudet / yhteentoimivuus	Missä määrin tiedostomuoto on sidottu esimerkiksi tiettyyn laitteistoon tai ohjelmistoon?
Standardointi	Missä määrin tiedostomuoto on käynyt läpi perusteellisen standardointiprosessin?

Arvioinnissa on käytetty seuraavaa asteikkoa:

A	Arviointikriteeri täyttyy hyvin
A[€]	Arviointikriteeri täyttyy hyvin; siihen liittyy kuitenkin kuluja (esim. maksullinen dokumentaatio)
B	Arviointikriteeri täyttyy kohtalaisesti
C	Arviointikriteeri ei täyty
A/C	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) muttei toisella sektorilla (esim. asiakirja-aineisto)
A/B	Arviointikriteeri täyttyy hyvin yhdellä sektorilla (esim. elektroniset vapaakappaleet) mutta vain kohtalaisesti toisella sektorilla (esim. asiakirja-aineisto)

Seuraavassa taulukossa 7 on määritelty kriteerit sekä niiden arvostelu tarkemmin.

Taulukko 7: Tiedostomuotojen arviointiohje

Kriteeri	Arviointiohje	Arvio
1. Avoimuus	Määrittelyt saatavissa yhdeltä tai useammalta seuraavista: (a) avoimen jäsenyyden järjestö [kuten W3C (World Wide Web Consortium) tai OMG (Object Management Group)] (b) kansainvälinen standardointijärjestö (esim. ISO) (c) tuotannonalan avoimen jäsenyyden järjestö	A
	Määrittelyt saatavilla vain maksusta	A [€]
	Määrittelyt mahdollisesti saatavilla useista lähteistä (ei saatu vahvistettua)	B
	Määrittelyjä jakelee tai niiden jakelua valvoo yksi kaupallinen toimija tai pieni kaupallisten toimijoiden joukko.	C
2. Käyttö PAS-standardina	Merkittävä osa (yli puolet) organisaatioista käyttää tai aikoo käyttää tiedostomuotoa säilytyskelpoina tiedostomuotona	A
	Jotkut organisaatioista (alle puolet) käyttävät tai aikovat käyttää tiedostomuotoa säilytyskelpoina tiedostomuotona	B
	Yksikään organisaatio ei käytä eikä aio käyttää tiedostomuotoa säilytyskelpoina tiedostomuotona	C
3. Vakaus / yhteensopivuus		
(a) Alas-/ylöspäin-yhteensopivuuden taso	Tiedostomuodon versio on alaspäin yhteensopiva, jos se sisältää kaikki aiemman version toiminnallisuudet. Tiedostomuodon versio on ylöspäin yhteensopiva, jos sen avulla voi ongelmitta tallentaa sisältöä, joka on tarkoitettu tiedostomuodon uudemmalle versiolle (kääntäen: ohjelmisto, joka on suunniteltu tulkitsemaan tai näyttämään tiedostomuodon aiempaa versiota, pystyy tulkitsemaan tai näyttämään myös tiedostomuodon nykyistä versiota).	
	Hyvä yhteensopivuus – tiedostomuoto on yhteensopiva niin ylös- kuin alaspäinkin.	A
	Kohtalainen yhteensopivuus – tiedostomuoto on vain alaspäin yhteensopiva.	B
	Huono yhteensopivuus – tiedostomuoto ei ole yhteensopiva ylös- eikä alaspäin.	C
(b) Korruptoitumisen sieto	Korruptoitumisen sieto tarkoittaa sitä, että tiedostomuoto sietää sisällön bitti- tai tavutason satunnaisia muutoksia.	
	Hyvä sietokyky – Muutokset eivät juurikaan tai lainkaan vaikuta tiedostomuodon näyttämiseen tai tulkintaan; tai tiedostomuoto sisältää menetelmiä, joilla muutokset havaitaan ja/tai korjataan	A

Kriteeri	Arviointiohje	Arvio
	Kohtalainen sietokyky – Muutokset vaikuttavat tiedostomuodon näytettävyyteen mutta eivät tulkittavuuteen; tiedostomuoto voi jossakin määrin palautua muutoksista.	B
	Huono sietokyky – Kaikki muutokset vaikuttavat näytettävyyteen ja tulkittavuuteen.	C
(c) Versiopäivitysten määrä	Tiedostomuodon vakaus ilmaistuna uusien versioiden tai laajennusten lukumääränä; tiedostomuodon käyttö johdannaisissa ja/tai tuotannonalakohtaisissa sovelluksissa.	
	Suuri vakaus	A
	Keskitason vakaus	B
	Epävakaus	C
4. Riippumattomuus, yhteentoimivuus	Korkea riippumattomuus & yhteentoimivuus Korkea riippumattomuus & keskitason yhteentoimivuus Keskitason riippumattomuus & korkea yhteentoimivuus	A
	Korkea riippumattomuus & alhainen yhteentoimivuus Keskitason riippumattomuus & yhteentoimivuus Keskitason riippumattomuus & alhainen yhteentoimivuus	B
	Alhainen riippumattomuus & yhteentoimivuus Alhainen riippumattomuus & keskitason yhteentoimivuus Alhainen riippumattomuus & korkea yhteentoimivuus	C
5. Standardointi	Tiedostomuotoa sääntelee jonkin seuraavista hallitsema muodollinen prosessi: ▪ Avoimen jäsenyyden organisaatio (W3C, OMG yms.) ▪ Kansainvälinen standardiorganisaatio (esim. ISO) ▪ Tuotannonalakohtainen avoimen jäsenyyden organisaatio	A
	Tiedostomuotoa sääntelee dokumentoitu prosessi, jonka on luonut yksittäinen kaupallinen toimija tai pieni kaupallisten toimijoiden joukko; tai tiedostomuotoa koskevia prosesseja ei ole dokumentoitu.	C

B.2. Yhteenveto säilytyskelpoisten tiedostomuotojen arvioinnista

Taulukossa 8 on vedetty yhteen säilytyskelpoisten tiedostomuotojen arvioinnit.

Taulukko 8: Tiedostomuotojen arviointiohje

Sisältö	Tiedostomuoto	Avoimuus	Käyttö PAS-standardina	Alas-/ylöspäin yhteensopivuuden taso	Korruptoitumisen sieto	Versiopäivitysten määrä	Riippumattomuus / yhteentoimivuus	Standardointi
TEKSTI	Comma Separated Values (CSV)	A	A	A		A	A	A
	Electronic Publications (EPUB)	A	B	C		A	B	A
	Extensible Hypertext Markup Language (XHTML)	A	B	B		A	A	A
	Extensible Markup Language (XML)	A	A	A		A	A	A
	Hypertext Markup Language (HTML)	A	A	B		A	A	A
	Industry Foundation Classes (IFC)							
	Open Document Format (ODF)	A	A	B		B	A	A
	PDF for long-term preservation (PDF/A)	A [€]	A	B		A	A	A
	Tekstitiedosto (plain text)	A [€]	A	B		A	A	A
ÄÄNI	Audio Interchange File Format (AIFF), PCM-koodattu	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
ELÄVÄ KUVA	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 -sekvenssi	A [€]	A	A	A	A	A	A
	MPEG-4 AVC – Advanced Video Coding (AVC)	A	B	A		A	A	A
	MPEG-H HEVC – High Efficiency Video Coding (HEVC)	A	B	B	B	B	A	A
KUVA	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
	WebP Image Format (WEBP)	A	B			A	A	A
VERKKOARKISTO	Web ARChive Format (WARC)	A	A		B	A	A	A
PAIKKATietoaineistot	Geographic Tagged Image File Format (GeoTiff)	A	A			A	A	A
	Geography Markup Language (GML)							
	Geopackage	A	B	A		A	A	A

Sisältö	Tiedostomuoto	Avoimuus	Käyttö PAS-standardina	Alas-/ylöspäin yhteensopivuuden taso	Korruptoitumisen sieto	Versiopäivitysten määrä	Riippumattomuus / yhteentoimivuus	Standardointi
	Keyhole Markup Language (KML)							
TIETOKANNAT	Software Independent Archiving of Relational Databases (SIARD)	A ¹²	B	B	C	B	B	A
TUTKIMUSAINEISTOT	SPSS Portable (POR)	A [€]	A	A	A	A	A	C
TIETEELLISET OHJELMISTOT	Matlab	A/C	B	B	C	B	B	C
	Hierarchical Data Format (HDF5)	A	B	B	A	A	B	A

¹² Määrittely on ilmaiseksi ja julkisesti saatavilla. Tiedostomuodon kehitystä kuitenkin kontrolloin ei-kaupallinen organisaation (Swiss Federal Archives), jolla ei ole avointa jäsenyyspolitiikkaa.

LIITE C. MUIDEN KUIN SÄILYTYS- JA SIIRTOKELPOISTEN TIEDOSTOMUOTOJEN KÄSITTELY PAS-PALVELUSSA

Tässä liitteessä määritellään mitä mahdollisuuksia ja millä ehdoin hyödyntävä organisaatio voi siirtää PAS-palveluun tiedostoja, jotka eivät ole säilytys- tai siirtokelpoisia tiedostomuotoja. Tällaiset tapaukset voidaan jakaa karkeasti kolmeen tapaukseen:

- (1) Säilytyskelpoisen tiedoston rinnalle halutaan säilyttää vastaava tiedosto alkuperäisessä tiedostomuodossa, jossa alkuperäinen tiedostomuoto ei ole säilytys- tai siirtokelpoinen. Tämä tehdään alkujaan digitaalisen aineiston autenttisuuden takaamiseksi.
- (2) Säilytyskelpoisen tiedoston rinnalle halutaan säilyttää tiedosto säilytyskelpoisesta johdetussa versiossa, joka ei ole säilytys- tai siirtokelpoinen. Tämä tehdään, koska johdettu tiedosto on aineiston käytön kannalta oleellinen ja sen uudelleen luominen ei ole mahdollista tai hyvin epäkäytännöllistä.
- (3) Tiedostomuotoa ei ole mahdollista, tai on hyvin epäkäytännöllistä, muuntaa säilytys- tai siirtokelpoiseen tiedostomuotoon, mutta tiedosto kuitenkin katsotaan tarpeelliseksi säilyttää.

C.1. Vaatimukset ja suositukset

Tapauksesta riippumatta, PAS-palvelu tarjoaa muille kuin säilytys- ja siirtokelpoisille tiedostoille vain bittitason säilyttämistä, jossa huolehditaan ainoastaan aineiston eheydestä ja turvallisesta säilytyksestä. Näissä tapauksissa ymmärtämisen säilyttämisen jää täysin hyödyntävän organisaation vastuulle.

Koska ymmärrettävyyden säilyttäminen jää hyödyntävän organisaation vastuulle, PAS-palvelu varmistaa aineiston oikeellisuuden ja hyödyntävän organisaation kyvykkyyden aineiston ymmärrettävyyden säilyttämiseen vaatimalla joko säilytyskelpoisen tiedostomuodon rinnalle (tapaukset 1 ja 2) tai soveltamalla tiukat kriteerit aineiston oikeellisuuden varmistamiseksi (tapaus 3). Tapauksissa 1 ja 2 PAS-palvelu säilyttää ymmärrettävyyden säilytyskelpoiselle tiedostomuodolle ja sen ominaisuuksille.

Taulukossa 9 on esitetty vaatimukset ja suositukset yllä mainituissa tapauksissa (+=suositeltavaa; *=vaaditaan). Jos tiedosto ei ole säilytys- tai siirtokelpoinen, tiedoston tulee täyttää taulukossa 9 esitetyt vaatimukset sekä mahdollisimman monta suosituksesta. Niillä varmistetaan aineiston eheys ja ymmärrettävyys säilytykseen siirrettäessä sekä pyritään käyttökelpoisuuden säilymiseen mahdollisimman pitkään. Siirto- ja säilytyskelpoisten tiedostomuotojen osalta vaatimusten mukaisuus on tarkistettu jo tiedostomuotoja valittaessa.

Taulukko 9: Vaatimukset ja suositukset

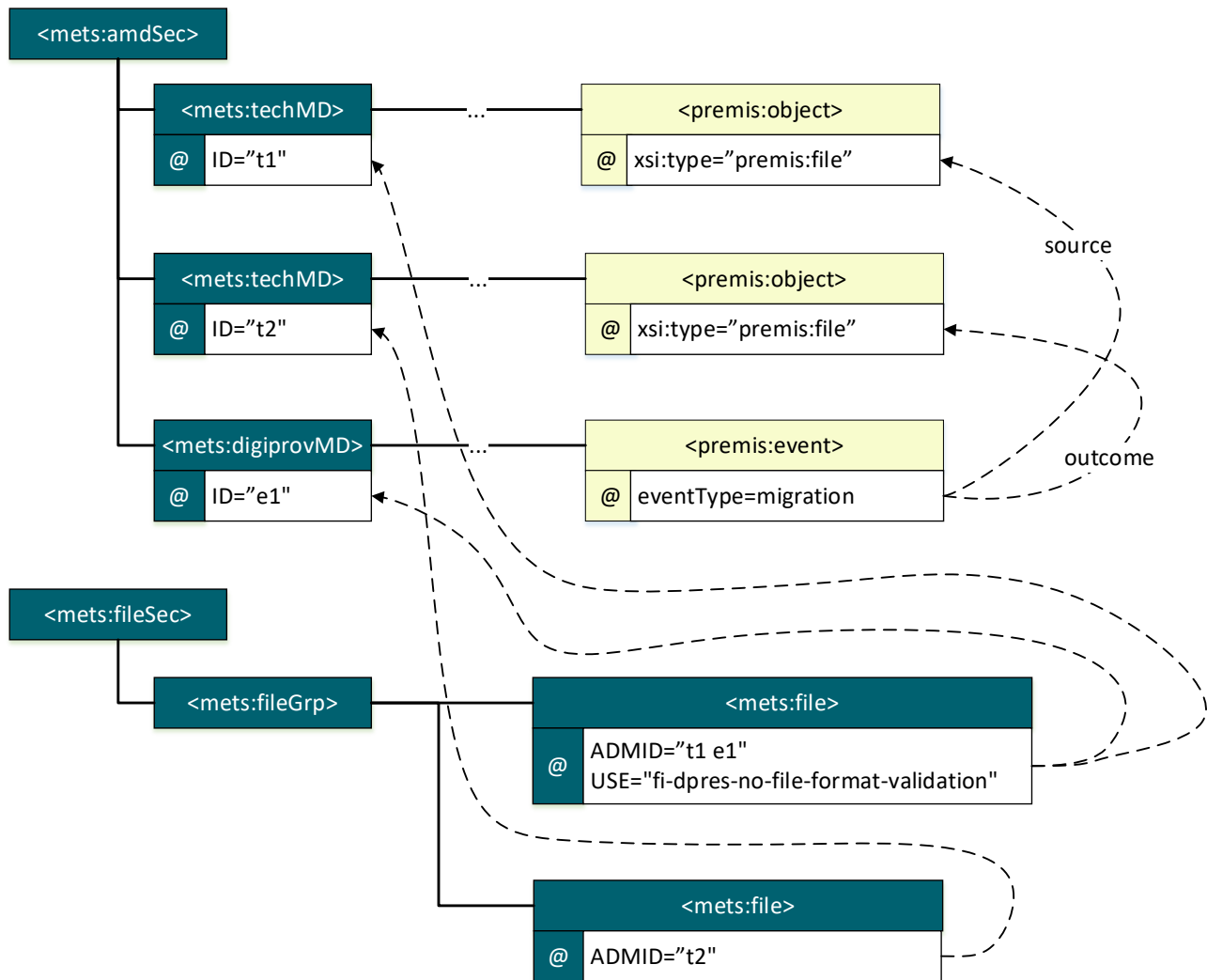
Numero	Vaatus / suositus	(1)	(2)	(3)
V1	PAS-palvelu on hyväksynyt tiedostomuodon siirtämisen säilytykseen. ▪ Hyödyntävä organisaatio arvioi yhdessä PAS-palvelun kanssa, onko tiedostomuodon käyttäminen perusteltua vai voitaisiinko tiedostot muuntaa säilytyskelpoiseen muotoon. ▪ Hyväksyntä on aina sopimuskohtaista.	*	*	*
V2	Tiedostomuoto on tuettu vähintään yhdessä yleisesti saatavilla olevassa ohjelmassa. ▪ Ohjelma voi olla maksullinen tai maksuton. Mikäli tiedostojen käsittelyyn tarvitaan erikoisohjelma, tulee ohjelman nimi ja linkki ohjelman kotisivulle mainita.	+	+	*

Numero	Vaatus / suositus	(1)	(2)	(3)
	Mikäli vain hyödyntävän organisaation itse kehittämä ohjelma tukee tiedostomuotoa, ohjelma ja sen lähdekoodi on liitettävä mukaan aineistokokonaisuuteen.			
V3	Tiedostomuodon rakenne on dokumentoitu. Mikäli rakenne ei ole tiedossa esimerkiksi valmistajakohtaisten suljettujen tiedostomuotojen tapauksessa, dokumentaation tulee sisältää mahdollisimman tarkat tiedot tiedostomuodosta.	+	+	*
V4	Mikäli tiedostomuodon tulkinta edellyttää metatietoja, jotka eivät ole pääteltävissä itse tiedostosta, ne on tallennettava erikseen koneluettavaan muotoon tiedostojen yhteyteen.	+	+	*
V5	Tiedostomuoto voidaan tunnistaa koneellisesti.	+	+	*
V6	Tiedostomuodolla pitää olla MIME tyyppi.	+	+	*
V7	Tiedostomuodolla on tunnus PRONOM-tiedostomuotokirjastossa [PRONOM]. Mikäli tunnusta ei ole valmiiksi, se tulee hakea PRONOM-palvelun ohjeiden mukaisesti.	+	+	+
V8	Tiedostomuoto on alalla laajasti käytetty.	+	+	+
V9	Tiedostomuoto on riippumattoman organisaation tai alan yhteisön standardoima. Standardoidut tiedostomuodot voidaan mahdollisesti hyväksyä myös siirto- tai säilytyskelloiksi. Asiasta tulee keskustella PAS-palvelun kanssa.	+	+	+
V10	Tiedostomuodon eheyden tarkistamiseen on olemassa validaattori. Validointi suoritetaan normaalisti PAS-palvelun vastaanotossa. Mikäli validaattori ei esim. ohjelmistoteknisistä syistä toimi PAS-palvelussa, validointi voidaan suorittaa myös hyödyntävän organisaation omassa järjestelmässä ennen säilytykseen siirtoa.	+	+	+
V11	Tiedostomuodon käsittelyyn on olemassa skripti tai toimintaohje, jonka avulla tallennettuja tietoja voidaan lukea johonkin tunnettuun sovellusohjelmaan. Skripti tai ohje tulee sisällyttää säilytettävään aineistokokonaisuuteen.	+	+	+

C.2. Metatietovaatimukset

Jos säilytyskelloisen tiedoston rinnalle halutaan säilyttää vastaava tiedosto alkuperäisessä tiedostomuodossa, joka ei ole säilytys- tai siirtokelpoinen (tapaus 1), PITÄÄ nämä tiedostot linkittää toisiinsa METS/PREMIS tasolla käyttäen "migration"- tai "normalization"-tapahtumaa. Tässä tapahtumassa säilytyskelloinen tiedostomuoto on tapahtuman "outcome" ja alkuperäinen tiedostomuoto "source", ja

alkuperäisen tiedostomuodon METS file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa "fi-dpres-no-file-format-validation". Tämä on havainnollistettu kuvassa 2¹³.



Kuva 2: Säilytyskelvoksen tiedoston rinnalla säilytettävä tiedosto alkuperäisessä tiedostomuodossa

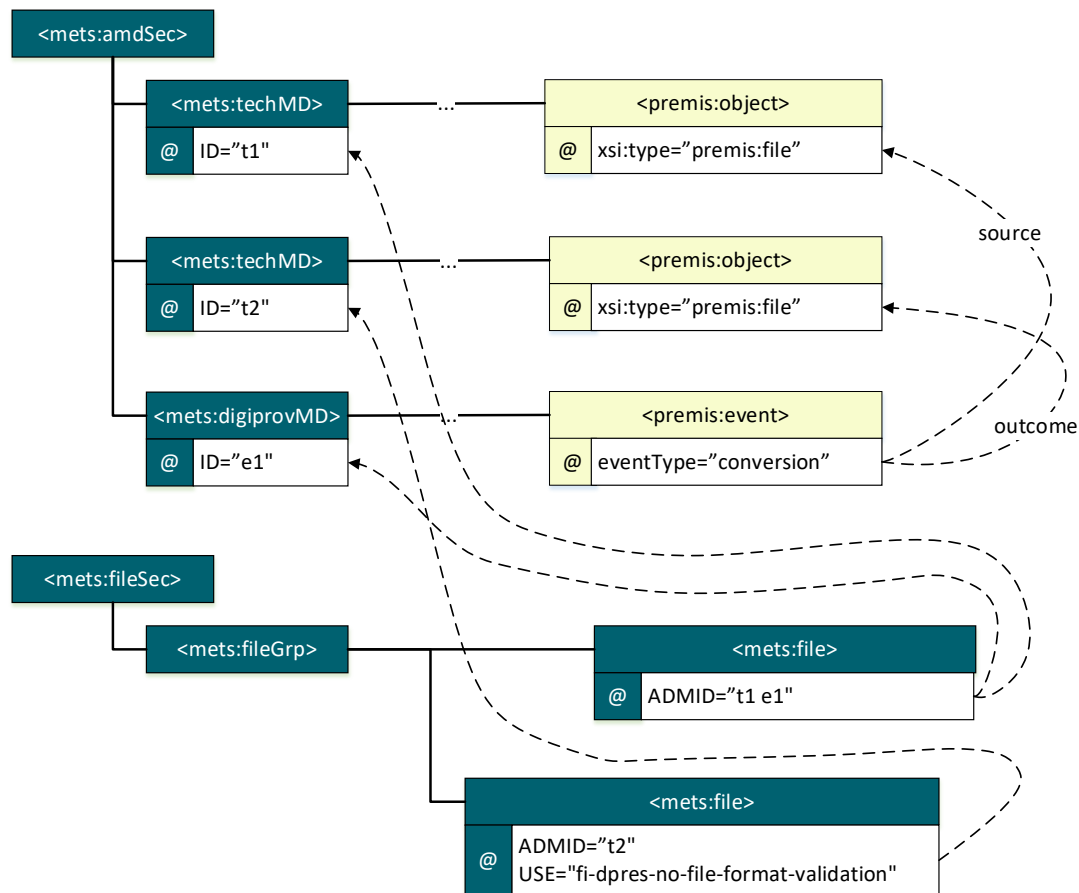
Jos säilytyskelvoksen tiedoston rinnalle halutaan säilyttää tiedosto säilytyskelvoksesta johdetussa versiossa, joka ei ole säilytys- tai siirtokeelpoinen (tapaus 2), on nämä tiedostot linkitettävä toisiinsa käyttäen "conversion" -tapahtumaa. Tässä tapahtumassa säilytyskelvoinen tiedosto on tapahtuman "source" ja rinnalle talletettava tiedostomuoto "outcome", ja muunnetun tiedostomuodon METS-dokumentin file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa "fi-dpres-no-file-format-validation". Tämä on havainnollistettu kuvassa 3¹⁴.

Jos tiedostomuotoa ei ole mahdollista, tai on hyvin epäkäytännöllistä, muuntaa säilytys- tai siirtokeelpoiseen tiedostomuotoon, mutta tiedosto kuitenkin katsotaan tarpeelliseksi säilyttää (tapaus 3), METS-dokumentin file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa "fi-dpres-file-format-identification". Tämä on havainnollistettu kuvassa 4.

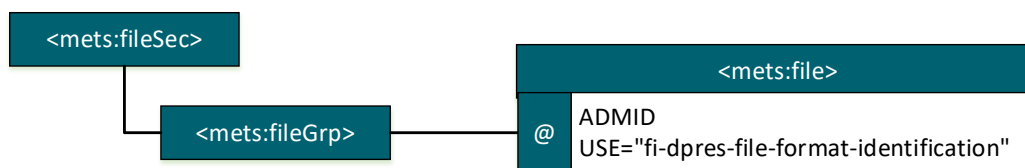
¹³ Konkreettinen XML-tason esimerkki tällaisesta METS-dokumentista löytyy osoitteesta <https://digitalpreservation.fi/support/examples>

¹⁴ Konkreettinen XML-tason esimerkki tällaisesta METS-dokumentista löytyy osoitteesta <https://digitalpreservation.fi/support/examples>

Kaikissa ylläolevissa tapauksissa, jos alkuperäinen tiedostomuoto on säiliömuoto, joka sisältää siirto- tai säilytyskelpoisen kuva- tai ääniraidan, niin tällöin kuva- tai ääniraidalta ei vaadita teknisen metatiedon lohkoa.



Kuva 3: Säilytyskelpoisen tiedoston rinnalla säilytettävä tiedosto johdetussa tiedostomuodossa



Kuva 4: Alkuperäistä tiedostoa ei ole mahdollista muuntaa säilytyskelpoiseen tiedostomuotoon

LIITE D. PAIKALLISET XML-SKEEMATIEDOSTOT

Tässä liitteessä määritellään mitä mahdollisuuksia ja millä ehdoin hyödyntävä organisaatio voi siirtää PAS-palveluun XML-skeematiedostoja, jotka ovat säilytettävän aineiston kannalta oleellisia. Jos säilytettävä aineisto käyttää XML-skeemaa rakenteensa kuvaamiseen, on itse skeematiedosto oltava hyödynnettävissä aineiston yhteydessä, sekä validoinnissa että säilytyksen suunnittelun ja säilytystoimenpiteiden (kuten migraatioiden) aikana.

PAS-palvelu ylläpitää katalogia yleisistä kansallisista- ja kansainvälisistä XML-skeemoista. Tämän katalogin skeemoja hyödynnetään aineiston prosessoinnin yhteydessä. Jos aineisto käyttää sellaista XML-skeemaa, joka ei ole PAS-palvelun ylläpitämässä katalogissa, on hyödyntävällä organisaatiolla kaksi vaihtoehtoa:

- (1) XML-skeematiedosto toimitetaan erikseen PAS-palvelulle, joka lisää sen skeemakatalogiin. Tästä pitää aina sopia erikseen PAS-palvelun kanssa, ja käytännössä tämä voidaan tehdä vain julkaistuille ja yleisesti useamman organisaation tai tahon käytössä oleville skeemoille. Nämä skeemat on myös oltava julkaistavissa, koska PAS-palvelun skeemakatalogi on julkinen.
- (2) XML-skeematiedostot toimitetaan osana aineiston siirtopakettia. Tämä lähestymistapa suositellaan paikallisille, lähinnä yhden organisaation sisäisesti käytössä oleville skeematiedostoille. Tämä vaihtoehto kuvataan tarkemmin alla.

XML-skeematiedostot, jotka toimitetaan siirtopaketissa, on oltava valideja XML-tiedostoja, ja niitä validoidaan ja säilytetään PAS-palvelussa muun aineiston ohella. Niitä on siirtopaketin metatiedoissa kuvailtava teknisillä metatiedoilla aivan kuten muu säilytettävä aineisto (PREMIS-metatiedoilla).

Koska XML-tiedostot viittaavat skeemoihin nimiavaruuksien ja URL-polkujen avulla, jotka eivät vastaa siirtopaketin sisäisiä relatiivisia tiedostopolkuja, on siirtopaketin mukana toimitettavat paikalliset XML-skeemat voitava yhdistää XML-tiedostoihin siirtopaketin metatietojen avulla. Tätä varten PITÄÄ siirtopaketin metatietoihin luoda PREMIS-representaatio (`premis:object xsi:type="premis:representation"`), joka sisältää linkkejä skeematiedostojen ja XML-tiedostojen välillä. Tämä representaatio sisältää yhden ympäristölohkon (`premis:environment`), joka sisältää linkkejä skeemojen ja XML-tiedostojen välillä. Ympäristölohkon PITÄÄ sisältää `<premis:environmentPurpose>`-elementin, jonka arvo on "xml-schemas".

Jokaiselle siirtopaketin mukana olevalle XML-skeemalle luodaan oma riippuvuuslohko (`premis:dependency`) ympäristölohkon sisällä. Lohkoissa kuvataan, kuinka siirtopaketin mukana lähetetty XML-skeemat voidaan yhdistää XML-tiedostoissa käytettyihin skeemapolkuihin. Riippuvuuslohkon `<premis:dependencyName>`-elementissä PITÄÄ antaa paikallisen XML-skeematiedoston relatiivinen polku siirtopaketin juuresta. Riippuvuuslohkon `<premis:dependencyIdentifier>`-elementissä kuvataan se skeemapolku, jota käytetään XML-tiedostoissa (`xsi:schemaLocation` tai `xsi:noNamespaceSchemaLocation`).

Esimerkki

```
<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>

```

Jotta XML-skeematiedosto(t) voidaan semanttisesti erotella varsinaisesta säilytettävästä aineistosta siirtopaketissa, on suositeltavaa, että skeematiedostot merkitään siirtopaketin metatiedoissa ns. oheismateriaalina. Tämä tapahtuu lisäämällä XML-skeematiedostot siirtopaketin METS-dokumentin <fileSec>-lohkossa omaan <fileGrp>-elementtiin, jonka attribuutin USE arvoksi SAA laittaa ”fi-dpres-xml-schemas”.

Esimerkki

```

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

```

Skeemalinkkejä sisältävän PREMIS-representaation SAA laittaa omaan rakennekarttaan (<mets:structMap>), jonka ensimmäisen <div>-elementin attribuutin TYPE arvoksi tulee ”fi-dpres-supplementary”.

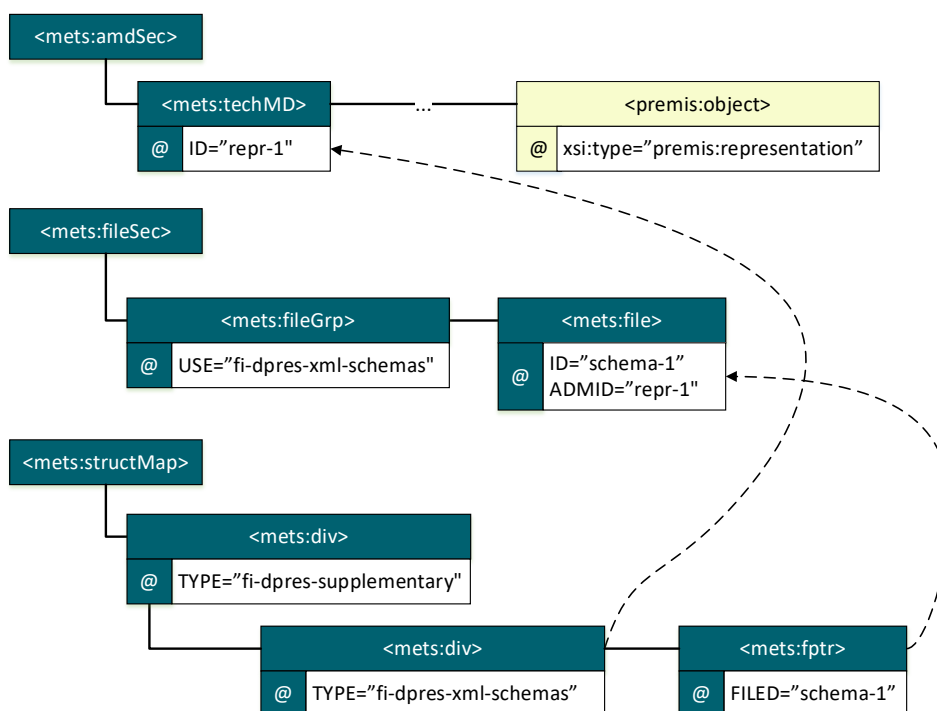
Esimerkki

```

<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>

```

XML-skeematiedosto <fileSec>-lohkossa ja rakennekartassa on havainnollistettu kuvassa 5.



Kuva 5: XML-skeematiedosto osana siirtopakettia

LÄHTEET

- [ADDML] Archival Data Description Markup Language. The National Archives of Norway. <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. <https://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. August 1991. <https://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 RIFF File Reference. Microsoft Corporation. <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. European Broadcasting Union (EBU). May 2011. <https://tech.ebu.ch/docs/tech/tech3285.pdf>
- [DNG_11] Digital Negative (DNG) Specification, Version 1.1.0.0. Adobe Systems Incorporated. February 2005. https://web.archive.org/web/20051108165239/http://www.adobe.com/products/dng/pdfs/dng_spec.pdf
- [DNG_12] Digital Negative (DNG) Specification, Version 1.2.0.0. Adobe Systems Incorporated. April 2008. http://web.archive.org/web/20090521113533/http://www.adobe.com/products/dng/pdfs/dng_spec_1_2_0_0.pdf
- [DNG_13] Digital Negative (DNG) Specification, Version 1.3.0.0. Adobe Systems Incorporated. June 2009. https://web.archive.org/web/20210803212020/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://web.archive.org/web/20220130173101/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://web.archive.org/web/20220122234044/https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.5.0.0.pdf
- [ECMA_376] ECMA International. Office Open XML file formats, 5th edition. ECMA-376. December 2021. <https://ecma-international.org/publications-and-standards/standards/ecma-376/>
- [EPS] Encapsulated PostScript File Format Specification. Adobe Developer Support. Version 3.0. Adobe Systems Incorporated. May 1992. https://web.archive.org/web/20170818010030/http://www.images.adobe.com/content/dam/Adobe/en/devnet/postscript/pdfs/5002.EPSF_Spec.pdf
- [EPUB_2.0.1] EPUB version 2.0.1. International Digital Publishing Forum. <http://idpf.org/epub/201>
- [EPUB_3] EPUB 3.3. W3C Recommendation. October 2024. <https://www.w3.org/TR/epub-33/>
- [FI_API] PAS-palveluiden rajapinnat. Kansalliset pitkäaikaissäilytyspalvelut. <urn:nbn:fi-fe2020100578097>
- [FI_META] Aineistojen ja niiden metatietojen paketointi pitkäaikaissäilytykseen. Kansalliset pitkäaikaissäilytyspalvelut. <urn:nbn:fi-fe2020100578093>
- [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. June 1987.
<https://www.w3.org/Graphics/GIF/spec-gif87.txt>
- [GIF_89a] GRAPHICS INTERCHANGE FORMAT(tm), Version 89a. CompuServe Incorporated. July 1990.
<https://www.w3.org/Graphics/GIF/spec-gif89a.txt>
- [HDF5] HDF5 API Reference: Specifications. The HDF Group.
https://support.hdfgroup.org/documentation/hdf5/latest/_spec.html
- [HTML_4.01] HTML 4.01 Specification. W3C Recommendation. December 1999.
<https://www.w3.org/TR/html401/>
- [HTML5] HTML. Living Standard. Web Hypertext Application Technology Working Group (WHATWG)
<https://html.spec.whatwg.org>
- [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
- [IFC_4.0.2.1] Industry Foundation Classes (IFC) 4.0.2.1. buildingSMART International Limited.
https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2_TC1/HTML/
- [IFC_4.3.2.0] Industry Foundation Classes (IFC) 4.3.2.0. buildingSMART International Limited.
https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/
- [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_16739:2018] International Organization for Standardization. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema. ISO 16739-1:2018
- [ISO_16739:2024] International Organization for Standardization. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema. ISO 16739-1:2024
- [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_23008-2] International Organization for Standardization. Information technology — High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding. ISO/IEC 23008-2
- [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_23736] International Organization for Standardization. Information technology — Digital publishing — EPUB 3.0.1. ISO/IEC 23736:2020 (parts 1-6)
- [ISO_28500:2009] International Organization for Standardization. Information and documentation — WARC file format. ISO 28500:2009
- [ISO_28500:2017] International Organization for Standardization. Information and documentation — WARC file format. ISO 28500:2017
- [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_30135] International Organization for Standardization. Information technology — Digital publishing — EPUB3. ISO/IEC TS 30135:2014 (parts 1-7)
- [ISO_32000-2] International Organization for Standardization. Document management — Portable document format — Part 2: PDF 2.0. ISO 32000-2:2020
- [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] OGC KML 2.3. Open Geospatial Consortium. August 2015. <http://www.opengis.net/doc/IS/kml/2.3>
- [LAC] Trusted Digital Repository, Local Digital Format Registry (LDFR), File Format Guidelines for Preservation and Long-term Access, Version 1.0. Library and Archives Canada.
- [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. March 2001. <https://developer.apple.com/standards/qtff-2001.pdf>
-

[MS_DOC]	[MS-DOC]: Word (.doc) Binary File Format. Microsoft Corporation. November 2024. https://msopenspecs.azureedge.net/files/MS-DOC/%5bMS-DOC%5d.pdf
[MS_PPT]	[MS-PPT]: PowerPoint (.ppt) Binary File Format. Microsoft Corporation. November 2024. https://msopenspecs.azureedge.net/files/MS-PPT/%5bMS-PPT%5d.pdf
[MS_XLS]	[MS-XLS]: Excel Binary File Format (.xls) Structure. Microsoft Corporation. August 2024. https://msopenspecs.azureedge.net/files/MS-XLS/%5bMS-XLS%5d.pdf
[ODF_1.0]	Open Document Format for Office Applications (OpenDocument) v1.0. OASIS Standard. May 2005. https://docs.oasis-open.org/office/v1.0/OpenDocument-v1.0-os.pdf
[ODF_1.1]	Open Document Format for Office Applications (OpenDocument) v1.1. OASIS Standard incorporating Approved Errata 01. July 2013. http://docs.oasis-open.org/office/v1.1/OpenDocument-v1.1.html
[ODF_1.2]	Open Document Format for Office Applications (OpenDocument) Version 1.2. OASIS Standard. September 2011. https://docs.oasis-open.org/office/v1.2/OpenDocument-v1.2.pdf
[ODF_1.3]	Open Document Format for Office Applications (OpenDocument) Version 1.3. OASIS Standard. April 2021. https://www.oasis-open.org/2021/06/16/opendocument-v1-3-oasis-standard-published/
[OGC_1.3.0]	Open Geospatial Consortium. OGC GeoPackage Encoding Standard. Version 1.3.0. 2021. http://www.geopackage.org/spec130/
[OGC_1.3.1]	Open Geospatial Consortium. OGC GeoPackage Encoding Standard. Version 1.3.1. November 2021. https://www.geopackage.org/spec131/
[PNG]	Portable Network Graphics (PNG) Specification (Third Edition). W3C Candidate Recommendation Draft. July 2024. https://www.w3.org/TR/png/
[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 technical registry PRONOM. The National Archives, UK. https://www.nationalarchives.gov.uk/aboutapps/pronom/
[RFC_2046]	Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types. Request for Comments: 2046. Network Working Group. November 1996. https://datatracker.ietf.org/doc/rfc2046/
[RFC_2119]	Key words for use in RFCs to Indicate Requirement Levels. Request for Comments: 2119. Network Working Group. March 1997. https://datatracker.ietf.org/doc/rfc2119/
[RFC_3190]	RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio. Request for Comments: 3190. Network Working Group. January 2002. https://datatracker.ietf.org/doc/rfc3190/
[RFC_3551]	RTP Profile for Audio and Video Conferences with Minimal Control. Request for Comments: 3551. Network Working Group. July 2003. https://datatracker.ietf.org/doc/rfc3551/
[RFC_4180]	Common Format and MIME Type for Comma-Separated Values (CSV) Files. Request for Comments: 4180. Network Working Group. October 2005. https://datatracker.ietf.org/doc/rfc4180/
[RFC_9043]	FFV1 Video Coding Format Version 0, 1, and 3. Request for Comments: 9043. Internet Engineering Task Force (IETF). August 2021. https://datatracker.ietf.org/doc/rfc9043/
[RFC_9559]	Matroska Media Container Format Specification. Request for Comments: 9559. Internet Engineering Task Force (IETF). October 2024. https://datatracker.ietf.org/doc/rfc9559/
[RFC_9649]	WebP Image Format. Request for Comments: 9559. Internet Engineering Task Force (IETF). November 2024. https://datatracker.ietf.org/doc/rfc9649/
[RFC_9639]	Free Lossless Audio Codec. Request for Comments: 9639. December 2024. Internet Engineering Task Force (IETF). https://datatracker.ietf.org/doc/rfc9639/
[SIARD_2.1.1]	SIARD Format Specification 2.1.1. May 2019. https://github.com/DILCISBoard/SIARD/tree/master/SIARD%202.1.1/format/2019-05-15

[SIARD_2.2]	SIARD Format Specification 2.2. August 2021. https://siard.dilcis.eu/SIARD_2-2/format/2021-08-31/SIARD_2-2.pdf
[SMPTE_268]	Society of Motion Picture and Television Engineers. For File Format for Digital Moving-Picture Exchange (DPX), Version 2.0. SMPTE ST 268-1:2014
[SMPTE_314]	Society of Motion Pictures & Television Engineers. Standard for Television. Data Structure for DV-Based Audio, Data and Compressed Video 625 and 50 Mb/s. SMPTE ST 314:2005
[SMPTE_370]	Society of Motion Pictures & Television Engineers. Standard for Television. Data Structure for DV-Based Audio, Data and Compressed Video at 100 Mb/s 1080/60i, 1080/50i, 720/60p, 720/50p. SMPTE ST 370:2013
[SMPTE_377]	Society of Motion Pictures & Television Engineers. Material Exchange Format (MXF) - File Format Specification. SMPTE ST 377-1:2019
[SMPTE_421]	Society of Motion Pictures & Television Engineers. Standard for Television - VC-1 Compressed Video Bitstream Format and Decoding Process. SMPTE ST 421:2013
[SPSS]	SPSS Portable File, ASCII encoding. Library of Congress. https://www.loc.gov/preservation/digital/formats/fdd/fdd000468.shtml
[SVG]	Scalable Vector Graphics (SVG) 1.1 specification (Second Edition). W3C Recommendation. August 2011. https://www.w3.org/TR/SVG11/
[TIFF]	TIFF Revision 6.0 Final. Adobe Developers Association. June 1992. https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf
[UNICODE]	The Unicode Standard, Version 16.0.0. The Unicode Consortium. September 2024. https://www.unicode.org/versions/Unicode16.0.0/
[VIDEOMD]	Video Technical Metadata Extension Schema. https://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html
[WARC_1.0]	The WARC Format 1.0. International Internet Preservation Consortium. https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.0/
[WARC_1.1]	The WARC Format 1.1. International Internet Preservation Consortium. https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/
[WAV]	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
[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/