



Kansallisarkisto

Vastaanotettavia tiedostoja koskevat vaatimukset



Sähköisen arkistoinnin palvelu



Sisälllys

1. OHJEEN TARKOITUS	1
1.1. VIITTAUKSET MUIHIN OHJEISIIN JA DOKUMENTTEIHIN	1
2. TIEDOSTOJA KOSKEVAT YLEISET VAATIMUKSET	2
3. VASTAANOTETTAVAT TIEDOSTOMUODOT	2
4. TIEDOSTO- JA HAKEMISTONIMIÄ KOSKEVAT RAJOITUKSET	3
5. TEKSTITIEDOSTOSSA SALLITUT MERKISTÖT	4
LIITE 1 KANSALLISARKISTOON VASTAANOTETTAVAT TIEDOSTOMUODOT	5
LIITE 2 TEKNISET VAATIMUKSET AV-TIEDOSTOILLE	8
1. VASTAANOTETTAVIEN ÄÄNIAINEISTOJEN TIEDOSTOFORMAATIT (PÄIVITETTY 23.12.2022):	8
2. VASTAANOTETTAVIA VIDEOAINEISTOJEN TIEDOSTOFORMAATIT (PÄIVITETTY 23.12.2022):	9

1. Ohjeen tarkoitus

Tässä ohjeessa kuvataan Kansallisarkistoon vastaanotettavia tiedostoja koskevat yleiset vaatimukset.

Tiedostoja koskevat aineistotyyppikohtaiset vaatimukset löytyvät aineistotyyppikohtaisista ohjeista (kts. kohta 1.1)

Ohjeen liitteenä on luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista. Luettelo on päivittyvä.

1.1. Viittaukset muihin ohjeisiin ja dokumentteihin

Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin

Dokumentti	Sisältö
Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista (LIITE 1)	Päivittyvä luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista.
Teknisen vaatimukset AV-tiedostoille (LIITE 2)	Kansallisarkistoon vastaanotettavia AV-tiedostomuotojen teknisiä vaatimuksia tarkentava päivittyvä ohje.
SÄHKE2-aineistot - siirtopaketin muodostaminen	Ohjeessa kuvataan, miten muodostetaan Kansallisarkistoon siirrettävä SÄHKE2-tietomallin mukainen siirtopaketti.
Digitoidut kuva-aineistot - siirtopaketin muodostaminen	Ohjeessa kuvataan, miten muodostetaan Kansallisarkistoon siirrettävä paketti viranomaisen itse digitoimista aineistoista.
Rakenteiset tietoaineistot - siirtopaketin muodostaminen	Ohjeessa kuvataan, miten rakenteisesta datasta muodostetaan Kansallisarkistoon siirrettävä siirtopaketti.
Muut digitaaliset tietoaineistot - siirtopaketin muodostaminen	Ohjeessa kuvataan, miten aineistosta, joka ei sovellu muissa siirtorakenteissa siirrettäväksi, muodostetaan Kansallisarkistoon siirrettävä siirtopaketti.
Kansallisarkiston vaatimukset hävittämiseen tähtäävään digitointiin (KA/3357/07.01.02.04.02/2019)	Kansallisarkiston vaatimukset digitoinnille, joka mahdollistaa A4 ja foliokokoisten asiakirjojen analogisen ilmentymän hävittämisen.

Dokumentti	Sisältö
Kansallisarkiston määräykset arkistoitavien erikoisaineistojen digitointiin (KA/15780/07.01.02.04.02/2020)	Kansallisarkiston määräykset arkistoitavien erikoisaineistojen, esim. videot ja kuvakokoelmat, digitointiin.

2. Tiedostoja koskevat yleiset vaatimukset

Siirrettävissä tiedostoissa **pitää** huomioida seuraavat rajoitukset:

- Tiedostojen **pitää** olla Kansallisarkistoon vastaanotettavassa tiedostomuodossa.
- Tiedostojen **pitää** olla eheitä, hyvin muodostettuja¹ ja valideja².
- Tiedostoissa **ei saa** käyttää teknisiä salausrakenteita tai muita mekanismeita, jotka estävät aineiston avaamisen, säilyttämisen, konvertoinnin ja myöhemmän käytön tai vaikuttavat asiakirjan sisältöön.
- Tiedostoja tai tiedostokokoelmia **ei saa** pakata (esimerkiksi ZIP-paketti). Pakkaaminen estää tiedostojen käsittelyn vastaanoton yhteydessä.
- Tiedostosta **ei saa** puuttua sen esittämiseen ja hyödyntämiseen liittyviä osia (esimerkiksi fontit).
- Tiedostojen **pitää** olla itsenäisesti käsiteltävissä. Ne **eivät saa** olla riippuvaisia ulkopuolisista tietovarannoista tai palveluista.
- Tiedostoissa mahdollisesti olevat makrot tai vastaavat sovelluskoodit **eivät saa** estää tiedoston käsittelyä tai muuttaa tietosisältöä tiedostoa avattaessa.

Kansallisarkiston määräysten mukaisesti digitoitujen aineistojen yksityiskohtaiset vaatimukset löytyvät Kansallisarkiston digitointiohjeista. Katso Taulukko 1. Viittaukset muihin ohjeisiin ja dokumentteihin.

3. Vastaanotettavat tiedostomuodot

Luettelo Kansallisarkistoon vastaanotettavista tiedostomuodoista päivitetään säännöllisesti ja uusia tiedostomuotoja lisätään listalle.

¹ Esim. xml-tiedoston tulee noudattaa xml-formaatin syntaksia, CSV-tiedoston kenttien lukumäärä tulee olla sama jokaisella tiedoston rivillä.

² Tiedoston tulee noudattaa dokumenttityypin määräyksiä, esimerkiksi PDF/A-tiedoston tulee vastata ISO-standardia ja XML-tiedoston käytettyä skeemaa.

Asiakirja otetaan Kansallisarkistoon vastaan pääsääntöisesti vain yhdessä tiedostomuodossa. On suositeltavaa siirtää asiakirja alkuperäisessä tiedostomuodossaan aina kun se on mahdollista; yleisesti ottaen konversio tulisi tehdä ainoastaan silloin, kun alkuperäinen tiedostomuoto ei ole Kansallisarkistoon vastaanotettavien tiedostoformaattien joukossa.

Mikäli tiedostoista tuotetaan elinkaaren aikaisemmassa vaiheessa esimerkiksi PDF/A-muotoinen versio esimerkiksi julkaisua ja jakelua varten, myös alkuperäinen tiedostomuoto tulisi säilyttää.

Tiedostomuodot:

- Kansallisarkisto suosittaa, että siirtokelpoisia tiedostomuotoja ei konvertoida toiseen tiedostomuotoon siirtoa varten.

4. Tiedosto- ja hakemistonimiä koskevat rajoitukset

Tiedostojen ja hakemistojen nimeämistä koskevat yksityiskohtaiset ohjeet ja rajoitukset löytyvät siirtopakettien **aineistotyypikohtaisista** muodostamisohjeista.

Taulukkoon 4 ja taulukkoon 5 on koottu tiedostonimien ja hakemistojen nimien pituutta ja sallittuja merkkejä koskevat yleiset rajoitukset.

Taulukko 4. Rakenteiset tietoaineistot, muut digitaaliset tietoaineistot ja digitoidut kuva-aineistot -siirtorakenteiden tiedostonimiä ja hakemistoja koskevat rajoitukset

Tiedostonimen maksimipituus	256 merkkiä
Sallitut merkit	kirjaimet a-z, A-z numerot 0-9
Merkkikoko	Hakemistojen nimet ovat merkkikokoriippuvaisia (engl. <i>case-sensitive</i>). Tiedostojen päätteet <i>tulee</i> esittää pienin kirjaimin. ESIMERKKI oikein 0001.csv, väärin 0001.CSV

Taulukko 5. SÄHKE2-tietoaaineistot -siirtorakenteen tiedostonimiä ja hakemistoja koskevat erityisohjeet

Tiedostonimen maksimipituus	256 merkkiä
Sallitut merkit	kirjaimet a-ö, A-Ö Huom! Skandinaavisia merkkejä voidaan käyttää ainoastaan SÄHKE2-muotoisen aineiston yhteydessä
Merkkikoko	Hakemistoissa ja sahke.xml-tiedostossa esitetyt tiedot, myös tiedostonimet ja tiedostopäätteet, ovat merkkikokoriippuvaisia (engl. <i>case-sensitive</i>). Tiedoston nimi tai sen päätte voidaan esittää isoin kirjaimin, mutta viittauksen tiedostoon tulee sahke.xml-tiedostossa vastata tiedostopäätteen merkkikokoa. <i>ESIMERKKI: Tiedoston nimi on F0055720.pdf, joten siihen kohdistuva viittaus F0055720.PDF on virheellinen.</i>

5. Tekstitiedostossa sallitut merkistöt

Sallitut merkistöt	<ul style="list-style-type: none">• ISO-8859-15• UNICODE: UTF-8, UTF-16, UTF-32
--------------------	--

Huom! Yksikin tiedostoon sisältyvä sallitun merkistön ulkopuolinen merkki aiheuttaa virheen vastaanotossa. Jos merkistö muunnetaan ennen siirtoa, varmista, ettei tiedostoon jää käytetyn merkistön ulkopuolisia, ei-sallittuja merkkejä.

LIITE 1 Kansallisarkistoon vastaanotettavat tiedostomuodot

Kansallisarkistoon vastaanotettavat tiedostomuodot (päivitetty 23.12.2022)

Tiedostoformaatti	PRONOM-tunniste (PUID)	Voidaan siirtää rakenteessa
JPEG part 1 versio 1.00; [ISO_10918-1]	fmt/42	Digitoidut tietoaineistot (2019 ja 2021), SÄHKE2, Muut digitaaliset tietoaineistot
JPEG part 1 versio 1.01; [ISO_10918-1]	fmt/43	Digitoidut tietoaineistot (2019 ja 2021), SÄHKE2, Muut digitaaliset tietoaineistot
JPEG part 1 versio 1.02; [ISO_10918-1]	fmt/44	Digitoidut tietoaineistot (2019 ja 2021), SÄHKE2, Muut digitaaliset tietoaineistot
TIFF 6.0; [TIFF]	fmt/353	Digitoidut tietoaineistot (2019 ja 2021), SÄHKE2, Muut digitaaliset tietoaineistot
XML versio 1.0; [XML_1.0]:	fmt/101	Rakenteiset tietoaineistot, SÄHKE2
JSON	x-fmt/817	Rakenteiset tietoaineistot, SÄHKE2
CSV	x-fmt/18	Rakenteiset tietoaineistot, SÄHKE2
Microsoft Word (97–2003)	fmt/40	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Word (2007 onwards); [ISO_29500-1]	fmt/412	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Excel (8)	fmt/61	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Excel (8X)	fmt/62	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Excel 2003 (11.0)	fmt/62	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Excel (2007 onwards); [ISO_29500-1]	fmt/214	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
Microsoft Powerpoint (97–2003)	fmt/126	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"

1.3.2023

Microsoft Powerpoint (2007 onwards); [ISO_29500-1]	fmt/215	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.2; [PDF_1.2]	fmt/16	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.3; [PDF_1.3]	fmt/17	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.4; [PDF_1.4]	fmt/18	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.5; [PDF_1.5]	fmt/19	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.6; [PDF_1.6]	fmt/20	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF 1.7; [ISO_32000-1]	fmt/276	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odt versio 1.0; [ISO_26300:2006]	fmt/136	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odt versio 1.1	fmt/290	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odt versio 1.2; [ISO_26300:2015]	fmt/291	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
ods versio 1.0; [ISO_26300:2006]	fmt/137	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
ods versio 1.1	fmt/294	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
ods versio 1.2; ISO_26300:2015]	fmt/295	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odp versio 1.0; [ISO_26300:2006]	fmt/13	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odp versio 1.1	fmt/292	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
odp versio 1.2; ISO_26300:2015]	fmt/293	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"

1.3.2023

PDF/A-1a; [ISO_19005]	fmt/95	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-1b; [ISO_19005]	fmt/354	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-2a; [ISO_19005-2]	fmt/476	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-2b; [ISO_19005-2]	fmt/477	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-2u; [ISO_19005-2]	fmt/478	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-3a; [ISO_19005-3]	fmt/479	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-3b; [ISO_19005-3]	fmt/480	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
PDF/A-3u; [ISO_19005-3]	fmt/481	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
TXT [RFC2046]	x-fmt/111	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
TXT (ISO 88591-15, UTF-8, UTF-16 tai UTF-32)	x-fmt/111	SÄHKE2, Muut digitaaliset tietoaineistot, Rakenteiset tietoaineistot - "documentation"
SIARD versio 2.1	fmt/1196	Rakenteiset tietoaineistot – "master"
AV-formaatit (ks. liite 2)		SÄHKE2, Muut digitaaliset tietoaineistot

LIITE 2 Vastaanotettavat ääni- ja videotiedostot ja niiden tekniset vaatimukset AV-tiedostoille
1. Vastaanotettavien ääniaineistojen tiedostoformaatit (päivitetty 23.12.2022):

Vastaanotettava formaatti	Versio ja standardi	Tunniste	Premis formatName	Huomaus	Tiedostopäätteet
Audio Interchange File Format (AIFF, AIFF-C)		x-fmt/135 x-fmt/136	audio/x-aiff	Vain häviötön LPCM koodattu ääni.	AIFF, AIF
Broadcast Wave Format (WAV, BWF)	Versio 2	fmt/527 fmt/141	audio/x-wav	Voi sisältyä MKV ja AVI -säiliömuotoihin. Vain häviötön LPCM koodattu ääni.	WAV
Free Lossless Audio Codec (FLAC)	1.2.1	fmt/279	audio/flac	Voi sisältyä AVI, MKV tai MOV -säiliömuotoihin.	FLAC
MPEG-4 AAC	ISO_14496-3	fmt/199	audio/mp4	Voi sisältyä AVI, MKV, MJ2, MP4 ja MOV -säiliömuotoihin.	MP4
Moving Pictures Expert Group (MPEG) MPEG-1 layer-3, MPEG-2 layer-3 (MP3)	MPEG-1 layer-3 ISO 11172-3 MPEG-2 layer-3 ISO 13818-3	MPEG-1 layer-3 fmt/134 MPEG-2 layer-3 fmt/134	audio/mpeg	Tämä tiedostomuoto voi sisältyä AVI, MKV, MP4 ja MOV -säiliömuotoihin.	MP3
Windows Media Audio (WMA)	WMA 9	fmt/132	audio/x-ms-wav	Tämä tiedostomuoto voi sisältyä AVI -säiliömuotoon.	WMA

Audio Video Interleave [AVI]		fmt/5	video/x-msvideo	Säiliömuoto, voi sisältää ääntä, videota tai molempia.	AVI
QuickTime [MOV]		x-fmt/384	video/quicktime	Säiliömuoto, voi sisältää ääntä, videota tai molempia.	MOV

2. Vastaanotettavia videoaineistojen tiedostomuodot (päivitetty 23.12.2022):

Vastaanotettava formaatti	Versio ja standardi	Tunniste	Premis formatName	Huomautus	Tiedostopäätteet
Matroska [MKV] FFV1 versio 3 [FFV1]	FFV1 versio 3 [FFV1]	fmt/569	video/x-ffv		MKV
Motion JPEG 2000 JPEG 2000 Part 1	ISO_15444-1	fmt/337 x-fmt/392	video/jpeg2000		MJ2, MJP2
MPEG-4 AVC – Advanced Video Coding (AVC)	ISO-14496-10	fmt/199	video/mp4	Voi sisältyä AVI, MKV, MP4 ja MOV -säiliömuotoihin.	MPG, MP4
Digital Video (DV ja sen variantit)	[IEC_61834, SMPTE_314, SMPTE_370];	x-fmt/152	video/dv	Voi sisältyä AVI ja MOV -säiliömuotoihin. DV-formaatti on videon raakamuoto.	DV
Moving Pictures Expert Group (MPEG-1, MPEG-2)	MPEG-1; [ISO_11172-2] MPEG-2; [ISO_13818-2]	fmt/649 fmt/640	video/mpeg	Voi sisältyä AVI, MKV ja MOV -säiliömuotoihin.	MPG
Windows Media Video (WMV)	WMV 9; [SMPTE_421]	fmt/133	video/x-ms-wmv		WMV
Audio Video Interleave [AVI]		fmt/5	video/x-msvideo		AVI



QuickTime [MOV]		x-fmt/384	video/quicktime	Huom! Ei Apple ProRes-videota.	MOV
--------------------	--	-----------	-----------------	-----------------------------------	-----