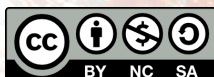


SÄILYTYS- JA SIIRTOKELPOISET TIEDOSTOMUODOT

v1.8.0

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



Lisenssi

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

Tämän määrittymisen 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	Säilytys- ja siirtokelpoisten tiedostomuotojen määrittely ja määrittelyn ylläpitäminen	6
1.4	Tiedostomuotojen tekniset piirteet	6
1.5	Tiedostomuotokirjasto	7
2	SÄILYTYSKELPOISET 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)	10
2.1.4	Extensible Markup Language (XML)	10
2.1.5	Hypertext Markup Language (HTML)	10
2.1.6	Open Document Format (ODF)	11
2.1.7	PDF for long-term preservation: PDF-Archive (PDF/A)	11
2.1.8	Tekstitiedosto (Plain text)	12
2.2	Ääni	12
2.2.1	Audio Interchange File Format (AIFF), LPCM-koodattu	12
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)	13
2.2.5	MPEG-4 AAC – Advanced Audio Coding (AAC)	14
2.2.6	Waveform Audio Format (WAV)	14
2.3	Elävä kuva	14
2.3.1	Digital Picture Exchange (DPX)	14
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)	15
2.4	Kuva	16
2.4.1	Digital Negative (DNG)	16
2.4.2	Joint Photographic Experts Group (JPEG)	16
2.4.3	Joint Photographic Experts Group JPEG 2000 (JP2)	17
2.4.4	Portable network graphics (PNG)	17
2.4.5	Scalable Vector Graphics (SVG)	17
2.4.6	Tagged Image File Format (TIFF)	17
2.5	Verkkoarkisto	18
2.5.1	Web ARChive Format (WARC)	18
2.6	Paikkatietoaineistot	19
2.6.1	Geographic Tagged Image File Format (GeoTiff)	19
2.6.2	Geography Markup Language (GML)	19
2.6.3	Keyhole Markup Language (KML)	19
2.7	Tietokannat	20
2.7.1	Software Independent Archiving of Relational Databases (SIARD)	20
2.8	Tilastoaineistot	20
2.8.1	SPSS Portable (POR)	20
2.9	Tieteelliset ohjelmistot	20
2.9.1	Matlab	20
2.9.2	Hierarchical Data Format (HDF5)	21
3	SIIRTOKELPOISET TIEDOSTOMUODOT	22
3.1	Teksti	22
3.1.1	Microsoft Office Suite	22

3.1.2	Portable Document Format (PDF)	23
3.2	Ääni	23
3.2.1	Audio Interchange File Format (AIFF-C)	23
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).....	24
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)	25
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)	26
4	SÄILIÖMUODOT	27
5	TEKNISET METATIEDOT	30
5.1	ADDML	30
5.1.1	Skeeman soveltaminen	30
5.1.2	Esimerkki	30
5.2	AudioMD	31
5.2.1	Skeeman soveltaminen	31
5.2.2	Esimerkki	32
5.3	VideoMD	33
5.3.1	Skeeman soveltaminen	33
5.3.2	Esimerkki	35
5.4	MIX	36
5.4.1	Skeeman soveltaminen	36
5.4.2	Esimerkki	36
6	KONTROLLOITU SANASTO	38
6.1	Tekstitiedoston sallitut merkistöt	41
6.2	Tiedostomuodon tarkentaminen	41
LIITE A.	TIEDOSTOMUODOISTA	42
LIITE B.	TIEDOSTOMUOTOJEN PYSYVYYDEN ARVIOINTI	43
B.1.	Arviointikriteerit.....	43
B.2.	Yhteenvedo säilytyskelpoisten tiedostomuotojen arvioinnista.....	46
LIITE C.	MUIDEN KUIN SÄILYTYS- JA SIIRTOKELPOISTEN TIEDOSTOMUOTOJEN KÄSITTELY PAS-PALVELUSSA	47
C.1.	Vaatimukset ja suositukset	47
C.2.	Käytännön ohjeet.....	48
LÄHTEET	51

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. Määrittelyssä on hyödynnetty Kanadan kansalliskirjaston ja -arkiston selvityksiä [LAC] ja muita kansainvälisiä selvityksiä.

1.1 PAS-palvelut

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

Kulttuuriperintöaineistojen pitkäaikaissäilyttämiseen tuotettu PAS-palvelu (Kulttuuriperintö-PAS-palvelu) takaa kirjastojen, arkistojen ja museoiden keskeisten kansallisten digitalisten tietovarantojen pitkäaikaissäilyttämisen. Digitaalisilla kulttuuriperintöaineistoilla tarkoitetaan sekä digitoituja että digitaaliseen muotoon tuotettuja kulttuuriperintöaineistoja: lakisääteisen säilyttämisen piiriin kuuluvia digitaalisia kulttuuriaineistoja, kansalliseen kulttuuriperintöön kuuluvaa digitaalista asiakirjallista aineistoa sekä aineellisen ja henkisen kulttuuriperinnön säilyttämisestä vastaavien, opetus- ja kulttuuriministeriön hallinnonalalla toimivien organisaatioiden muita pitkäaikaissäilytyksen piiriin kuuluvia digitaalisia tietovarantoja.

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

1.2 Säilytys- ja siirtokelpoiset tiedostomuodot PAS-palveluissa

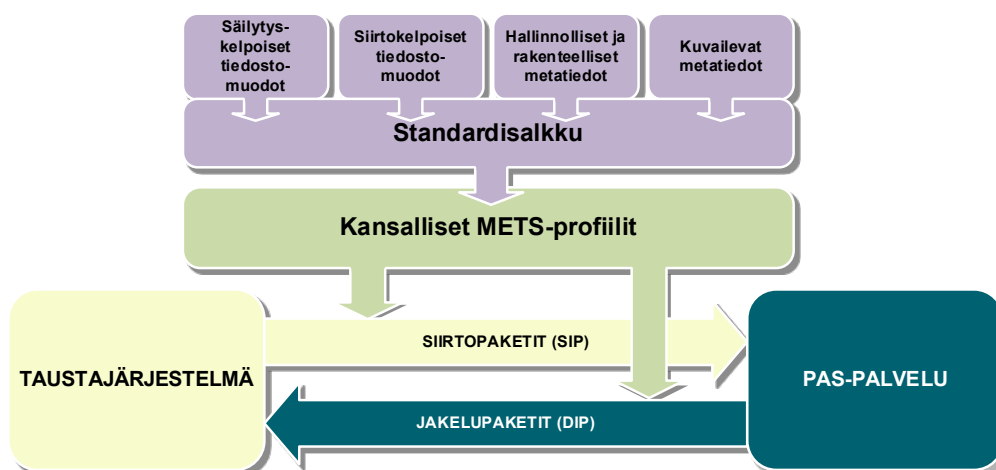
PAS-palveluiden tukemia tiedostomuotoja on kahdenlaisia: *säilytyskelpoisia* ja *siirtokelpoisia*. Säilytyskelpoisten tiedostomuotojen arvioidaan olevan käyttökelpoisia vielä pitkään, kun taas siirtokelpoiset tiedostomuodot ovat arvion mukaan niitä, joissa pitkäaikaissäilytettävää aineistoa on runsaasti tallennettu. Mikäli säilytettävän aineiston tiedostomuoto ei ole säilytys- eikä siirtokelpoinen, tiedostot tulee muuntaa suoraan säilytyskelpoiseen tiedostomuotoon, jos mahdollista. Tällaisista muunnoksista vastaa ensisijaisesti hyödyntävä organisaatio, jolle keskitetty PAS-palvelu voi tarjota muunnoksiin liittyvää tukea (esim. tietoa soveltuvista työkaluista). Ennen muunnoksien tekemistä on suositeltavaa olla yhteydessä PAS-palveluun, jotta tiedostot muunnetaan mahdollisimman tarkoituksen mukaiseen tiedostomuotoon. On tärkeää, että tiedostoa ei muunneta ensin vain siirtokelpoiseen muotoon, koska tällöin jouduttaisiin tekemään vielä uusi, ylimääräinen muunnos 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.

1.3 Säilytys- ja siirtokelpoisten tiedostomuotojen määrittely ja määrittelyn ylläpitäminen

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 tiedostumuotoja, jotka ovat ns. de facto – eli teollisuusstandardeja. Liitteessä B on käsitelty valittujen tiedostomuotojen pysyvyyden arviointia tarkemmin.

Säilytys- ja siirtokelpoisten tiedostomuotojen määrittely on sovittu kuuluvan standardisalkkuun (kuva 1), jonka ylläpidon yhteydessä tulee myös määrittelyä tarkistaa. Määrittely tulee muuttumaan ajan myötä, erityisesti uusien tiedostomuotojen levitessä käyttöön ja nykyisten tiedostomuotojen vanhentuessa.



Kuva 1: Standardisalkun, säilytys- ja siirtokelpoisten tiedostomuotojen ja tietopakettien suhde

Tässä dokumentissa on määritelty säilytys- ja siirtokelpoisia tiedostumuotoja 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. Esimerkiksi PDF, joka on luokiteltu tekstitiedostomuodoksi, voi sisältää kuvan. Tiedoston todellinen sisältö tulee huomioida säilytys suunnitelmassa.

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. Keskimmäisen 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.

1.4 Tiedostomuotojen tekniset piirteet

PAS-palvelu asettaa kaikille säilytettävälle tiedostoille yleisiä rajoituksia, jotka koskevat kaikkia tiedostumuotoja, 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.

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.

- 3) Tiedostoja EI SAA allekirjoittaa digitaalisesti. Digitaalisen allekirjoituksen käyttäminen 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 linkitetyt aineistot, jotka vaikuttavat tiedoston esittämiseen.

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 tämän kaltaisten 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 tiedon alkuperäisen aineiston kuvasuhteesta.
- 3) Metatiedot, jotka halutaan säilyvän ja ylläpidettävän 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.

1.5 Tiedostomuotokirjasto

Tiedostomuotokirjasto (Format registry) on PAS-palvelun osa, joka sisältää tietoa tiedostomuodoista. Yleisimmin tunnettuja tiedostomuotokirjastoja ovat muun muassa PRONOM³ ja Yhdysvaltojen 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

² <http://digitalpreservation.fi/specifications/vocabularies/unknown-values>

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

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

vanhentumassa oleva digitaalinen objekti muunnetaan. Valitun tiedostomuodon yhteydessä viitataan sen tarkempaan kuvaukseen ja käytetään tiedostomuodolle PRONOMin pysyvää tunnistetta (PUID, PRONOM Persistent Unique Identifier) silloin kun se on mahdollista.

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 määrittäksessä 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]; 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 Forumin (IDPF) toimesta ja uusin versio (3.2) W3C:n toimesta. Standardissa julkaisun sisällön rakenne paketoidaan säiliöön ja tekstin muotoilu sekä ulkoasu muodostetaan lukulaitteessa. Tiedostomuoto on ns. rakenteinen nippuformaatti.

Hyväksyttävät versiot:

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

EPUB versio 3.0.0: [EPUB_3.0.0]; PRONOM: fmt/483

EPUB versio 3.0.1: [EPUB_3.0.1]; PRONOM: fmt/483

EPUB versio 3.2: [EPUB_3.2]; 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-sovitus, jonka viimeisin W3C suosittelema versio on XHTML 5.0.

Hyväksyttävät versiot:

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

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

XHTML versio 5.0; [HTML5]; PRONOM: fmt/471

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.

Hyväksyttävät versiot:

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 15445:2000, versiota 5.0 [HTML5].

Hyväksyttävät versiot:

HTML versio 4.01; [ISO_15445]; PRONOM: fmt/100

HTML versio 5.0; [HTML5]; PRONOM: fmt/471

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 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 26300:2006 ja versio 1.2 ISO-standardina ISO/IEC 26300:2015. Teknisesti Open document on ZIP-pakattu arkisto eli ns. yksinkertainen nippuformaatti, joka sisältää useita tiedostoja ja hakemistoja. Dokumentin sisältö, esitystyylit, metatieto ja asetukset on eroteltu omiksi XML-tiedostoikseen.

Yleisimmät ODF-standardia noudattavat tiedostotyypit ovat mm.:

- .odt, tekstinkäsittely
- .ods, taulukkolaskenta
- .odp, esitys
- .odg, piirto
- .odf, kaava

Hyväksyttävät versiot:

.odt versio 1.0; [ISO_26300:2006]; PRONOM: [fmt/136](#)
 .odt versio 1.1; PRONOM: [fmt/290](#)
 .odt versio 1.2; [ISO_26300:2015]; PRONOM: [fmt/291](#)
 .ods versio 1.0; [ISO_26300:2006]; PRONOM: [fmt/137](#)
 .ods versio 1.1; PRONOM: [fmt/294](#)
 .ods versio 1.2; [ISO_26300:2015]; PRONOM: [fmt/295](#)
 .odp versio 1.0; [ISO_26300:2006]; PRONOM: [fmt/138](#)
 .odp versio 1.1; PRONOM: [fmt/292](#)
 .odp versio 1.2; [ISO_26300:2015]; PRONOM: [fmt/293](#)
 .odg versio 1.0; [ISO_26300:2006]; PRONOM: [fmt/139](#)
 .odg versio 1.1; PRONOM: [fmt/296](#)
 .odg versio 1.2; [ISO_26300:2015]; PRONOM: [fmt/297](#)
 .odf versio 1.0; [ISO_26300:2006].
 .odf versio 1.2; [ISO_26300:2015]

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.

2.1.7 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 visuaalisen ulkoasun säilyttämiseksi. PDF/A-tiedostosta on riisuttu ne ominaisuudet, joita ei voida luotettavasti säilyttää, kuten muun muassa skriptit, äänet, elävä kuva, salaus sekä hyperlinkit dokumentin ulkopuolelle. PAS-palvelu suosittelee versioita PDF/A-1a ja PDF/A-1b, jotka ovat standardoitu ISO-standardina ISO 19005:2005, versioita PDF/A-2a, PDF/A-2b ja PDF/A-2u, jotka ovat standardoitu ISO-standardina ISO 19005-2:2011, ja

versioita PDF/A-3a, PDF/A-3b ja PDF/A-3u, jotka ovat standardoitu ISO-standardina ISO 195005-3:2012. 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 opotetuille liitetiedostoille.

Hyväksyttävät versiot:

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

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.

2.1.8 Tekstitiedosto (Plain text)

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

Hyväksyttävät versiot:

[RFC2046]; 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; [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. Säilytuskelpoinen 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 Version 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.

Hyväksyttävät versiot:

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

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Matroska [MATROSKA]

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) [LPCM] on tietyn tyyppinen PCM, jossa kvantisointitasot ovat lineaarisesti yhdenmukaiset ja jota käytetään tyypillisesti säiliömuotojen ääniraidoissa.

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Audio Video Interleave [AVI]

DV (raakamuoto) [IEC_61834, SMPTE_314, SMPTE_370]

Material Exchange Format [MXF]

Motion JPEG 2000 [MJ2]

Matroska [MATROSKA]

Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/L16, audio/L8, 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 14496-3:2009), osana MPEG-2- ja MPEG-4-määrittäjiä. 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 [MXF]

Quicktime MOVie [MOV]

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

audio/mp4

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. Säilytyskelvokseen 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_LOC, WAV_IBM_MS]; 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 [DPX].

Hyväksyttävät versiot:

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

Sanasto PREMIS 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 [FFV1]

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Matroska [MATROSKA]

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 (MJ2) 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 [MXF]

Motion JPEG 2000 [MJ2]

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

MPEG-tiedostomuodot voivat sisältyä säiliömuotoihin:

Material Exchange Format [MXF]
 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/mp4

Pakollisten teknisten metatietojen metatietoskeema:

VideoMD

2.4 Kuva

2.4.1 Digital Negative (DNG)

DNG on Adoben kehittämä valmistajariippumaton ja rojaliton kuvatiedostomuoto, johon digitaalikameran RAW-raakakuvatiedostot saa muunnettua häviöttömästi. DNG on säilytyskelpoinen tiedostomuoto edellyttäen, että siirtopaketissa tulee lisäksi DNG-kuvatiedostoa vastaava TIFF- tai JPEG2000-tiedosto. DNG-tiedostomuodolle on haettu ISO-standardointia.

Hyväksyttävät versiot:

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/tiff

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.4.2 Joint Photographic Experts Group (JPEG)

Joint Photographic Experts Groupin kehittämä JPEG-standardi (ISO/IEC 10918-1:1994) viittaa sekä tiedostomuotoon että pakkausmenetelmään. JPEG-pakkausmenetelmä on standardoitu häviöllinen kuvien pakkausmenetelmä täysväri- (RGB) ja mustavalkokuville.

Hyväksyttävät versiot:

JPEG part 1 versio 1.00; [ISO_10918-1]; PRONOM: fmt/42
 JPEG part 1 versio 1.01; [ISO_10918-1]; PRONOM: fmt/43
 JPEG part 1 versio 1.02; [ISO_10918-1]; PRONOM: fmt/44
 JPEG/EXIF versio 2.0; PRONOM; x-fmt/398
 JPEG/EXIF versio 2.1; PRONOM; x-fmt/390
 JPEG/EXIF versio 2.2; PRONOM; x-fmt/391
 JPEG/EXIF versio 2.2.1; PRONOM; fmt/645
 JPEG/EXIF versio 2.3
 JPEG/EXIF versio 2.3.1
 JPEG/EXIF versio 2.3.2

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- värimaailmoja.

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 15948:2004.

Hyväksyttävät versiot:

PNG; [ISO_15948]; 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 versiot⁵:

SVG 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 32-bittisiä CMYK- tai 24-bittisiä RGB-kuvia.

Hyväksyttävät versiot:

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

⁵ Tiedosto ei saa sisältää animaatioita eikä interaktiivista toiminnallisuutta.

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/tiff

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.5 Verkkoarkisto

2.5.1 Web ARChive Format (WARC)

WARC (Web ARChive) tiedostomuoto määrittelee menetelmän useiden digitaalisten resurssien yhdistämiseksi yhdeksi arkistointitiedostoksi. ISO-standardoitu (ISO 28500:2009) WARC on The Internet Archiven kehittämän ARC-tiedostomuodon kehittyneempi versio. Se tukee ARC-tiedostomuotoa paremmin verkkoaineistojen haravointiin, jakeluun ja tiedonvaihtoon liittyviä tarpeita.

Hyväksyttävät versiot:

WARC versio 0.17⁶ PRONOM: fmt/289

WARC versio 1.0; [ISO_28500]; PRONOM: fmt/289

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.

2.5.1.1 Internet Archive (ARC)

Kuten WARC, myös ARC-tiedostomuoto määrittelee menetelmän useiden digitaalisten resurssien yhdistämiseksi yhdeksi arkistointitiedostoksi. ARC-muoto on säilytyskelppoinen tiedostomuoto vain erikoistapauksissa, eikä sitä tule ottaa käyttöön uusiin paketoiteihin. Uusiin paketoiteihin tulee käyttää WARC-tiedostomuotoa.

Hyväksyttävät versiot:

ARC versio 1.0; [ARC]; PRONOM: x-fmt/219

ARC versio 1.1; PRONOM: fmt/410

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

application/x-internet-archive

Pakollisten teknisten metatietojen metatietoskeema:

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

⁶ WARC versio 0.17 on säilytyskelppoinen tiedostomuoto vain erikoistapauksissa, eikä sitä tule ottaa käyttöön uusiin paketoiteihin.

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 1.0 [GEOTIFF]; PRONOM: fmt/155

Sanasto PREMIS formatName semanttisessa rakenneyksikössä:

image/tiff

Pakollisten teknisten metatietojen metatietoskeema:

MIX

2.6.2 Geography Markup Language (GML)

Geography Markup Language (GML) on XML-pohjainen merkintäkieli erilaisten paikkatietoa sisältävien piirteiden esittämiseen. GML on standardoitu ISO-standardina ISO 19136:2007 [ISO_19136].

Hyväksyttävät versiot:

GML 3.2.1 [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 Keyhole Markup Language (KML)

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

Hyväksyttävät versiot:

KML 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 toimittaja riippumattomassa muodossa. Tiedostomuoto perustuu XML:ään ja SQL-kieleen, ja sen on kehittänyt Sveitsin liittovaltion arkisto (Swiss Federal Archives).

Hyväksyttävät versiot:

SIARD 2.0 [SIARD2.0]; PRONOM: fmt/995
 SIARD 2.1 [SIARD2.1]; PRONOM: fmt/1196

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 tilastomatemaattisilla ohjelmilla. POR-tekstitiedostossa SPSS-ohjelmasta viety data tallennetaan ASCII-muodossa.

Hyväksyttävät versiot:

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

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 (HDF5)

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 1.10 [HDF5]; PRONOM⁷: 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.

⁷ Eri PRONOM tunnisteet viittaavaan tiedostomuodon Superblocking eri versioihin. Versiolle 3 ei vielä ole PRONOM tunnistetta.

3 SIIRTOKELPOISET TIEDOSTOMUODOT

Tässä luvussa määritellään PAS-palveluun siirrettävissä tietopaketeissa hyväksyttävät tiedostomuodot säilytuskelpoisten tiedostomuotojen lisäksi. Siirtokelpoisiksi hyväksytään vain sellaisia tiedostomuotoja, joita käytetään useassa PAS-palvelua hyödyntävässä organisaatiossa ja joissa pitkäaikais säilytettävää aineistoa on runsaasti tallennettu. PAS-palvelu voi muuntaa säilytysuunnitelman asettamien vaatimusten ja ehtojen mukaisesti siirtokelpoisessa muodossa vastaanotetut tiedostot säilytuskelpoiseen tiedostomuotoon ennen säilyttämisen aloittamista. Jos hyödyntävä organisaatio muuntaa aineistonsa tiedostomuotoja, tulee muunnokset aina suorittaa säilytuskelpoisiin 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 määrittäksessä 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 29500-1:2012). Microsoft Office Suite -ohjelmistosta (versiosta 97 ylöspäin) siirtokelpoisia tiedostomuotoja ovat:

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

Hyväksyttävät versiot:

Microsoft Word 97 (8.0); PRONOM: fmt/40
 Microsoft Word 98 (8.5); PRONOM: fmt/40
 Microsoft Word 2000 (9.0); PRONOM: fmt/40
 Microsoft Word 2002 (10.0); PRONOM: fmt/40
 Microsoft Word 2003 (11.0); PRONOM: fmt/40
 Microsoft Word 2007 (12.0); [ISO_29500-1]; PRONOM: fmt/412
 Microsoft Word 2010 (14.0); [ISO_29500-1]; PRONOM: fmt/412
 Microsoft Word 2013 (15.0); [ISO_29500-1]; PRONOM: fmt/412
 Microsoft Word 2016 (16.0); [ISO_29500-1]; PRONOM: fmt/412
 Microsoft Excel 97 (8.0); PRONOM: fmt/61
 Microsoft Excel 99 (9.0); PRONOM: fmt/62
 Microsoft Excel 2001 (10.0); PRONOM: fmt/62
 Microsoft Excel 2003 (11.0); PRONOM: fmt/62
 Microsoft Excel 2007 (12.0); [ISO_29500-1]; PRONOM: fmt/214
 Microsoft Excel 2010 (14.0); [ISO_29500-1]; PRONOM: fmt/214
 Microsoft Excel 2013 (15.0); [ISO_29500-1]; PRONOM: fmt/214
 Microsoft Excel 2016 (16.0); [ISO_29500-1]; PRONOM: fmt/214
 Microsoft Powerpoint 1997 (8.0); PRONOM: fmt/126
 Microsoft Powerpoint 1999 (9.0); PRONOM: fmt/126
 Microsoft Powerpoint 2001 (10.0); PRONOM: fmt/126
 Microsoft Powerpoint 2003 (11.0); PRONOM: fmt/126
 Microsoft Powerpoint 2007 (12.0); [ISO_29500-1]; PRONOM: fmt/215
 Microsoft Powerpoint 2010 (14.0); [ISO_29500-1]; PRONOM: fmt/215

Microsoft Powerpoint 2013 (15.0); [ISO_29500-1]; PRONOM: fmt/215
Microsoft Powerpoint 2016 (16.0); [ISO_29500-1]; PRONOM: fmt/215

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

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

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 1.7 on standardoitu ISO-standardina ISO 32000-1:2008.

Hyväksyttävät versiot:

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

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 säilytyskelpoinen muoto (ks. luku 2.2.1), mutta PAS-palvelu hyväksyy pakatun AIFF-tiedoston siirtokelepoisena 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 soittimissa. MP3:n pakkausmenetelmä hävittää äänestä ihmisen kuuloalueen ulkopuoliset taajuudet ja pakkaa tiedoston. MPEG-1 layer-3 on standardoitu ISO-standardina ISO/IEC 11172-3:1993. MPEG-2 layer-3 on standardoitu ISO-standardina ISO/IEC 13818-3:1995.

Hyväksyttävät versiot:

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

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

Tämä tiedostomuoto voi sisältyä säiliömuotoihin:

Audio Video Interleave [AVI]

Material Exchange Format [MXF]

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 määrittästä ei ole julkaistu.

Hyväksyttävät versiot:

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

MPEG-tiedostomuodot voivat sisältyä säiliömuotoihin:

Audio Video Interleave [AVI]
 Material Exchange Format [MXF]
 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 kuvalle. WMV 9 on standardoitu Society of Motion Picture and Television Engineers (SMPTE)-standardina SMPTE 421M, joka tunnetaan paremmin tunnisteena VC-1.

Hyväksyttävät versiot:

WMV 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 Adoben kehittämä PostScript tiedostomuoto 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 1987a; [GIF_87a]; PRONOM: fmt/3

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

Sanasto PREMISin formatName semanttisessa rakenneyksikössä:

image/gif

Pakollisten teknisten metatietojen metatietoskeema:

MIX

4 SÄILIÖMUODOT

Säiliömuotoiset digitaaliset objektit sisältävät osia, jotka ovat kuin digitaalisia objekteja, mutta jotka eivät ole itsenäisiä tiedostoja. Esimerkiksi elokuvatiedosto (säiliömuotoinen digitaalinen objekti) voi sisältää yhden tai useamman kuva- tai ääniraidan (digitaalisen objektin osan). Tällöin voidaan tarvita toisaalta digitaalisen objektin metatietoja ja toisaalta 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 <http://digitalpreservation.fi/specifications/examples>.

Jokaiselle digitaalisen objektin osalle esitetään oma PREMIS-muotoinen metatieto-objekti omiin METS:n techMD-lohkoihinsa. Sen tyyppin PITÄÄ olla "bitstream", ja sen PITÄÄ sisältää digitaalisen objektin osan tunnisteen, 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/mp4</premis:formatName>
            </premis:formatDesignation>
            <premis:formatRegistry>
              <premis:formatRegistryName>
                PRONOM
              </premis:formatRegistryName>
              <premis:formatRegistryKey>
                fmt/199
              </premis:formatRegistryKey>
            </premis:formatRegistry>
          </premis:format>
        </premis:objectCharacteristics>
      </premis:object>
    </mets:xmlData>
  </mets:mdWrap>
</mets:techMD>
```

Digitaalisen objektin osan muut hallinnolliset ja kuvailevat metatiedot esitetään määritysten 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:

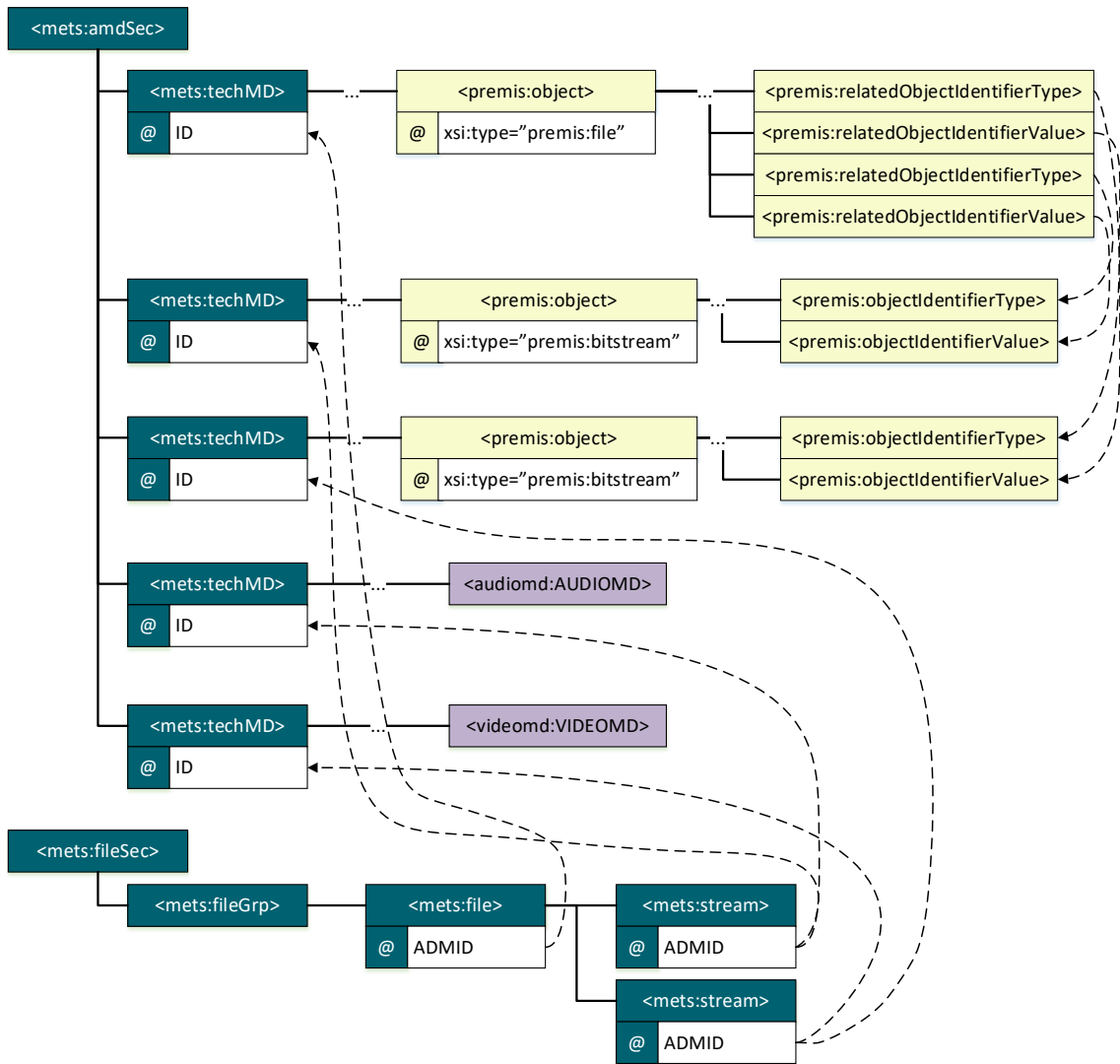
```
<premis:relationship>
  <premis:relationshipType>structural</premis:relationshipType>
  <premis:relationshipSubType>includes</premis:relationshipSubType>
  <premis:relatedObjectIdentification>
    <premis:relatedObjectIdentifierType>UUID
    </premis:relatedObjectIdentifierType>
    <premis:relatedObjectIdentifierValue>
      387e9e6c-4596-4baf-b316-49cc6aflcba2
    </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 2 mukaisesti.



Kuva 2: 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="http://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>
```

```

        </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 käytetty 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 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.

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) tai (:unap), jos tietoa ei voida kohtuullisella vaivalla selvittää. Tuntematon arvona (:unap) vain pakkaamattoman äänen osalta.

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="http://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:compression>
  </amd:fileData>
</amd:AUDIOMD>
```



```

    <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 käytetty elementin tai attribuutin esiintymät ja veloitteet seuraavin lyhentein:

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

Taulukossa 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.

Elementti/ Attribuutti	Esiintymä/ veloite	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.
--<color>	T, P	Vapaaehtoinen videoMD-skeemassa, mutta pakollinen PAS-palvelussa

Elementti/ Attribuutti	Esiintymä/ velvoite	Säännöt ja suositukset
--<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="http://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 käytetty 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 on lueteltu vain PAS-palvelussa vaadittavat 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 mukaisesti⁸.

Elementti/ Attribuutti	Esiintymä/ velvoite	Säännöt ja suositukset
<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.
<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="http://www.loc.gov/mix/v20">
  <mix:BasicDigitalObjectInformation>
    <mix:Compression>
      <mix:compressionScheme>JPEG 2000 Lossless</mix:compressionScheme>
    </mix:Compression>
  </mix:BasicDigitalObjectInformation>
  <mix:BasicImageInformation>
```

⁸ 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.

```
<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

Seuraavassa 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 1: Tiedostomuodot

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
CSV	text/csv		x-fmt/18
EPUB	application/epub+zip	2.0.1	fmt/483
		3.0.0	fmt/483
		3.0.1	fmt/483
		3.2	fmt/483
XHTML	application/xhtml+xml	1.0	fmt/102
		1.1	fmt/103
		5.0	fmt/471
XML	text/xml	1.0	fmt/101
HTML	text/html	4.01	fmt/100
		5.0	fmt/471
ODT	application/vnd.oasis.opendocument.text	1.0	fmt/136
		1.1	fmt/290
		1.2	fmt/291
ODS	application/vnd.oasis.opendocument.spreadsheet	1.0	fmt/137
		1.1	fmt/294
		1.2	fmt/295
ODP	application/vnd.oasis.opendocument.presentation	1.0	fmt/138
		1.1	fmt/292
		1.2	fmt/293
ODG	application/vnd.oasis.opendocument.graphics	1.0	fmt/139
		1.1	fmt/296
		1.2	fmt/297
ODF	application/vnd.oasis.opendocument.formula	1.0	
		1.2	
PDF/A	application/pdf	A-1a	fmt/95
		A-1b	fmt/354
		A-2a	fmt/476
		A-2b	fmt/477
		A-2u	fmt/478
		A-3a	fmt/479
		A-3b	fmt/480
		A-3u	fmt/481
Plain text	text/plain		x-fmt/111
AIFF	audio/x-aiff	1.3	x-fmt/135
BWF	audio/x-wav	2	fmt/527
FLAC	audio/flac	1.2.1	fmt/279
LPCM	audio/L16, audio/L8, audio/L20, audio/L24		
AAC	audio/mp4		fmt/199
WAV	audio/x-wav		fmt/141
DPX	image/x-dpx	2.0	fmt/541
FFV1	video/x-ffv	3	

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
JPEG2000	video/jpeg2000		x-fmt/392
AVC	video/mp4		fmt/199
DNG	image/tiff	1.3 1.4	fmt/438 fmt/730
JPEG JPEG/EXIF	image/jpeg	1.00 1.01 1.02 2.0 2.1 2.2 2.2.1	fmt/42 fmt/43 fmt/44 x-fmt/398 x-fmt/390 x-fmt/391 fmt/645
JP2	image/jp2		x-fmt/392
SVG	image/svg+xml	1.1	fmt/92
TIFF	image/tiff	6.0	fmt/353
PNG	image/png	1.2	fmt/13
ARC ⁹	application/x-internet-archive	1.0 1.1	x-fmt/219 fmt/410
WARC	application/warc	0.17 ¹⁰ 1.0	fmt/289 fmt/289
GEOTIFF	image/tiff	1.0	fmt/155
GML	application/gml+xml	3.2.1	fmt/1047
KML	application/vnd.google-earth.kml+xml	2.3	fmt/244
SIARD	application/x-siard	2.0 2.1	fmt/995 fmt/1196
POR	application/x-spss-por		fmt/997
Matlab	application/matlab	7 7.3	fmt/806 fmt/828
HDF5	application/x-hdf5	1.10	fmt/807 fmt/286 fmt/287
MS Word	application/msword application/vnd.openxmlformats-officedocument.wordprocessingml.document	8.0 8.5 9.0 10.0 11.0 12.0 14.0 15.0 16.0	fmt/40 fmt/40 fmt/40 fmt/40 fmt/40 fmt/412 fmt/412 fmt/412 fmt/412
MS Excel	application/vnd.ms-excel application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	8.0 9.0 10.0 11.0 12.0 14.0 15.0 16.0	fmt/61 fmt/62 fmt/62 fmt/62 fmt/214 fmt/214 fmt/214 fmt/215

⁹ Tiedostomuoto on sallittu vain erikoistapauksissa.

¹⁰ Tiedostomuodon versio sallittu vain erikoistapauksissa.

Tiedostomuoto	formatName	formatVersion	formatRegistryKey
MS Powerpoint	application/vnd.ms-powerpoint	8.0	fmt/126
		9.0	fmt/126
		10.0	fmt/126
		11.0	fmt/126
	application/vnd.openxmlformats-officedocument.presentationml.presentation	12.0	fmt/215
		14.0	fmt/215
		15.0	fmt/215
		16.0	fmt/215
PDF	application/pdf	1.2	fmt/16
		1.3	fmt/17
		1.4	fmt/18
		1.5	fmt/19
		1.6	fmt/20
		1.7	fmt/276
AIFF-C	audio/x-aiff		x-fmt/136
MP3	audio/mpeg	1	fmt/134
		2	fmt/134
WMA	audio/x-ms-wma	9	fmt/132
DV	video/dv		x-fmt/152
MPEG	video/mpeg	1	fmt/649
		2	fmt/640
WMV	video/x-ms-wmv	9	fmt/133
EPS	application/postscript	3.0	fmt/124
GIF	image/gif	1987a	fmt/3
		1989a	fmt/4

Taulukossa 2 on kuvattu PAS-palvelun hyväksymät säiliömuodot ja mitä ääni- ja kuvaraitojen muotoja saa sisällyttää säiliöihin. Muodot ovat taulukossa eroteltu säilytys- ja siirtokelpoisiin muotoihin.

Taulukko 2: Säiliömuodot

Säiliömuoto	formatName	Säilytuskelpoinen		Siirtokelpoinen		format-RegistryKey
		Ääni	Kuva	Ääni	Kuva	
ASF	video/x-ms-asf			WMA	WMV	fmt/131
AVI	video/avi	LPCM		MP3	DV, MPG/H262	fmt/5
DV	video/dv	LPCM			DV	
Matroska v4 ¹¹	video/x-matroska	FLAC, LPCM	FFV1			fmt/569
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	AAC	AVC		MPG/H262	fmt/585
MPEG-4	video/mp4	AAC	AVC	MP3	MPG/H262	fmt/199
MXF	application/mxf	AAC, LPCM	AVC, JP2	MP3	DV, MPG/H262	
MJPEG2000	video/mj2	LPCM	JP2			fmt/337
MOV	video/quicktime	AAC, LPCM	AVC, JP2	MP3	DV, MPG/H262	x-fmt/384

¹¹ Matroska säiliömuodosta hyväksytään vain versio 4

6.1 Tekstiedoston sallitut merkistöt

Tekstimuotoisessa tiedostossa (CSV (2.1.1), XHTML (2.1.3), XML (2.1.4), HTML (2.1.5), plain text (2.1.8), 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>
```

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 *säiliöformaatit*. 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 säiliöformaatit voivat sisältää vain tässä määrityksessä lueteltuja säilytys- ja siirtokelpoisia tiedostomuotoja. Yhdysvaltojen kongressin kirjasto (Library of Congress) jakaa säiliöformaatit 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 itsensä selittävä 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 – tiedostonnimien 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 paketoita 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 tästä tyyppistä ovat METS ja MPEG-21.

LIITE B. TIEDOSTOMUOTOJEN PYSYVYYDEN ARVIOINTI

Tiedostomuotojen valinta on tärkeää pitkäaikaissäilytyksen onnistumiseksi. 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 yleistyvän 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 Yhdysvaltojen 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?
Standardisuus	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 määritellyt kriteerit sekä niiden arvostelu tarkemmin.

Kriteeri	Arviointiohje	Arvio
1. Avoimuus	Määritykset saatavissa yhdeltä tai useammalta seuraavista: (a) avoimen jäsenyyden järjestö [kuten W3C (World Wide Web Consortium) tai OMG (Object Management Group)] (b) kansainvälinen standardintijärjestö (esim. ISO) (c) tuotannonalan avoimen jäsenyyden järjestö	A
	Määritykset saatavilla vain maksusta	A [€]
	Määritykset mahdollisesti saatavilla useista lähteistä (ei saatu vahvistettua)	B
	Määrityksiä 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. Standardisuus	Tiedostomuotoa sääntelee jonkin seuraavista hallitsema muodollinen prosessi: <ul style="list-style-type: none"> ▪ 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 säilytuskelpoisten tiedostomuotojen arvioinnista

Sisältö	Tiedostomuoto	Avoimuus	Käyttö PAS-standardina	Alas-/ylöspäin yhteensopivuuden taso	Korruptoitumisen sieto	Versiopäivitysten määrä	Riippumattomuus / yhteentoimivuus	Standardisuus
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
	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	A	B	A		A	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
VERKKOARKISTO	Web ARChive Format (WARC)	A	A		B	A	A	A
PAIKKATIETOAINIESTOT	Geographic Tagged Image File Format (GeoTiff)	A	A			A	A	A
	Geography Markup Language (GML)							
	Keyhole Markup Language (KML)							
TIETOKANNAT	Software Independent Archiving of Relational Databases (SIARD)	A ¹²	B	B	C	B	B	A
TUTKIMUSAINIESTOT	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 Määrittäminen on ilmaista 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 bittitason säilyttämistä, jossa huolehditaan ainoastaan aineiston eheydestä ja turvallisesta säilytyksestä. Näissä tapauksissa vastuu ymmärtämisen säilyttämisestä 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 3 on esitetty vaatimukset ja suositukset yllä mainituissa tapauksissa (+=suositeltavaa; *=vaaditaan). Jos tiedosto ei ole säilytys- tai siirtokelpoinen, tiedoston tulee täyttää taulukossa 3 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 3: Vaatimukset ja suositukset

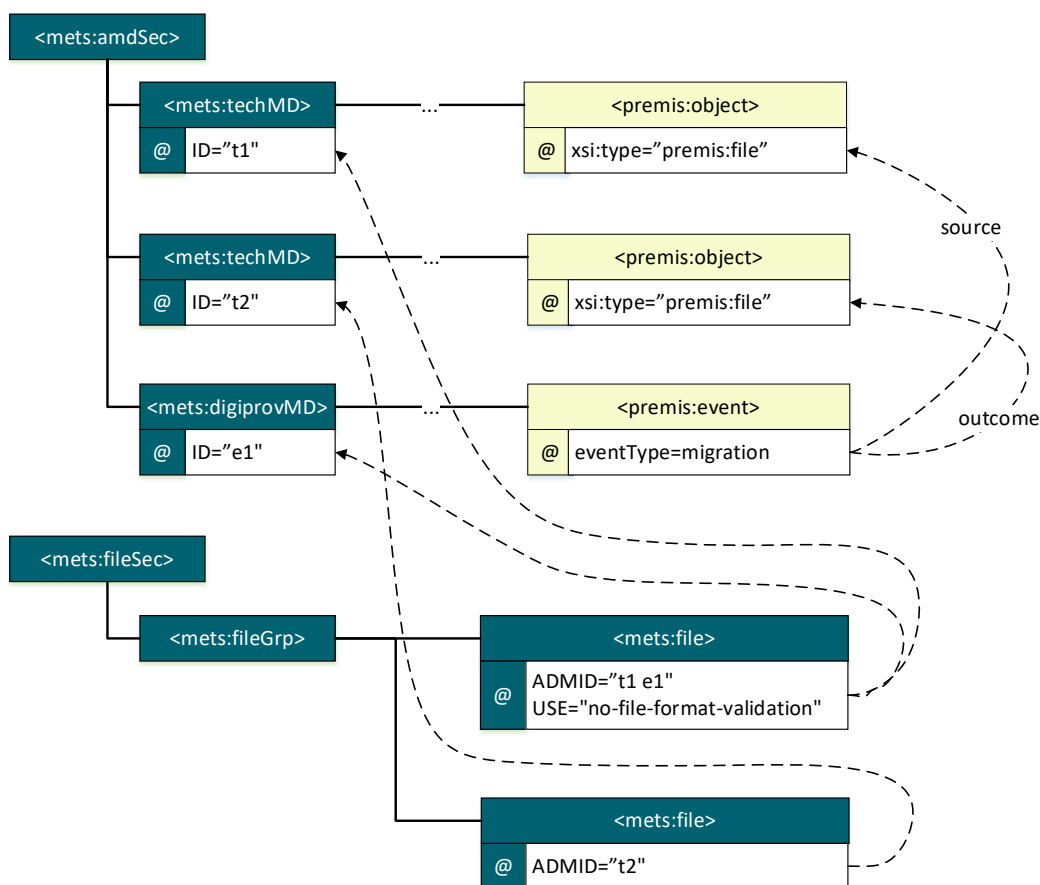
		(1)	(2)	(3)
V1	<p>PAS-palvelu on hyväksynyt tiedostomuodon siirtämisen säilytykseen.</p> <ul style="list-style-type: none"> ▪ Arvioidaan yhdessä PAS-palvelun kanssa, onko tiedostojen säilyttäminen perusteltua vai voitaisiinko ne muuntaa säilytyskelpoiseen muotoon. ▪ Hyväksyntä on aina sopimuskohtaista. 	*	*	*
V2	<p>Tiedostomuoto on tuettu vähintään yhdessä yleisesti saatavilla olevassa ohjelmassa.</p> <ul style="list-style-type: none"> ▪ Ohjelma voi olla maksullinen tai maksuton. Mikäli tiedostojen käsittelyyn tarvitaan erikoisohjelma, tulee ohjelman nimi ja linkki ohjelman kotisivulle mainita. ▪ Mikäli vain hyödyntävän organisaation itse kehittämä ohjelma tukee tiedostomuotoa, ohjelma ja sen lähdekoodi on liitettävä mukaan aineistokokonaisuuteen. 	+	+	*

		(1)	(2)	(3)
V3	Tiedostomuodon rakenne on dokumentoitu. <ul style="list-style-type: none"> 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]. <ul style="list-style-type: none"> 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. <ul style="list-style-type: none"> Standardoidut tiedostomuodot voidaan mahdollisesti hyväksyä myös siirto- tai säilytyskelpoisiksi. Asiasta tulee keskustella PAS-palvelun kanssa. 	+	+	+
V10	Tiedostomuodon eheyden tarkistamiseen on olemassa validaattori. <ul style="list-style-type: none"> 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. <ul style="list-style-type: none"> Skripti tai ohje tulee sisällyttää säilytettävään aineistokokonaisuuteen. 	+	+	+

C.2. Käytännön ohjeet

Jos säilytyskelpoisen tiedoston rinnalle halutaan säilyttää vastaava tiedosto alkuperäisessä tiedostomuodossa, jossa alkuperäinen tiedostomuoto ei ole säilytys- tai siirtokelpoinen (tapaus 1), PITÄÄ nämä tiedostot linkitettävä toisiinsa METS/PREMIS tasolla käyttäen ”migration” -tapahtumaa. Tässä tapahtumassa säilytyskelpoisen tiedostomuoto on tapahtuman ”outcome” ja alkuperäinen tiedostomuoto ”source”, ja alkuperäisen tiedostomuodon METS file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa ”no-file-format-validation”. Tämä on havainnollistettu kuvassa 3¹³.

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

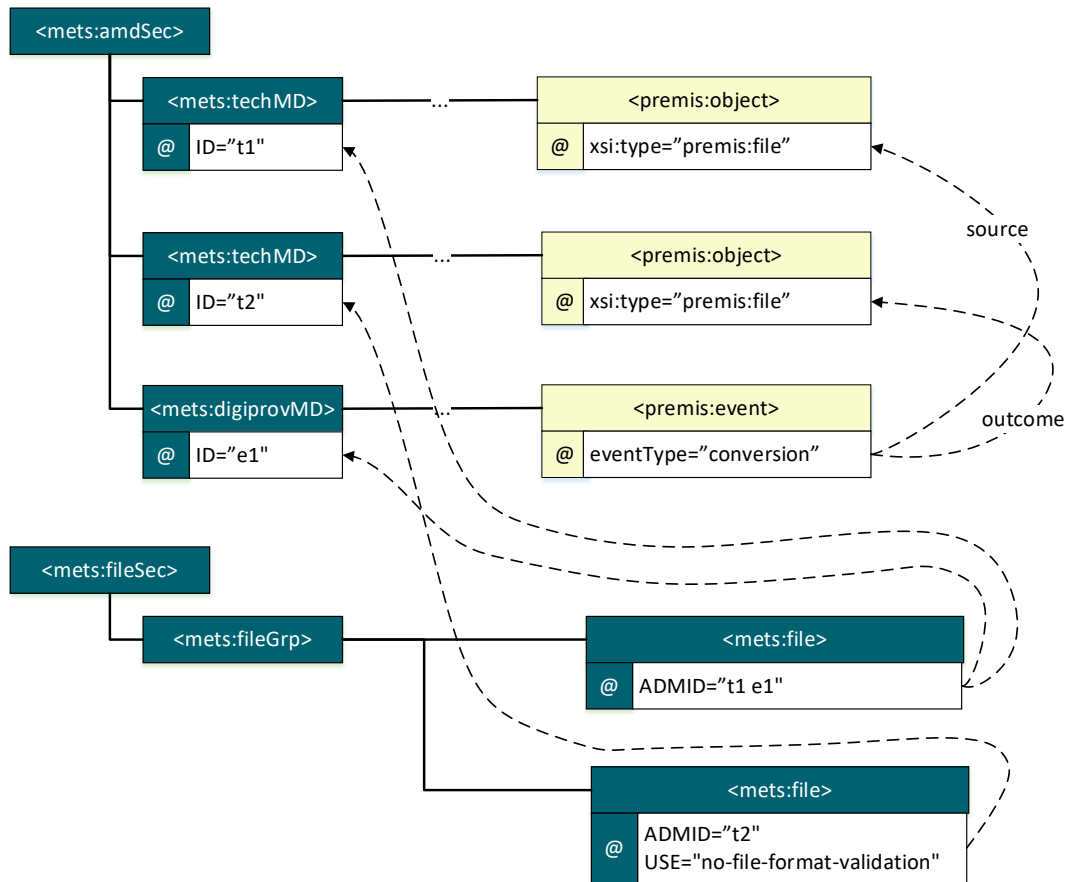


Kuva 3: Säilytyskelpoisen tiedoston rinnalla säilytettävä tiedosto alkuperäisessä tiedostomuodossa

Jos säilytyskelpoisen tiedoston rinnalle halutaan säilyttää tiedosto säilytyskelpoisesta johdetussa versiossa, ei ole säilytys- tai siirtokelpoinen (tapaus2), on nämä tiedostot linkitettävä toisiinsa käyttäen "conversion" -tapahtumaa. Tässä tapahtumassa säilytyskelpoisen tiedosto on tapahtuman "source" ja rinnalle talletettava tiedostomuoto "outcome", ja muunnetun tiedostomuodon METS-dokumentin file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa "no-file-format-validation". Tämä on havainnollistettu kuvassa 4¹⁴.

Jos tiedostomuotoa ei ole mahdollista, tai on hyvin epäkäytännöllistä, muuntaa säilytys- tai siirtokelpoiseen tiedostomuotoon, mutta tiedosto kuitenkin katsotaan tarpeelliseksi säilyttää (tapaus 3), METS-dokumentin file-elementissä attribuutin USE arvona PITÄÄ käyttää arvoa "file-format-identification". Tämä on havainnollistettu kuvassa 5.

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



Kuva 4: Säilytyskelpoisen tiedoston rinnalla säilytettävä tiedosto johdetussa tiedostomuodossa



Kuva 5: Alkuperäistä tiedostoa ei ole mahdollista muuntaa säilytyskelpoiseen tiedostomuotoon

LÄHTEET

- [ADDML] Archival Data Description Markup Language. <https://www.arkivverket.no/forvaltning-og-utvikling/regelverk-og-standarder/andre-arkivstandarder/addml-archival-data-description-markup-language>
- [AIFF] Audio Interchange File Format: "AIFF". A Standard for Sampled Sound Files. Version 1.3. Apple Computer, Inc. <http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf>
- [AIFF-C] Audio Interchange File Format AIFF-C. A revision to include compressed audio data. Apple Computer, Inc. Draft 26.8.1991. <http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-C.9.26.91.pdf>
- [ARC] ARC File Format. Version 1.0. Internet Archive, 15.9.1996. <http://archive.org/web/researcher/ArcFileFormat.php>
- [ASF] Advanced Systems Format (ASF) Specification. Revision 01.20.03. Microsoft Corporation, December 2004. http://download.microsoft.com/download/7/9/0/790feca-f64a-4a5e-a430-0bccdab3f1b4/ASF_Specification.doc
- [AUDIOMD] Audio Technical Metadata Extension Schema <http://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html>
- [AVI] AVI (Audio Video Interleave). Library of Congress. <http://www.digitalpreservation.gov/formats/fdd/fdd000059.shtml>
- [BWF] EBU Tech 3285 - Specification of the Broadcast Wave Format (BWF) - A format for audio data files in broadcasting. Version 2.0. Geneva May 2011. <http://tech.ebu.ch/docs/tech/tech3285.pdf>
- [CSV] Common Format and MIME Type for Comma-Separated Values (CSV) Files. Request for Comments: 4180. October 2005. <https://tools.ietf.org/html/rfc4180>
- [DNG_13] Digital Negative (DNG) Specification, Version 1.3.0.0. Adobe Systems Incorporated. June 2009. http://www.images.adobe.com/content/dam/Adobe/en/products/photoshop/pdfs/dng_spec.pdf
- [DNG_14] Digital Negative (DNG) Specification, Version 1.4.0.0. Adobe Systems Incorporated. June 2012. http://www.adobe.com/content/dam/Adobe/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf
- [DPX] Society of Motion Picture and Television Engineers. ST 268:2003 For File Format for Digital Moving-Picture Exchange (DPX), Version 2.0
- [EPS] Encapsulated PostScript File Format Specification. Adobe Developer Support. Version 3.0. Adobe Systems Incorporated. 1.5.1992. http://partners.adobe.com/public/developer/en/ps/5002.EPSF_Spec.pdf
- [EPUB_2.0.1] EPUB version 2.0.1. International Digital Publishing Forum. <http://idpf.org/epub/201>
- [EPUB_3.0.0] EPUB version 3.0.0. International Digital Publishing Forum. <http://idpf.org/epub/30>
- [EPUB_3.0.1] EPUB version 3.0.1. International Digital Publishing Forum. <http://idpf.org/epub/301>
- [EPUB_3.2] EPUB 3.2. Final Community Group Specification 08 May 2019. <https://www.w3.org/publishing/epub/epub-spec.html>
- [FFV1] FFV1 Video Coding Format Version 0, 1, and 3. <https://tools.ietf.org/html/draft-ietf-cellar-ffv1-06>
- [FLAC] FLAC Format Specification — The formal specification. <http://flac.sourceforge.net/format.html>
- [GEOTIFF] GeoTIFF Format Specification, GeoTIFF Revision 1.0, Specification Version: 1.8.2, November, 1995. <http://gis-lab.info/docs/geotiff-1.8.2.pdf>
- [GIF_87a] GIF(tm). GRAPHICS INTERCHANGE FORMAT(tm). A standard defining a mechanism for the storage and transmission of raster-based graphics information. CompuServe Incorporated. 15.6.1987. <http://www.w3.org/Graphics/GIF/spec-gif87.txt>
- [GIF_89a] GRAPHICS INTERCHANGE FORMAT(tm), Version 89a. CompuServe Incorporated, Columbus, Ohio. 31.7.1990. <http://www.w3.org/Graphics/GIF/spec-gif89a.txt>

- [HDF5] The HDF Group: HDF5 file format home page. <https://support.hdfgroup.org/HDF5/>
- [HTML5] HTML5. A vocabulary and associated APIs for HTML and XHTML. W3C Recommendation 28 October 2014. <https://www.w3.org/TR/html5/>
- [IEC_61834] International Electrotechnical Commission. Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems). IEC 61834:1998-2008
- [ISO_10918-1] International Organization for Standardization. Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines. ISO/IEC 10918-1:1994
- [ISO_11172-1] International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 1: Systems. ISO/IEC 11172-1:1993
- [ISO_11172-2] International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 2: Video. ISO/IEC 11172-2:1993
- [ISO_11172-3] International Organization for Standardization. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio. ISO/IEC 11172-3:1993
- [ISO_13818-1] International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information: Systems. ISO/IEC 13818-1:2000
- [ISO_13818-2] International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information: Video. ISO/IEC 13818-2:1996
- [ISO_13818-3] International Organization for Standardization. Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio. ISO/IEC 13818-3:1995
- [ISO_14496-10] International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding. ISO/IEC 14496-10:2010
- [ISO_14496-14] International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 14: MP4 file format. ISO/IEC 14496-14:2003
- [ISO_14496-3] International Organization for Standardization. Information technology — Coding of audio-visual objects — Part 3: Audio. ISO/IEC 14496-3:2009
- [ISO_15444-1] International Organization for Standardization. Information technology — JPEG 2000 image coding system: Core coding system. ISO/IEC 15444-1:2004
- [ISO_15444-3] International Organization for Standardization. Information technology — JPEG 2000 image coding system: Motion JPEG 2000. ISO/IEC 15444-3:2007
- [ISO_15445] International Organization for Standardization. Information technology — Document description and processing languages — HyperText Markup Language (HTML). ISO/IEC 15445:2000
- [ISO_15948] International Organization for Standardization. Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2004
- [ISO_19005] International Organization for Standardization. Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1). ISO 19005:2005
- [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 19005-2:2012
- [ISO_19136] International Organization for Standardization. Geographic information -- Geography Markup Language (GML). ISO 19136:2007

- [ISO_26300:2006] International Organization for Standardization. Information technology — Open Document Format for Office Applications (OpenDocument) v1.0. ISO/IEC 26300:2006
- [ISO_26300:2015] International Organization for Standardization. Information technology — Open Document Format for Office Applications (OpenDocument) v1.2. ISO/IEC 26300:2015
- [ISO_28500] International Organization for Standardization. Information and documentation — WARC file format. ISO 28500:2009
- [ISO_29500-1] International Organization for Standardization. Information technology — Document description and processing languages — Office Open XML File Formats — Part 1: Fundamentals and Markup Language Reference. ISO/IEC 29500:2012
- [ISO_32000-1] International Organization for Standardization. Document management — Portable document format — Part 1: PDF 1.7. ISO 32000-1:2008
- [ISO_8859-15] International Organization for Standardization. Information technology — 8-bit single-byte coded graphic character sets — Part 15: Latin alphabet No. 9. ISO/IEC 8859-15:1999
- [KML] Open Geospatial Consortium. OGC KML 2.3. <http://docs.openeospatial.org/is/12-007r2/12-007r2.html>
- [LAC] Library and Archives Canada. Trusted Digital Repository, Local Digital Format Registry (LDFR), File Format Guidelines for Preservation and Long-term Access, Version 1.0
- [LPCM] Linear Pulse Code Modulated Audio (LPCM). Library of Congress. February 19, 2008
- [MATROSKA] Matroska Specifications. IETF Draft. October 2019. <https://tools.ietf.org/html/draft-ietf-cellar-matroska-04>
- [METS] Metadata Encoding and Transfer Standard. <http://www.loc.gov/standards/mets/>
- [MIX] NISO Data Dictionary Technical Metadata for Digital Still Images. <http://www.loc.gov/standards/mix/>
- [MOV] QuickTime File Format. Classic Version of the QuickTime File Format Specification. Apple Computer, Inc. 1.3.2001. <http://developer.apple.com/standards/qtff-2001.pdf>
- [MXF] Material Exchange Format (MXF) - File Format Specification (Revision of SMPTE 377-1-2009).
- [PDF_1.2] Portable Document Format Reference Manual. Version 1.2. Adobe Systems Incorporated. November 12, 1996. <http://www.pdf-tools.com/public/downloads/pdf-reference/pdfreference12.pdf>
- [PDF_1.3] PDF reference, second edition. Adobe Portable Document Format. Version 1.3. Adobe Systems Incorporated. July, 2000. <http://partners.adobe.com/public/developer/en/pdf/PDFReference13.pdf>
- [PDF_1.4] PDF reference, third edition. Adobe Portable Document Format. Version 1.4. Adobe Systems Incorporated. December, 2001. http://wwwimages.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference.pdf
- [PDF_1.5] PDF reference, fourth edition. Adobe Portable Document Format. Version 1.5. Adobe Systems Incorporated. Aug, 2003. http://wwwimages.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference15_v5.pdf
- [PDF_1.6] PDF reference, fifth edition. Adobe Portable Document Format. Version 1.6. Adobe Systems Incorporated. Nov, 2004. http://wwwimages.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference16.pdf
- [PREMIS] Data Dictionary for Preservation Metadata: PREMIS version 2.2. July 2012. <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>

[PRONOM]	The National Archives, UK. The technical registry PRONOM. http://www.nationalarchives.gov.uk/aboutapps/pronom/
[RFC2046]	Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types. Request for Comments: 2046. November 1996
[SIARD2.0]	SIARD Format Specification 2.0. Swiss National Standard eCH-0165. 2016. https://www.ech.ch/dokument/90686c1d-35d1-4ce8-bf8e-ec0ff923c53c
[SIARD2.1]	SIARD Format Specification 2.1, Swiss National Standard eCH-0165. 2018. https://www.bar.admin.ch/dam/bar/it/dokumente/kundeninformation/siard_formatbeschreibung.pdf.download.pdf/siard_format_descriptioning.pdf
[SMPTE_314]	Society of Motion Pictures & Television Engineers. Standard for Television - Data Structure for DV-Based Audio, Data and Compressed Video ó25 and 50 Mb/s. SMPTE 314M-1999
[SMPTE_370]	Society of Motion Pictures & Television Engineers. Standard for Television - ata Structure for DV-Based Audio, Data and Compressed Video at 100 Mb/s 1080/60i, 1080/50i, 720/60p, 720/50p. SMPTE 370M-2006.
[SMPTE_421]	Society of Motion Pictures & Television Engineers. Standard for Television - VC-1 Compressed Video Bitstream Format and Decoding Process. ST 0421-2006.
[SVG]	W3C: Scalable Vector Graphics (SVG) 1.1 specification (Second Edition). https://www.w3.org/TR/SVG11/
[TIFF]	Adobe Developers Association. TIFF Revision 6.0 Final — June 3, 1992. http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf
[UNICODE]	The Unicode Consortium. The Unicode Standard, Version 6.0.0, (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6). http://www.unicode.org/versions/Unicode6.0.0/
[VIDEOMD]	Video Technical Metadata Extension Schema. http://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html
[WAV_IBM_MS]	Multimedia Programming Interface and Data Specifications 1.0. Issued as a joint design by IBM Corporation and Microsoft Corporation. August 1991. http://www.kk.ij4u.or.jp/~kondo/wave/mpidata.txt
[WAV_LOC]	"WAVE Audio File Format". Library of Congress. http://www.digitalpreservation.gov/formats/fdd/fdd000001.shtml
[XHTML_1.0]	XHTML 1.0 The Extensible HyperText Markup Language (Second Edition). A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000, revised 1 August 2002. http://www.w3.org/TR/xhtml1/
[XHTML_1.1]	XHTML 1.1 — Module-based XHTML — Second Edition. W3C Recommendation 23 November 2010. http://www.w3.org/TR/xhtml11/
[XML_1.0]	Extensible Markup Language (XML) 1.0 (Fifth Edition. W3C Recommendation 26 November 2008. http://www.w3.org/TR/xml/