

# Arkistoitavien asiakirjojen muuttaminen sähköiseen muotoon



**KANSALLISARKISTO**

Ohje

14.5.2025

KA/8675/28.00/2025



## Sisällys

|     |                                                            |    |
|-----|------------------------------------------------------------|----|
| 1   | Sisältö ja tarkoitus .....                                 | 1  |
| 1.1 | Rajaukset ja viittaukset muihin asiakirjoihin .....        | 2  |
| 2   | Sähköiseen muotoon muuttamisen periaatteet .....           | 4  |
| 3   | Metatiedot, kuvailu ja dokumentointi .....                 | 5  |
| 4   | Skannaus.....                                              | 5  |
| 5   | Tiedostomuotoja koskevat suositukset .....                 | 6  |
| 5.1 | Kuvatiedostoja koskevat huomiot .....                      | 7  |
| 5.2 | PDF/A-tiedostomuotoa koskevat huomiot.....                 | 8  |
| 6   | Autenttisuuden ja eheyden varmistaminen .....              | 9  |
| 7   | Optinen tekstintunnistus (OCR) ja koneluettava muoto ..... | 10 |
|     | Ohjeen sanasto .....                                       | 11 |



## 1 Sisältö ja tarkoitus

Asiakirja sisältää Kansallisarkiston ohjeet ja suositukset analogisessa muodossa laadittujen ja saapuvien arkistoitavien asiakirjojen sähköiseen muotoon muuttamisesta. Ohje on tarkoitettu toimijoille, jotka suunnittelevat ja toteuttavat arkistolain (831/1994) nojalla arkistoitavaksi määrättyjen asiakirjojen sähköiseen muotoon muuttamista. Sen tarkoitus on tukea pitkäaikaissäilytykseen soveltuvien, käytettävien ja ymmärrettävien digitaalisten ilmentymien tuottamisessa. Toimijan on suunniteltava ja määriteltävä, miten se soveltaa ohjeen sisältöä ja siinä kuvattuja periaatteita omassa toiminnassaan.

Julkisen hallinnon tiedonhallintalaista annetun lain (tiedonhallintalaki, 906/2019, 19 §) mukaan muussa kuin sähköisessä muodossa viranomaiselle saapuva asiakirja on muutettava sähköiseen muotoon, jos asiakirja on säädetty pysyvästi säilytettäväksi taikka lailla tai lain nojalla arkistoitavaksi. Tiedonhallintalaki on tältä osin tullut sovellettavaksi 1.1.2022 alkaen, ja sitä mukailee Kansallisarkiston 22.12.2021 arkistoitavien asiakirjojen muodosta antama määräys (KA/15906/07.01.01.00/2021). Sen perusteella asiakirjat, jotka ovat muodostuneet 1.1.2022 alkaen, arkistoidaan yksinomaan sähköisessä muodossa<sup>1</sup>, ja analogisena muodostuvat asiakirjat muutetaan sähköiseen muotoon Kansallisarkiston ajantasaisten vaatimusten mukaisesti.

Arkistolaissa pysyvästi säilytettävillä asiakirjoilla tarkoitetaan asiakirjoja ja tietoaaineistoja, jotka niiden alkuperäisen käyttötarkoituksen mukaisen säilytysajan päättymisen jälkeen arkistoidaan lain tai lain nojalla annetun seulontapäätöksen mukaisesti. Määräajan säilytettävien ja alkuperäisessä käyttötarkoituksessa pysyvästi säilytettävien asiakirjojen käsittelyä ohjeistaa tiedonhallintalain perusteella Julkisen hallinnon tiedonhallintalautakunta.

### Voimassaolo

Toistaiseksi, ohjeen päiväyksestä eteenpäin.

### Säädökset, joihin ohjeen antaminen perustuu

Laki Kansallisarkistosta (1145/2016, 2 §)

Arkistolaki (831/1994, 11 §)

Laki sähköisestä asioinnista viranomaistoiminnassa (13/2003, 21 §)

Määräys arkistoitavien asiakirjojen muodosta 22.12.2021 (KA/15906/07.01.01.00/2021)

---

<sup>1</sup> Asiakirjat, joiden analogisella säilytysmuodolla on kulttuurihistoriallista arvoa, arkistoidaan digitoinnin jälkeen sekä analogisessa että digitaalisessa muodossa. Tällä hetkellä muodostuvat analogiset asiakirjat ovat pääasiassa erilaisia paperitulosteita, joilla ei lähtökohtaisesti ole kulttuurihistoriallista arvoa. Kulttuurihistoriallisen arvon arviointi tehdään 1.1.2022 alkaen muodostuville asiakirjoille vain silloin, kun toimija on tunnistanut tälle perustellun tarpeen.



## 1.1 Rajaukset ja viittaukset muihin asiakirjoihin

Ohjetta sovelletaan, kun tällä hetkellä muodostuvia analogisia arkistoitavia asiakirjoja muutetaan sähköiseen muotoon esimerkiksi asiankäsittelyn tai palvelujen tiedonhallinnan yhteydessä. Toimijalle jo muodostuneiden analogisten arkistollisten kokonaisuuksien takautuvaan digitointiin sovelletaan Kansallisarkiston ohjetta Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet (KA/20919/30.01.00/2024).<sup>2</sup>

Takautuvaa digitointia ja muodostuvien asiakirjojen sähköiseen muotoon muuttamista koskevat erot liittyvät toimintatapaan, todistusvoimaisuuden ja tietosisällön eheyden varmistamiseen sekä tietoaisteistojen muodostamiseen:

- Muodostuvien asiakirjojen sähköiseen muotoon muuttaminen:
  - Asiakirjat liitetään muodostumassa olevan digitaalisen tietoaisteistokokonaisuuden osaksi.
  - Valtaosa asiakirjoista on tasalaatuisia paperitulosteita. Digitaaliset ilmentymät tallennetaan yleensä tietojärjestelmiin samaan asiayhteyteen alkujaan digitaalisten asiakirjojen kanssa. Autenttisuus ja eheys varmistetaan sähköiseen muotoon muuttamisen ja tallentamisen yhteydessä.
- Analogisten arkistokokonaisuuksien takautuva digitointi:
  - Asiakirjat ovat yleensä muodostuneet pitkän ajan kuluessa, ja niiden tyypit, muoto, koko, materiaalit ja kunto voivat vaihdella huomattavasti.
  - Asiakirjojen todistusvoimaisuus ja käytettävyys perustuvat alkuperäiseen arkistointijärjestykseen ja luettelointitietoihin. Autenttisuuden ja eheyden varmistamiseksi niiden luoma yhteys muodostumis- ja käsittelykontekstiin on säilytettävä, vaikka asiakirjojen metatietoja digitoinnin yhteydessä rikastettaisiin muilla tavoilla.

Sähköiseen muotoon muuttamisen suunnittelun ja toteuttamisen yhteydessä on arvioitava myös asiakirjan muodon tai tietosisällön asettamia vaatimuksia (ks. luku [5 Tiedostomuotoja koskevat suositukset](#)). Arkistoitavien ja Kansallisarkistoon vastaanotettavien asiakirjojen tiedostomuotoja, siirto-rakenteita, vaadittuja metatietoja ja vastaanotettavia tiedostoja koskevat vaatimukset ovat Kansallisarkiston kulloinkin ajantasaisissa siirto-ohjeissa. Niitä noudatetaan, kun digitaalisia tietoaisteistoja siirretään Kansallisarkistoon.

<sup>2</sup> Ohjeessa on huomioitu arkistosektorilla yleisesti käytössä olevat standardit sekä muiden kansallisarkistojen laatuvaatimukset digitoinnille sekä pitkäaikaissäilytyspalveluiden (PAS-palvelut) kansalliset vaatimukset.

**Taulukko 1. Viittaukset muihin asiakirjoihin**

| ASIAKIRJA                                                                                                                                                                                                                         | SISÄLTÖ                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet 19.12.2024 (KA/20919/30.01.00/2024)                                                                                                                        | Arkistoitavien eli arkistolain (831/1994) mukaan pysyvästi säilytettäväksi määrättyjen asiakirjojen digitointia koskevat Kansallisarkiston vaatimukset. Ohjetta sovelletaan, kun toimijalle jo muodostuneita arkistollisia kokonaisuuksia digitoidaan takautuvasti. Ohjetta sovelletaan lisäksi tällä hetkellä saapuvien tai laadittavien analogisten audiovisuaalisten aineistojen (äänitteet ja kuvanauhat), negatiivien, lasinegatiivien, diojen, röntgenfilmien sekä mikrofilmien ja -korttien sähköiseen muotoon muuttamisessa.                   |
| Arkistoitavien tietoaineistojen perusmetatiedot -suositus 20.02.2024 (KA/20011/30.00.00/2023)                                                                                                                                     | Suosituksessa määritellään arkistoinnin edellyttämän metatiedon vähimmäistasoa. Organisaation on suunniteltava arkistometatiedon toteutustapa ja hyödyntäminen suhteessa tiedonhallinnassa ja tiedon elinkaaren hallinnassa vaadittavien muiden toimintatapojen, metatietojen, tietojen ja dokumentoinnin kokonaisuuteen. Kun tietoaineistoja siirretään Kansallisarkistoon, noudatetaan Kansallisarkiston voimassa olevia siirto-ohjeita.                                                                                                             |
| Suositus tietoturvallisuuden vähimmäisvaatimuksista. Valtioavarainministeriön julkaisuja – 2024:19.                                                                                                                               | Julkisen hallinnon tiedonhallintalautakunnan suositus opastaa tiedonhallintalain asettamien tietoturvallisuuden vähimmäisvaatimusten täyttämässä.                                                                                                                                                                                                                                                                                                                                                                                                      |
| Julkishallinnon pysyvästi säilytettävien digitoitujen ja digitoitavien asiakirjojen säilytysmuoto ja analogisen säilytysmuodon kulttuurihistoriallisen arvon arviointitietojen pysyvä säilytys 8.12.2023 (KA/15927/28.02.01/2023) | Päätöksessä määrätään ennen 1.1.2022 muodostuneet pysyvästi säilytettävät asiakirjat arkistoitaviksi yksinomaan digitaalisessa muodossa lukuun ottamatta analogiselta säilytysmuodoltaan kulttuurihistoriallisesti arvokkaita asiakirjoja. Toimijat vastaavat 1.1.2024 alkaen kulttuurihistoriallisen arvon arvioinnista noudattaen päätöksen kriteereitä. Arviointimenettelyä ja arviointitietojen säilytystä koskevia päätöksen osia sovelletaan ilman aikarajaa. Analogiset asiakirjat digitoidaan päätöksessä esitettyjen reunaehtojen mukaisesti. |
| Suositus SÄHKE2-metatietomallin hyödyntämisestä (KA/12266/07.01.01.00/2022) versio 2.0 <sup>3</sup>                                                                                                                               | Suositus SÄHKE2-metatietomallin hyödyntämisestä käsittelee SÄHKE2-määrityksiä tiedon arkistoinnin näkökulmasta sekä kommentoi SÄHKE2-metatietomallia (versio 1.3) lainsäädännön muutoksiin liittyvillä huomioilla.                                                                                                                                                                                                                                                                                                                                     |
| Määräys valtion viranomaisten arkistoitavien asiakirjojen siirtämisestä Kansallisarkistoon 18.5.2021 (KA/26492/03.099/2020)                                                                                                       | Määräyksessä määrätään Kansallisarkistoon siirrettävien arkistoitavien asiakirjojen muodosta ja siirtämisestä valtionhallinnon osalta. Arkistoitavat asiakirjat siirretään Kansallisarkistoon digitaalisessa muodossa, ellei Kansallisarkisto ole toisin määrännyt.                                                                                                                                                                                                                                                                                    |

<sup>3</sup> SÄHKE2 (Sähköisten asiakirjallisten tietojen käsittely, hallinta ja säilyttäminen, AL/ 9815/07.01.01.00/2008) oli Kansallisarkiston määräys. Määräyksen voimassaolo on päätynyt 31.12.2022, eikä sitä enää ylläpidetä.



## 2 Sähköiseen muotoon muuttamisen periaatteet

Arkistoitavien asiakirjojen sähköiseen muotoon muuttamista koskevat vaatimukset perustuvat:

- digitaalisen pitkäaikaissäilytyksen<sup>4</sup> näkökulmasta suositeltuihin tiedostomuotoihin
- tiedostojen laadun ja eheyden varmistamiseen
- tietosisällön todistusvoimaisuuden varmistamiseen
- tietoaaineistojen pitkäaikaisen intellektuaalisen<sup>5</sup> ja fyysisen hallinnan varmistamiseen.

Asiakirjasta tuotetaan arkistointia varten teknisiltä ominaispiirteiltään korkealaatuinen digitaalinen ilmentymä, joka toistaa alkuperäisen analogisen asiakirjan sisältämän tietosisällön mahdollisimman tarkasti. Tätä varten toimijan on määriteltävä riittävät toimintatavat ja tekniset ratkaisut sekä huolehdittava henkilöstön ohjeistamisesta ja laadun valvonnasta. Toimija voi tuottaa muissa tiedostomuodoissa olevia ilmentymiä muita käyttötarkoituksia varten.<sup>6</sup>

Perusprosessissa analogisen asiakirjan kaikki sivut skannataan. Skannauksen yhteydessä muodostunut data tallennetaan valittuun tiedostoformaattiin, jonka jälkeen tiedosto tai tiedostot voidaan tallentaa haluttuun tietojärjestelmään. Metatietojen ja tiedostojen muodostumista sekä tiedostojen teknistä validointia voi automatisoida esimerkiksi työnkulkujen avulla.

Asiakirjojen analogisilla ilmentymillä ei ole sähköiseen muotoon muuttamisen jälkeen arkistollista arvoa, ja ne voidaan digitaalisten ilmentymien todistusvoimaisuuden ja säilymisen varmistamisen jälkeen hävittää toimijan määrittelemän ajan kuluttua.

Jos asiakirjaa ei voida asianmukaisesti muuttaa sähköiseen muotoon esimerkiksi sen kunnon tai muodon vuoksi, se arkistoidaan alkuperäisessä analogisessa muodossaan. Mikäli sähköiseen muotoon muuttaminen onnistuu myöhemmin, analoginen ilmentymä voidaan hävittää ja arkistoida asiakirja yksinomaan digitaalisessa muodossa.

<sup>4</sup> Digitaalisessa pitkäaikaissäilytyksessä aineistot muunnetaan tarvittaessa hallitusti uuteen säilytyskelpoiseen tiedostomuotoon, kun teknologiset muutokset sitä vuosien ja vuosikymmenten saatossa vaativat. Aineiston kaikkien keskeisten ominaisuuksien pitää säilyä luotettavasti. Tämä perustuu mm. tiedostomuotostandardien noudattamiseen.

<sup>5</sup> Intellektuaalinen hallinta tarkoittaa aineiston tiedollista ja käsitteellistä hallintaa, kuten kuvailua, metatietoja ja muuta aineiston mallintamista, joka auttaa hahmottamaan sen kokonaisuutta ja sisältöä. Lähde: Arkistosektorin aineistohallinnan arkkitehtuuri, versio 2.3, 30.3.2021. Viitattu 15.1.2025

<sup>6</sup> Nk. käyttökappaleet ovat asiakirjan erilaisia käyttötarkoituksia palvelevia ilmentymiä.



### 3 Metatiedot, kuvailu ja dokumentointi

Organisaation on linjattava, miten sähköiseen muotoon muuttaminen dokumentoidaan, millaista metatietoa tuotetaan ja mihin se tallennetaan. Kuvailevan ja hallinnollisen metatiedon lisäksi tulee huomioida sähköiseen muotoon muuttamista, muodostuneita tiedostoja ja niiden käsittelyhistoriaa dokumentoiva tekninen metatieto. Skannerien ja skannausohjelmistojen tuottamien metatietojen täydentämisessä voi hyödyntää myös automaattisia skriptejä.

Vaatimukset, linjaukset ja toimintatavat, joihin asiakirjan sähköiseen muotoon muuttaminen on perustunut, voivat vaikuttaa digitaalisten ilmentymien todistusvoimaisuuteen, ymmärrettävyyteen ja hallintaan. Ne pitää tarvittaessa kuvailla ja dokumentoida tarkoituksenmukaiselle tasolle asiakirjan tai tietoaaineiston kuvailutietoihin.<sup>7</sup>

**Esimerkki.** Asiakirja tai monisivuinen asiakirjan sivut on skannattu vain toiselta puolelta sen jälkeen, kun on varmistettu, ettei kääntöpuoli sisällä informaatiota. Ratkaisu on dokumentoitava, jos se vaikuttaa todistusvoimaisuuteen ja digitaalisten ilmentymien ymmärrettävyyteen. Ilman alkuperäistä analogista ilmentymää ei esimerkiksi aina voi arvioida, onko kääntöpuolen tai sivun puuttuminen vahinko vai tarkoituksellista.

Asiakirjasta on sen elinkaaren aikana tavallista tuottaa versioita tai ilmentymiä erilaisia käyttötarkoituksia tai käyttöliittymiä varten. Eri säilytysmuodoissa, tiedostomuodoissa, tiedostoversioissa tai sijainneissa olevien ilmentymien hallinta pitää varmistaa riittävillä metatiedoilla, kuvailulla, dokumentoinnilla ja muilla tarvittavilla keinoilla. Tiedon on säilyttävä oikeellisenä ja johdonmukaisena huolimatta siitä, miten ja mihin se on tallennettu.<sup>8</sup>

### 4 Skannaus

Asiakirja voidaan skannata kameralla kuvaamalla tai skannerilla skannaamalla. Skannauksessa voi hyödyntää päivittäisessä toiminnassa käytettäviä skannereita tai monitoimilaitteita. Laitteisto pitää

<sup>7</sup> Dokumentoinnista ja kuvailusta ks. Suositus Sähke2-metatietomallin hyödyntämisestä, luku 4.4. Asiakirjan alkuperäisyyden ja eheyden todentaminen ja saapumistapa sekä Arkistoitavien tietoaaineistojen perusmetatiedot -suositus (KA/20011/30.00.00/2023), luku 3.3 Aineistojen ja ilmentymien elinkaaritapahtumat ja kohdat 4.1.4 Elinkaaritapahtumat ja 4.1.7 Alkuperäisyys ja eheys ja 4.4.6 Ilmentymän elinkaaritapahtumat.

<sup>8</sup> Ks. ilmentymien hallinnasta esimerkiksi Kansallisarkiston Suositus Sähke2-metatietomallin hyödyntämisestä -suosituksen mukaisen metatietomallin metatiedot 4.8 Suhde, 4.11. Versio ja 6.1.2. Käyttötarkoitus.



kuitenkin mitoittaa aineistolle sopivaksi, siten että kuva-ala on riittävä, eikä skannaus aiheuta vahinkoa alkuperäiselle asiakirjalle. Tarvittaessa on hyödynnettävä eri aineistotyypeille soveltuvia skannereita ja laitteita.<sup>9</sup>

Autenttisuuden, todistusvoimaisuuden ja eheyden varmistamiseksi analoginen ilmentymä pitää muuttaa sähköiseen muotoon kokonaisuutena ja tietosisällöltään sellaisena kuin se on:

- Informaatiota ei saa jäädä muuntamatta digitaaliseen muotoon. Jokaisesta yksittäisestä tiedostosta pitää silmämääräisellä tarkastelulla saada kaikki sama informaatio kuin sen analogisesta vastineesta.
- Muodostuva kuvatiedosto ei saa sisältää sellaisia elementtejä, joita ei ole analogisessa vastineessa.
- Merkintöjä sisältäviä sivuja ei saa poistaa. Mikäli tietoaineistokokonaisuuden kääntöpuolet eivät sisällä mitään informaatiota, ne voidaan harkinnan jälkeen jättää skannaamatta.
- Kuvatiedosto ei saa olla epätarkka, vääristynyt tai mittasuhteiltaan muuttunut, jotta tietosisällön tulkinta ei vaarannu. Asiakirjan pitää olla kuvaushetkellä tasainen.

## 5 Tiedostomuotoja koskevat suositukset

Tiedostomuotoja, versioita ja tiedostojen pakkaamista koskevat ohjeet on koottu taulukkoon 2.

**Taulukko 2. Suositellut tiedostomuodot**

| Tiedostomuoto | Versio           | Pakkaaminen                                                        | Asiakirjan sisältö tai muoto                                  |
|---------------|------------------|--------------------------------------------------------------------|---------------------------------------------------------------|
| TIFF          | versio 6.0       | Pakkaamaton tai LZW                                                | Valokuvat ja muut kuvasisällöt, kuten kartat ja piirustukset. |
| JPEG          | part 1, versio 1 | JPEG-laatusa 90 %                                                  | Pääosin tekstimuotoinen sisältö.                              |
| PDF/A         | 1b tai 2b        | Ks. luku <a href="#">5.2 PDF/A-tiedostomuotoa koskevat huomiot</a> | Pääosin tekstimuotoinen sisältö.                              |

Sähköiseen muotoon muuttamisessa on suositeltavaa noudattaa Kansallisarkiston Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet (KA/20919/30.01.00/2024) -ohjeen aineistotyyppikohtaisia tiedostomuotovaatimuksia, teknisiä vaatimuksia ja raja-arvoja.

<sup>9</sup> Aineistotyyppikohtaisista ohjeista ks. ohje Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet 19.12.2024 (KA/20919/30.01.00/2024).



Suositteltu tiedostomuoto on analogisen asiakirjan muodosta tai tietosisällöstä riippuen häviöttömästi pakattu TIFF tai häviöllisesti pakattu JPEG. PDF/A-tiedostomuodon hyödyntämistä pitää arvioida suhteessa asiakirjan muotoon ja tietosisältöön.

PDF/A on PDF-tiedostomuodon ISO-standardoitu versio, joka on kehitetty erityisesti arkistoinnin ja pitkäaikaissäilyttämisen vaatimusten näkökulmasta. Kuvatiedostomuotoihin verrattuna siihen kuitenkin liittyy muun muassa seuraavia rajoitteita:

- Monisivuisen, pääosin tekstimuotoisen asiakirjan tallentaminen yhdeksi PDF/A-tiedostoksi parantaa yleensä asiakirjan käytettävyyttä päivittäisessä toiminnassa. PDF/A-tiedostomuoto vaikeuttaa kuitenkin tiedon jalostamista ja jatkokäsittelyä verrattuna suositeltuihin kuvatiedostomuotoihin. Tietosisällön käytettävyyttä voi parantaa myös tuottamalla PDF/A-muotoisen käyttö-kappaleen kuvatiedoston rinnalle.
- Tarkkuuteen, yksityiskohtien toistumiseen, sävyntoistoon sekä mittasuhteiden vääristymättömyyteen liittyvät vaatimukset korostuvat valokuvien ja muiden kuvasisältöjen (esimerkiksi kartat ja piirustukset) kohdalla. Niille suositeltu tiedostomuoto on kuvatiedostomuoto (TIFF 6.0, LZW). Mikäli kuvasisältöjä digitoidaan laajempina kokonaisuuksina, on suositeltavaa noudattaa Arkistoitavien asiakirjojen digitointia koskevat vaatimukset ja ohjeet (KA/20919/30.01.00/2024) -ohjetta.

Sähköiseen muotoon muuttamisen suunnittelussa ja toteutuksessa on huomioitava, ettei tallennusformaatin laatua voi enää myöhemmässä käsittelyssä parantaa.

## 5.1 Kuvatiedostoja koskevat huomiot

Sähköiseen muotoon muuttamisen yhteydessä asiakirjan sivut skannataan molemmilta puolilta, ja kummastakin puolesta muodostuu oma kuvatiedostonsa.

Muodostuvan kuvatiedoston pitää olla fyysisiltä mitoiltaan (n cm x n cm) 1:1-koossa analogisen asiakirjan kanssa resoluutiolla 300 ppi. Asiakirjan koko tai tietosisältö voivat vaatia resoluution nostamista. Tietosisällön pitää olla luettavissa digitaalisesta kuvasta samalla tavalla kuin alkuperäisestä analogisesta asiakirjasta. Resoluutiota ei saa muuttaa kuvatiedostoa käsiteltäessä, jotta analogisen asiakirjan fyysisten mittojen arviointi on digitaalisen ilmentymän perusteella mahdollista. Valokuvien kohdalla tavoitteena on aina ~A3-kokoinen (cm x cm) tiedosto resoluutiolla 300 ppi, joten niissä sovelletaan alla olevaa resoluutiotaulukkoa (Taulukko 3).

**Taulukko 3. Valokuvien resoluutiotaulukko**

| Analogisen asiakirjan koko | Resoluutio |
|----------------------------|------------|
| 6x9 cm                     | 1 600 ppi  |
| 9x12 cm                    | 1 200 ppi  |
| 13x18 cm                   | 900 ppi    |
| 18x24 cm                   | 600 ppi    |
| A3 (tai suurempi)          | 300 ppi    |

Kuvatiedosto pitää tallentaa 24-bittisenä RGB-kuvana. Siihen pitää liittyä käytettyä väriavaruutta vastaava ICC-väriprofiili<sup>10</sup>, jonka avulla varmistetaan, että kuvat näkyvät oikein eri laitteilla ja ohjelmistoilla. Jos väriprofiili lisätään tai sitä muunnetaan jälkikäteen, pitää varmistaa, että tämä tehdään oikein ja värintoisto vastaa alkuperäistä kuvaa.

**5.2 PDF/A-tiedostomuotoa koskevat huomiot**

PDF/A-muunnoksen yhteydessä on varmistettava, ettei mahdollisesti aiemmin luotuja kuvatiedostoja pakata uudelleen tai tiedostojen tekniset ominaisuudet muuten muutu käsittelyn yhteydessä. Mikäli skannaussovellus muodostaa PDF/A-tiedoston suoraan skannausprosessissa, tulee varmistua siitä, millaista muodostamistapaa sovellus hyödyntää.

On tärkeää varmistaa, että muodostunut tiedosto noudattaa formaatille PDF/A-standardissa asetettuja vaatimuksia (tiedoston tekninen validointi).

PDF/A-tiedoston pakkaaminen häviöllisesti tiedostokoon pienentämiseksi vaikuttaa alkuperäisen kuvatiedoston laatuun ja voi vaarantaa arkistoitavaksi määrätyn tiedon säilymisen. Huomaa myös, että LZW-pakkaus ei ole PDF/A-tiedostomuotostandardissa sallittu.

<sup>10</sup> Ohjeen KA/20919/30.01.00/2024 mukaiset väriprofiilit: TIFF: eciRGB v2, ProPhoto RGB, AdobeRGB (1 998) ja JPEG: sRGB, eciRGB v2, ProPhoto RGB tai AdobeRGB (1 998).



## 6 Autenttisuuden ja eheyden varmistaminen

Sähköiseen muotoon muuttamiseen pitää kuulua varmennus siitä, että digitaalisen ilmentymän tietosisältö on viranomaisessa tarkistettu ja varmistettu.<sup>11</sup> Toimijan pitää määritellä skannausta ja tiedostoja koskeva laadunvarmistus- ja validointiprosessi sekä toimintatavat digitaalisten ilmentymien todistusvoimaisuuden arviointiin. Asiakirjan digitaalisen ilmentymän alkuperäisyys (autenttisuus) ja eheys tarkistetaan asiakirjaa teknisesti ja laadullisesti tarkastelemalla, ja niitä koskevat merkinnät tallennetaan asiakirjan metatiedoiksi.<sup>12</sup>

Erilaisten validointityökalujen avulla voidaan tarkistaa, että muodostunut tiedosto noudattaa tiedostomuotonsa mukaista standardia (**tekninen validointi**). Tiedostomuodon tunnistaminen ja tiedostojen laadun valvonta voidaan toteuttaa esimerkiksi automatisoituina työnkulkuina tai vähintään testaamalla laitteiston tuottamien tiedostojen laatua säännöllisesti. Myös tiedostojen muodostumista voi automatisoida erikseen määriteltyjen työnkulkujen avulla.

Tiedostojen teknisen validoinnin (tekninen eheys) lisäksi sähköiseen muotoon muuttamisen yhteydessä pitää havainnoida tietosisällön säilymistä täydellisenä ja eheänä (**semanttinen eheys**). Sisällön on vastattava alkuperäistä tietoa. Siinä ei saa olla puutteita eikä se saa sisältää elementtejä, joita ei ole sen analogisessa vastineessa. Tätä ja mahdollisia laatueroja pitää havainnoida esimerkiksi silmäämällä. Mikäli asiakirja tallennetaan yhdeksi PDF/A tiedostoksi, muodostunut tiedosto pitää selata läpi ja varmistaa, ettei yhdistettyjen sivujen tietosisältö ole puutteellinen tai vioittunut.

Tietosisällön tarkistuksen ja varmennuksen perusteella pitää pystyä jälkikäteen todentamaan, onko sähköisessä muodossa olevaan asiakirjaan tehty muutoksia. Tiedostojen muuttumattomuutta voidaan seurata tarkistussumman eli eheystarkisteen avulla. Yleisiä eheysalgoritmeja ovat MD5, SHA-1, SHA-256 ja SHA-512. Toimintatapa tai elinkaaren vaihe, jossa eheystarkiste lasketaan, voi vaihdella hyödynnetyn tietojärjestelmän tai säilytysratkaisun mukaan.

<sup>11</sup> Tiedonhallintalain 19 §:n mukaan viranomainen vastaa siitä, että sähköiseen muotoon muutetun asiakirjan luotettavuus ja eheys varmistetaan. Sähköisestä asioinnista viranomaistoiminnassa annetun lain (13/2003) 21 §:n mukaan sähköinen asiakirja on arkistoitava niin, että sen alkuperäisyys ja säilyminen sisällöltään muuttumattomana voidaan myöhemmin osoittaa.

<sup>12</sup> Ks. mm. Suositus SÄHKE2-metatietomallin hyödyntämisestä, luku 4.4. Asiakirjan alkuperäisyyden ja eheyden todentaminen ja saapumistapa.



Sähköiseen muotoon muuttamisen yhteydessä voi tietoturvallisuussyistä tai muuten olla tarve hyödyntää sähköisiä allekirjoituksia, varmenteita tai erilaisia salaustekniikoita.<sup>13</sup> Arkistoitavien asiakirjojen autenttisuuden ja eheyden varmistaminen tulisi kuitenkin todentaa pääsääntöisesti metatietojen avulla. Kansallisarkistoon siirrettävän asiakirjan tulisi olla tulkittavissa ja hyödynnettävissä ilman varmenteisiin liittyviä mekanismeja. Myös sähköistä allekirjoitusta koskevat tiedot tulee arkistoida ensisijaisesti asiakirjan metatietona.<sup>14</sup>

Sähköisten allekirjoitusten tai leimojen yhteydessä on validoitava, että tiedosto on niiden lisäämisen jälkeen edelleen tiedostomuotostandardinsa mukainen.

## 7 Optinen tekstintunnistus (OCR) ja koneluettava muoto

Optisella tekstintunnistuksella (engl. Optical Character Recognition, OCR) voidaan tunnistaa skannattujen sivujen tai kuvien sisällöstä kirjainmuotoja (merkkejä), jotka muunnetaan tekstimuotoiseksi dataksi. Tekstintunnistuksen laatuun ja virheettömyyteen vaikuttavat alkuperäisen analogisen asiakirjan sisällön ja kunnon lisäksi skannauksen ja muodostuneen kuvan laatu.

Tiedonhallintalain 19 §:n mukaan viranomaisen on huolehdittava, että tietoaineisto on saatavilla ja hyödynnettävissä yleisesti käytettävässä koneluettavassa muodossa kuvailutietoineen, jos tietoaineisto voidaan saattaa alkuperäisestä muodosta suoraan koneluettavaan muotoon. Koneluettava muoto tarkoittaa sellaista rakennetta ja tiedostomuotoa, josta ohjelmistot pystyvät helposti yksilöimään, tunnistamaan ja poimimaan tietoja sekä niiden rakenteita. Tekstintunnistus yksin ei tarkoita sitä, että tietoaineisto on koneluettavassa muodossa.

Kansallisarkistoon siirrettäviltä sähköiseen muotoon muutetuilta asiakirjoilta ei edellytetä tekstintunnistusta tai koneluettavaa muotoa. Mikäli arkistointia varten muodostettu PDF/A-tiedosto<sup>15</sup> sisältää myös tekstintunnistetun sisällön, pitää tekstintunnistuksen yhteydessä välttää PDF/A-tiedostoksi tallennettujen alkuperäisten kuvatiedostojen laadun heikentämistä ja varmistaa, ettei tekstintunnistus vahingossa korvaa kuvatiedostoa tai sen sisältöä.

<sup>13</sup> Ks. esim. Julkisen hallinnon tiedonhallintalautakunnan Suositus tietoturvallisuuden vähimmäisvaatimuksista sähköiseen muotoon muuttamisen yhteydessä (s. 30). Viitattu 25.4.2025.

<sup>14</sup> Kansallisarkisto ei ota kantaa siihen, mikä on juridisessa mielessä riittävä allekirjoitustekniikka. Ks. Suositus Sähke2-metatietomallin hyödyntämisestä, luku 4.4. Asiakirjan alkuperäisyyden ja eheyden todentaminen ja saapumistapa ja luku 4.8. Sähköiset allekirjoitukset.

<sup>15</sup> Kuvatiedostoformaattit (TIFF, JPEG) eivät voi suoraan sisältää OCR:n tuottamaa tekstintunnistustietoa. OCR-tulos säilytetään niiden kohdalla erillisenä tiedostona, jonka yhteydet kuvatiedostoon voidaan varmistaa tunnisteilla, muilla metatiedoilla tai säilytysrakenteilla.



## Ohjeen sanasto

| Käsite                           | Selite                                                                                                                                                                                                                                                                                                                                 |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Analoginen asiakirja</b>      | Paperisessa tai muussa käsin kosketeltavassa muodossa laadittu, säilytetty ja käytettävä asiakirja (vrt. digitaalinen tieto tai asiakirja, joka muodostuu bittijonoista).                                                                                                                                                              |
| <b>Arkistointi</b>               | Osaksi asiakirjallista kulttuuriperintöä määriteltyjen asiakirjojen ja tietoaaineistojen säilyttäminen ja säilyvyyden ja käytettävyyden turvaaminen. Arkistoituja asiakirjoja ja tietoaaineistoja säilytetään pysyvästi eli käytännössä ilman aikarajaa. Arkistointi perustuu Kansallisarkiston seurantapäätöksiin tai lainsäädäntöön. |
| <b>Asiankäsittely</b>            | Viranomaisen tehtäviin liittyvien asioiden lainsäädännön mukainen ja järjestelmällinen hoitaminen eri käsittelyvaiheissa vireille tulosta ratkaisemiseen asiankäsittelyprosessin mukaisesti. Lähde: <u>Tiedonhallintalautakunnan suositusten sanasto</u> , viitattu 20.4.2025.                                                         |
| <b>Hävittäminen</b>              | Määräajan säilytettävän asiakirjan fyysinen (ml. tekninen) tuhoaminen. Arkistoitavan asiakirjan analogisen ilmentymän tuhoaminen.                                                                                                                                                                                                      |
| <b>Ilmentymä</b>                 | Asiakirjan digitaalinen tai analoginen ilmenemismuoto. Asiakirjalla voi olla elinkaarensa aikana useita ilmentymiä.                                                                                                                                                                                                                    |
| <b>Kuva-ala</b>                  | Kuvatiedostosta silmämääräisellä tarkastelulla näkyvä informaatio.                                                                                                                                                                                                                                                                     |
| <b>Kuvatiedosto</b>              | Digitointiprosessissa tuotettu bittikarttakuva.                                                                                                                                                                                                                                                                                        |
| <b>Käyttökappale</b>             | Käyttökappaleet ovat erilaisia käyttötarkoituksia ja käyttöliittymiä varten luotuja tiedostoja. Käytetään myös käsitettä jakelukappale. Lähde: FADGI, <u>käyttökappale</u> , viitattu 20.4.2025.                                                                                                                                       |
| <b>Palvelujen tiedonhallinta</b> | Tiedonhallinta, joka kohdistuu siihen tietoaaineistoon, joka jää tiedonhallintayksikön loogisen asiarekisterin ulkopuolelle. Lähde: <u>Tiedonhallintalautakunnan suositusten sanasto</u> , viitattu 20.4.2025.                                                                                                                         |
| <b>Säilytysmuoto</b>             | Asiakirjan digitaalinen ja/tai analoginen muoto. Asiakirjalla voi olla elinkaarensa aikana useita säilytysmuotoja.                                                                                                                                                                                                                     |
| <b>Tuhoaminen</b>                | Ks. Hävittäminen                                                                                                                                                                                                                                                                                                                       |
| <b>Väriprofiili</b>              | Väriprofiili määrittää, miten värit toistetaan tiettyssä laitteessa tai ohjelmistossa. Se tulkitsee kuvan värintoisto-ominaisuudet ja linkittää ne tiettyyn väriavaruuteen. Väriprofiili varmistaa, että värit näkyvät johdonmukaisesti ja oikein riippumatta käyttöympäristöstä.                                                      |