

Arhiivivormingute juhised asutustele

Sisukord

Sisukord	1
Taust	2
Digitaalsete arhivaalide loomise failivormingud	5
Tekstidokumentide vormingud	5
Tabelvormingud	8
Esitlusvormingud	9
E-kirjade vormingud	9
Soovitused üksikkirjade kohta	10
Soovitused e-kirjade kogumite kohta	10
Märgendfailide vormingud (JSON, XML ja HTML)	10
Pildifailide vormingud	11
Rastervormingud	11
Vektorvormingud ja CAD-vormingud	12
GIS-vormingud	13
Vektorvormingud	13
Rastervormingud	13
Filmi- ja videovormingud	14
Digitaalsete videodokumentide vormingud	14
Professionaalse digitaalse filmi (Digital cinema) vormingud	14
Filmide ja videote kasutuskooptide vormingud	16
Helivormingud	16
Üldsoovitused – digiallkirjastamine, pakkimine ja kaustastruktuur	16
Pakkimine, konteinerfailid	16
Digiallkirjastamine	17

Taust

Käesolev juhise on rakendatav eelkõige arhivaalidele ega puuduta arhiiviväärtuseta dokumente. Juhises ei käsitleta veebi ega sotsiaalmeedia arhiveerimise vorminguid, nende asjus palume pöörduda Rahvusraamatukogu poole. Samuti ei ole siin kajastatud analoogdokumentide digiteerimisel loodavate failide vorminguid. Digiteerimissoovituste kohta on loomisel eraldi juhise. Siin dokumendis toodud soovitused kehtivad ka veebikeskkondades (nt GoogleDocs, Prezi) loodud dokumentidele, sest dokumendi vormistamiseks ja arhiveerimiseks tuleb see siiski kindlas failivormingus alla laadida. Siin juhises kirjeldamata vorminguliikide asjus on Rahvusarhiiv valmis tegema asutustega koostööd, et välja selgitada parim võimalik lahendus.

Digitaalarhiivindus on riskihaldus, mis peab tagama failide kasutatavuse ajas muutunud tehnoloogilistes ja ühiskondlikes tingimustes. Rahvusarhiiv püüab maandada digiarhivaalide kasutuskõlbmatuks muutumise riski. Üks võimalus sellist riski maandada on piirduda digiarhivaalide säilitamisel ainult elujõulisemate failivormingutega.

Pikaajaliseks säilitamiseks eelistame selliseid vorminguid, mis on:¹

- kasutajate jaoks olulisi dokumendi omadusi säilitavad (näiteid leiab allpool);
- lihtsad, robustsed, et
 - kasutustarkvara kirjutamisega saaks eriolukorras hakkama ka Rahvusarhiiv,
 - eri tarkvaratootjad tõlgendaksid vormingu aluseks olevat spetsifikatsiooni ühtviisi,
 - fail oleks veakindel ja taluks bitiriknemisi²,
 - faili alamfailid (*embedded files*) või seosed väliste failidega oleksid hõlpsasti kontrollitavad ja säilitatavad,
 - spetsifikatsiooni versioon oleks stabiilne, väljakujunenud ja elujõuline ka lähitulevikus, võimaluse korral tagasiühilduv ning arvestaks vormingute vanemate versioonidega;
- tehniliselt ja õiguslikult avatud – krüpteerimata, pakkimata, DRM (*Digital Rights Management*)-kaitseta, litsentsipiiranguta;
- matemaatiliselt (või vähemalt inimtajutavalt³) ilma kadudeta;
- avatud spetsifikatsiooniga, et vormingu kirjeldus oleks avalik, litsentsi- ja kommertspiiranguteta ning eriolukorras õnnestuks spetsifikatsiooni abil faili lugeda või kasutustarkvara kirjutada;
- laia tarkvaratoega:
 - soovitatavalt avatud lähtekoodiga ja/või vabavaralise tarkvara toega,
 - faili loomise hetkel avatav mitme kasutustarkvaraga, mis tagab vähemalt mõne tarkvara eksisteerimise ka kaugemas tulevikus,

¹ Võrdluseks

- [Harvardi ülikooli kriteeriumid](#) (09.10.2020).
- [USA kongressi raamatukogu kriteeriumid](#) (09.10.2020).
- [DPC \(Digital Preservation Coalition\) kriteeriumid](#) (09.10.2020).
- [KOST kriteeriumid](#) (09.10.2020).

² Robustsuse erinevuse näited: ühe biti moondumisega TIFF-failis moondub üks rida piksleid, samal ajal kui JPG-failis muutub kogu pilt loetamatuks (<http://www.dlib.org/dlib/july08/buonora/07buonora.html> [09.10.2020]).

TXT-failis võib ühe biti moondumine kaasa tuua ühe tähe moonutuse, XML-failis (NB! ka ODT- ja DOCX-failis) kaob elemendi nime moondumisel XML-elementis sisalduv (halvemal juhul ka kõigi järgnevate elementide) info masinloetavus. Kui see tag osutub keskmisest olulisemaks, kogu dokumenti kirjeldavaks elemendiks, võib dokument (näiteks MS Office'i dokument) loetamatuks muutuda.

³ Matemaatiliselt kadudeta vorminguid on tehniliselt lihtsam kontrollida ja tagada dokumendi kvaliteedi säilimine ajas. Matemaatiliselt kadudega (*lossy compression*), kuid inimtajutavalt kadudeta pakitud vormingute puhul võib kvaliteet sõltuda konkreetse dokumendi sisust. Näiteks JP2-e saab pakkida nii matemaatiliselt kadudeta kui ka kadudega, vt selleteemalist uurimust <http://www.digitizationguidelines.gov/still-image/documents/JP2LossyCompression.pdf> (09.10.2020).

- eksisteerib tarkvara, millega tuvastada metaandmeid ja vormingut, kontrollida vormingu kvaliteeti ja spetsifikatsioonile vastavust, konverteerida teise vormingusse;
- laia kasutusega ja teadmusbaasiga:
 - mitmekesisus nii lõppkasutajate, digitaalarhiivide kui ka tarkvaratootjate poolest,
 - palju tarkvaratöe olemasolust huvitatud aktiivseid kasutajaid. See tagab olukorra, kus vajaduse korral luuakse kasutustarkvara ühiste jõududega või rahastatakse selle loomist ja arendamist,
 - eelnevale tuginedes ei saa automaatselt eelistada uusimaid vormingustandardeid;
- soodsate säilitamiskuludega:
 - failid on mahult väikesed,
 - säilitustarkvarad on eelistatult vabavaralised,
 - säilitusprotsess on lihtne.

Olulised omadused

Dokumentide **olulised omadused** on omadused, mis **peavad** ajas säilima selleks, et dokument oleks väärtuslik ka tulevikus (kasutatav ja kindlas kontekstis). Olulised omadused sõltuvad eelkõige dokumendiliigist⁴, kuid üldistatult on need järgmised:

- **autentsus, usaldusväärsus** (ehtsus, algupärasus) – selle omaduse säilimist saab tagada tehniliselt (digiallkiri, digitempel, kontrollsumma), metaandmetega (logid, EXIF⁵), protseduuriliselt (reeglid, dokumendihalduse või digisäilituse menetlusprotsess, dokumenteeritus ja auditid) jne;
- **sisu, terviklus, kvaliteet** – informatsioon ei kao ega moundu dokumendi elutsükli jooksul, vähemalt mitte sellisel määral, mis takistaks dokumendi sisu tajumist samal viisil, nagu autor seda kavandas. Eristada tuleb visuaalselt mittemärgatavaid ja rangemaid tehnilisi kadusid. Visuaalsetest kontrollidest lihtsam ja objektiivsem on kontrollida tehnilisi kadusid (näiteks kontrollsumмага). Siia kuulub ka IT-tehniline bitisäilitus varundamise ja taaste mõistes. Sisulise väärtuse säilitamiseks võib olla vajalik vormingut värskendada – konverteerida fail vananevast vormingust uuemasse. Seda omadust tuleb silmas pidada näiteks fotode puhul, kus tuleb jälgida, et vormingu muutmisel kvaliteet säiliks;
- **kasutatavus, dokumendi käitumine, kujundus** – dokumendi kasutusviisi säilitamisel peab arvestama, kas oluline on näiteks visuaalne kujundus (PDF/A, ODT, DOCX või piisab TXT-vormingust), taaskasutatavus (ODT, DOCX, TXT) või kasutatavus laialt levinud vabavaraliste tarkvaradega;
- **kontekst, sisemine ja välimine struktuur** – dokumendi seos selle loonud asutuse funktsioonide, tegevuste, toimingute, dokumendisüsteemi ja teiste dokumentidega. Selle omaduse säilimine tagatakse peamiselt metaandmete ja identifikaatorite (sh failinimed) dokumenteerimise ja säilitamisega. Kui fail sisaldab alamfaile (*embedded files*), siis on oluline ka sisemise struktuuri säilimine;
- **väärtus** – dokumendi tähtsus eri inimrühmadele ja asutustele ühiskonna toimimise kajastamiseks, üksikute toimingute ja tehingute tõestamiseks või uurimishuvi rahuldamiseks. Praktiline väärtus seisneb dokumendi loomise tinginud tegevuse edasises tagamises. See on olulisim omadus, mis moodustub teiste loetletud omaduste summana.

4 INSPECT-projekt oluliste omaduste teemal. JISC/BL/DPC Workshop'i kokkuvõtavad raportid (video, tarkvara ja vektorpiltide omadused): <https://www.webarchive.org.uk/wayback/archive/20140615130716/http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops.aspx> (09.10.2020).
 INSPECT-projektis kirjeldatud meetodika oluliste omaduste määramiseks ja näited (teksti, e-kirja, heli ja rasterpildi omadused): https://www.kdl.kcl.ac.uk/fileadmin/documents/digifutures/materials/preservation/DF09_prsrv_knight-definingSigProperties.pdf (09.10.2020).
 Archivemata. Levinumate aineseliikide olulised omadused: https://wiki.archivemata.org/Significant_characteristics (09.10.2020).

5 Exchangeable image file format.

Olulisi omadusi saab liigitada ka muudel viisidel⁶, kuid see ei muuda selle mõiste sisu. Näiteks INSPECT projektis⁷ liigitati neid selliselt: sisu, kontekst, esitus, käitumine ja struktuur.

Eksisteerib ka juhtumeid, kus erinevate oluliste omaduste säilitamiseks on vaja luua originaalfailist mitu eri vormingus säilitusfaili. Sellistel juhtudel tuleks võimaluse korral luua dokument ühes failivormingus, millest Rahvusarhiiv saab vajaduse korral pärast dokumendi vastuvõtmist luua teistes vormingutes failid

6 Artikkel oluliste omaduste definitsioonide ja liigituste kohta:

https://www.academia.edu/36970329/Metaphors_We_Work_By_Reframing_Digital_Objects_Significant_Properties_and_the_Design_of_Digital_Preservation_Systems (09.10.2020).

7 INSPECT-projekti tutvustavad slaidid: <https://www.dpconline.org/docs/miscellaneous/events/142-presentation-wilson/file> (09.10.2020).

Digitaalsete arhivaalide loomise failivormingud

Uute dokumendifailide loomisel soovitame kasutada „Arhiivieeskirja“ lisas loetletud arhiivivorminguid. Arhiivivormingud on valitud nii, et pikaajalisel arhiveerimisel oleks võimalik tagada dokumentide oluliste omaduste säilimine. Kui dokument on juba loodud ja dokumendihaldusesse hõlmatud, on konverteerimisel raskem tagada oluliste omaduste säilimist.

Üldine soovitus on kasutada kõikjal kodeeringut UTF-8 (*Unicode*, ISO 10646), sest see on ühtselt arusaadav ja tarkvaraliselt toetatud. Samuti tuleb arhiveeritavad failid luua pakkimata (või kadudeta pakitud) ning krüpteerimata kujul.

Käesoleva peatüki alapunktides on põhjalikumalt kirjeldatud arhiivivormingute kasutusalasid ja antud soovitusi failide loomise kohta.

Failivormingud on grupeeritud kolmeks:

*** eelistatavad – vormingud, mille arhiivinduslikud eelised teiste ees on selged;

** riskantsed, aga arvestatavad – vormingud, mis on laialt kasutusel, kuid millega kaasnevad arhiivinduslikud riskid ja arhiivile üleandmise järel võib olla vajalik konverteerida failid arhiivipüsivamasse vormingusse;

* mittesoovitavad või juhtumipõhiselt otsustatavad – vormingud, mis üldjuhul ei ole sobivad dokumentide oluliste omaduste säilitamiseks, kuid nende sobivus võib selguda kasutuskontekstiga põhjalikumalt tutvudes. Samuti vormingud, mille kasutajaskond on väiksem, ja seetõttu puudub Rahvusarhiivil eelnev kompetents soovitusel andmiseks. Siia gruppi kuuluvad ka käesolevas juhises nimetatud vormingud. Sellistel juhtudel kujundab Rahvusarhiiv seisukoha, lähtudes konkreetsest juhtumist ja konsulteerides vastava valdkonna asjatundjatega.

Tekstidokumentide vormingud

*** Eelistatavad vormingud:

- TXT – lihttekst
- ODT – *Open Document Format Text*
- DOCX – *MS Office Open Format XML Document*
- PDF/A-2a – *Portable Document Format*

** Riskantsed vormingud:

- PDF/A-1a, PDF/A-3a

* Mittesoovitavad või juhtumipõhiselt otsustatavad vormingud:

- PDF (üldstandard, mitte PDF/A)
- RTF – *Rich Text Format*
- DOC – *MS Word Document binary*

Lihtsaim tekstivorming on **Plain TXT**, mis on eelistatuim juhtudel, kui ei ole vaja säilitada teksti kujundust ja struktuuri. Selle vormingu eelis seisneb lihtsuses, avatuses ja laias tarkvaratoes.

Kujundusega tekstifailid on soovitatav luua kas **ODT**- või **DOCX**-vormingus (loodav näiteks OpenOffice'i, LibreOffice'i, StarOffice'i või uuemate MS Office Wordi tarkvaradega [alates 2010]). Vormingute ODT ja DOCX eelisteks on taaskasutatavus uute failide loomisel ja seetõttu nende konverteerimine PDF/A-vormingusse ei ole mõistlik. Kui PDF/A-vorming on mingil põhjusel vältimatu, sobib alamvorming **PDF/A-2a**, mis toetab PDF/A maksimumvõimalusi.

Lisasoovitused

- Kasutatud tekstifondid tuleb integreerida faili (või vältida vähe levinud fonte), vastasel korral võib fail osutada loetamatuks. Vormingute ODT ja DOCX puhul peab seda jälgima dokumendi looja, PDF/A puhul toimub nõuetele vastavuse kontroll enamasti PDFi eksportimise või konverteerimise tarkvara abil automaatselt.
- Tuleb arvestada, et dokumendis kasutatud väliseid viiteid ei ole enamasti võimalik arhiveerida. Lõika-ja-kleebi teel veebilehitsejast ODT-faili lisatud pildid võivad olla näiteks ODT-failis lingina originaalile, mitte originaalist sõltumatu ODT-elementina.
- Teksti lisatud linkide puhul tuleb arvestada nende võimalikku aegumist.
- Tekstidokumendis kasutatud mittetekstilised alamfailid (*embedded* - pildid, helid, videod) peaksid omakorda olema arhiivivormingus, vastasel korral võivad need tulevikus loetamatuks osutada. Näiteks on soovitatav luua joonised pildifaili arhiivivormingus (vt ptk „Pildifailide vormingud“).
- Vältida tuleks raskesti renderdatavaid omadusi (nt tekstidokumenti lisatud pool-läbipaistva või gradientse taustaga vektorgraafilised joonise elemendid).
- Dokumendi lisad (*appendixes*) on soovitatav luua eraldi, kuid samuti arhiivivormingus failidena. Dokumendi sees sisalduvad teised dokumendid ja lisad ei ole arhiveeritavad (nt DOCXi sisse lisatud teised DOCX- või XLSX-failid).
- Dokumendi sisu ei tohi olla krüpteeritud.
- Dokumendis sisalduvad makrod ei ole arhiveeritavad.
- Digiallkirjastatava või mitmeosalise dokumendi puhul võtta arvesse ka peatükki „Üldsoovitused – digiallkirjastamine, pakkimine ja kaustastruktuur“.

Riskid PDF/A kasutamisel

- Failis sisalduvad fotod võivad PDFiks konverteerimisel kasvatada faili mahu oodatust suuremaks. (Selle probleemi vastu aitab PDF/A-2 toetatud JPEG2000 kompressioon.)
- PDFiks ei pruugi konverteeruda jooniste läbipaistvad taustad, mille tulemuseks on kogu joonise asemel must riskülik. (Selline risk esineb PDF/A-1 korral, mitte aga kasutades PDF/A-2 vormingut.)
- Pealkirjade, sisukorra jms kujundus võib PDFiks konverteerimisel muutuda.

Rahvusarhiiv ei soovita dokumentide loomisel kasutada vorminguid RTF, DOC ega PDFi üldvormingut. Nende vormingutega kaasneb pikaajalisel säilitamisel risk, et kogu info ei pruugi olla tulevikus kuvatav autori poolt mõeldud kujul. Teised ülalloetletud vormingud on pikaajalise säilitamise jaoks sobilikud, sest põhinevad XMLil (mitte binaarvormingul) ja/või avatud spetsifikatsioonidel.

PDF, PDF/A-1, PDF/A-2, a, u, b

NB! PDF-vormingud sobivad ainult juhul, kui neisse luuakse fail dokumendi autori enda poolt. Seda tohiks teha ainult vananeva või riskantse originaalvormingu korral või kui seda nõuab asutuse asjaajamiskord. Kuna

PDF ei ole enamasti originaalvorming, peaks dokumendi autor ise veenduma, et originaalfailist PDF-faili loomisel ei kahjustuks dokumendi olulised omadused. Lähtevormingust PDF-versiooni loomisel tuleb veenduda, et konverteerumine toimuks vigadeta (nt arhiveerimistarkvara Archivemata ei ole masstöötlusena seda tagada suutnud⁸). Arhiiviväärtusega dokumentide konverteerimisel tuleb veenduda, et kogu sisu, eriti pildid, fondid ja värviruumid, oleksid kaasatud. Vajaduse korral tuleb piirata funktsioone, nagu JavaScript, integreeritud failid, heli- või videofailid⁹.

PDF või PDF/A

Tava-PDF on väga väheste piirangutega ja võib seetõttu sisaldada ka raskesti säilitatavaid elemente, mis võivad tulevikus osutada loetamatuks. PDF/A on arhiivinduse vajadustele vastavalt spetsifitseeritud PDF-standardi alamosa.

PDF/A-1 või PDF/A-2

Eelistatud on uuemal standardil põhinev **PDF/A-2**, sest edaspidised ümberkonverteerimised ühest PDFi alamvormingust teise ei pruugi õnnestuda¹⁰. Muud PDFi alaliigid, PDF/A-3¹¹, PDF/A-4¹², PDF/UA¹³, on vähe levinud ja pole näha, et need muutuks arhiivides *de facto* standardiks. Põhjalikuma ülevaate PDFi alaliikidest leiab näiteks DPC (*Digital Preservation Coalition*) raportist¹⁴.

PDF/A-2 täiendused võrreldes PDF/A-1ga¹⁵:

- PDF/A-2 lubab JPEG2000 kompressiooni, mis võimaldab sama suure failimahu juures paremat pildikvaliteeti;
- PDF/A-2 lubab läbipaistvust, PDF/A-1 mitte. Seetõttu lubab PDF/A-2 ka vesimärgistamist;
- PDF/A-2 lubab *embed*’ida teiste failide sisaldumist, mis on vastavuses standardiga PDF/A-1 või 2;
- PDF/A-2 lubab suuremaid failimahtusid (ka üle 10GB, PDF/A-1 mitte);
- PDF/A-2 lubab võimalust kihte („*optional content*“) näidata või peita nagu CAD-failides.

PDF/A-3¹⁶

See versioon on samaväärne PDF/A-2-ga, kuid lisab oma eelkäija (PDF/A-2) spetsifikatsioonile ühe ja väga olulise funktsiooni, mis annab võimaluse PDF/A-faili sisse liita mistahes vormingus alamfaile. PDF/A-2 seevastu võimaldab lisada ainult PDF/A-vormingus alamfaile.

Peamine erinevus PDF/A-1 ja PDF/A-2ga on hilisema PDF versiooni kasutamine. Seda versiooni ei saa kasutada ilma standardita ISO 32000-1¹⁷, mida see kasutab normatiivse viitena. PDF/A-3 versiooni lisavõimalused, mis vastavad standardile ISO 32000-1, on järgmised:

- Märgendatud PDF-i täiustused (juurdepääsu hõlbustamiseks)
- Tihendatud objekti ja XRefi voog (väiksema failimahu saamiseks)
- Tugi PDF/A-ühilduvate failimanuste, kaasnevate kogude ja PDF-pakettide manustamiseks
- Piltide läbipaistvuse tugi
- Piltide JPEG2000 tihendamise tugi

⁸ https://wiki.archivemata.org/Word_processing_files (09.10.2020).

⁹ <https://kost-ceco.ch/cms/pdf.html> (09.10.2020).

¹⁰ <https://ipres2017.ip/wp-content/uploads/15.pdf> (09.10.2020).

¹¹ <https://kost-ceco.ch/cms/PDF-A-3.html> (09.10.2020).

<https://www.ianus-fdz.de/it-empfehlungen/pdf-dokumente/> (09.10.2020).

¹² <https://www.pdfliib.com/pdf-knowledge-base/pdfa/the-pdfa-standards/> (09.10.2020).

¹³ <https://ipres2017.ip/wp-content/uploads/15.pdf> (09.10.2020).

¹⁴ <https://www.dpconline.org/docs/technology-watch-reports/1707-twr17-01-revised/file> (09.10.2020).

¹⁵

https://www.bar.admin.ch/dam/bar/en/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.pdf.download.pdf/archivable_file_formats.pdf (09.10.2020).

¹⁶ <https://www.loc.gov/preservation/digital/formats/fdd/fdd000360.shtml> (20.10.2020)

¹⁷ <https://www.loc.gov/preservation/digital/formats/fdd/fdd000277.shtml> (20.10.2020)

a, b või u vastavustase

- **a** (PDF/A-1a, 2a, 3a): tagab täieliku vastavuse PDF/A standardile (a = *accessible*¹⁸, *all PDF/A functions*), nõuab Unicode'i ühendamist (nagu ka allpool olev „u“ tase) ja lisaks kõigi elementide semantilist märgendamist; „**tavaline digisündinud**“.
- **b** (PDF/A-1b, 2b, 3b): lubatud on ainult PDF/A baasfunktsioonid (b = *basic*); „**ainult visuaal/väljatrükk**“¹⁹; sobilik analoogdokumentide digiteerimisel, aga ei võimalda lisada OCR²⁰ teksti.
- **u** (PDF/A-2u, 3u): nagu PDF/A-2b ja 3b, kuid lisaks mäpitakse (*mapping*) kogu tekst Unicode'i, nii et seda saab indekseerida, otsida ja muuta (u = *unicode*). PDF/A-1 seda alaliiki ei võimalda; „**otsitav**“; sobilik digiteeritud dokumentide puhul, võimaldades ka OCRitud tekstiotsingut. Samuti on see sobilik juhul, kui digisündinud dokumendi teisi omadusi (linke, interaktiivset sisukorda jms funktsioone) ei ole vaja säilitada.

Tabelvormingud

*** Eelistatavad vormingud:

- CSV – *Comma Separated Values* (lihttekst)
- ODS – *Open Document Format Spreadsheet*
- XLSX – *MS Office Open Format XML Spreadsheet*

** Riskantsed vormingud:

- XLS

* Juhtumipõhiselt otsustatavad vormingud:

- PDF
- PDF/A-2a

Kui tabelandmete puhul puudub vajadus säilitada andmetele lisaks veel funktsionaalsus, valemid või kujundus ning tabeli struktuur on lihtne, siis on eelistatuim vorming **CSV** (tekstifail, mille veergude eraldajaks on semikoolon). Keerukama struktuuriga andmete korral on soovitatav kasutada töövorminguks relatsioonilist andmebaasisüsteemi ja arhiveerimiseks sobivat vormingut SIARD.

Kui oluline on kujundus ja/või valemid, siis on soovitatav luua tabelfailid kas **ODS**- või **XLSX**-vormingus (loodav näiteks OpenOffice'i, LibreOffice'i, StarOffice'i või uuemate MS Office Exceli tarkvaradega [alates 2010]). ODS ja XLSX tabelvormingutele kehtivad samad lisasoovitused, mis on kirjeldatud peatükis „Tekstidokumentide vormingud“.

Lisasoovitused CSV kohta:

- Tekstiväljad tuleb piiritleda jutumärkidega.
- Väljade arv peab olema kõigil kirjetel ühesugune.
- Failis esimesel real peavad asuma veerunimed.

¹⁸ <https://ipres2017.jp/wp-content/uploads/15.pdf> (09.10.2020).

¹⁹ https://www.aquaforest.com/en/tiff_vs_pdf.asp (09.10.2020).

²⁰ OCR (*Optical Character Recognition*) ehk optiline tekstivastustus on trükitud või käsitsi kirjutatud teksti teisendamine masinloetavale kujule. Esmalt salvestatakse tekst skanneri või digikaamera abil pildi kujul arvutisse. Seejärel otsib tekstivastustarkvara pildilt üles tähemärgid ja teisendab need digitaalsele kujule, näiteks Unicode'i märkideks (Wikipedia).

Tabelandmete failivorminguks ei ole soovitatav kasutada PDFi ega selle alaliike, sest sellega kaob ära võimalus faili(andmeid) taaskasutada. Kui tabelvorming on kasutusel ainult teksti kujundamise ja korrastamise eesmärgil, siis tuleks lähtuda tekstidokumentidele kehtivatest soovitustest. Samas on originaalvormingud ODS või XLSX (ja ka XLS) aktsepteeritavad ilma PDFiks konverteerimata²¹.

Uute tabelfailide loomisel ei ole soovitatav kasutada XLS-vormingut. ODS- ja XLSX-vorming on pikaajalise säilitamise jaoks sobilikud, sest põhinevad avalikul, XMLil baseeruval spetsifikatsioonil, mitte binaarvormingul (XLS).

Esitlusvormingud

*** Eelistatavad vormingud:

- ODP – *Open Document Format Presentation*
- PPTX – *MS Office Open Format XML Presentation*
- PDF/A-2a

** Riskantsed vormingud:

- PPT

* Juhtumipõhiselt otsustatavad vormingud:

- PDF

Uute failide loomisel soovitame kasutada vorminguid **ODP**, **PPTX** või **PDF/A-2a** ja **PDF/A-2u**.

PDF/A vorminguks teisendamisel ei säili animatsioonid ja mitmekihilised elemendid, seega tuleks jälgida lähtefaili sisu ja võimaluse korral originaalfaili mitte konverteerida.

Siin kehtivad samad lisasoovitused, mis on loetletud peatükis „Tekstidokumentide vormingud“.

E-kirjade vormingud

*** Eelistatavad vormingud:

- EML – üksikkiri
- MBOX – kaust

** Riskantsed vormingud:

- PDF/A-2u

* Juhtumipõhiselt otsustatavad vormingud:

- MSG
- PST
- PDF

²¹ <https://kost-ceco.ch/cms/pdf-a-2-fuer-tabellenkalkulation.html> (09.10.2020).

Soovitused üksikirjade kohta

Soovitatav on dokumendihaldusesse hõlmatav (või arhiveeritav) fail luua kohealt arhiivivormingus **EML** (üksikkiri) või **MBOX** (kaust). Need vormingud on avatud standardiga ja tekstipõhised ning seetõttu on need hea veakindlusega ja tarkvaralise toega (eeldatavasti ka tulevikus). EML-failina on võimalik üksikuid e-kirju salvestada näiteks levinud meiliprogrammidega MozillaThunderbird ja MS Outlook. MBOXi perekonna alamvormingut MBOXrd võimaldab luua MozillaThunderbird.

E-kirjade kohta kehtivad sarnased lisasoovitused nagu peatükis „Tekstidokumentide vormingud“.

Juhul, kui asutuse tööprotsess või EDHSi töövoog näeb ette e-kirjade konverteerimise **PDF/A** vormingusse, siis tuleb veenduda, et kirja manuste hulgas ei oleks faile, mille olulised omadused võiksid konverteerimisel kaduma minna (vt vormingusooitusi vastavate dokumendiliikide juurest).

E-kirjade dokumendihaldusesse hõlmamise ja arhiveerimise puhul soovitage mõelda ka nende kureerimisele ja hindamisele: kuidas korrektselt reguleerida juurdepääsu delikaatsetele andmetele, automaatselt tuvastada kontonumbreid jms tundlikke andmeid, kuidas valida arhiveeritavat ainet, et mitte automaatselt kõike arhiveerimisele kaasata.

Soovitused e-kirjade kogumite kohta

Ka e-kirjade kogumite puhul on soovitatav kasutada nende dokumendihaldusesse hõlmamiseks ja arhiveerimiseks ainult eespool loetletud vorminguid. Näiteks ei ole soovitatavad Microsofti MSG ja PST, sest neist vormingutest arhiivivormingusse konverteerimine ei ole veavaba. Massarhiveerimisel on soovitatav PST-faili asemel arhiveerida otse meiliserverist üksik-e-kirjade arhiivivormingusse (nt tarkvara Mozilla Thunderbird abil MBOXi).

Märgendfailide vormingud (JSON, XML ja HTML)

Dokumendivahetuskihi DHX (endine DVK) kaudu vahetatavatele **XML** (*Extensible Markup Language*) vormingus dokumentidele on kehtestanud nõuded Riigikantselei, Majandus- ja Kommunikatsiooniministeerium²² ja Riigi Infosüsteemide Amet²³.

Lisasoovitused XMLi kohta

- XML-fail peab olema reeglipärasel kujul (*well-formed and valid*) vastavalt XML 1.0 spetsifikatsioonile²⁴.
- Soovitatav on siduda XML-dokument XML-skeemiga (*XML Schema, XSD*).
- XML-dokumendi kujunduse säilitamiseks tuleb tagada seos XML-laadilehtedega (*XML Stylesheet, XSL*), mis on vastavuses W3C soovitustega²⁵.

²² https://www.mkm.ee/sites/default/files/juhis_dokumendiliigi_xml_andmekirjelduse_koostamiseks.pdf (09.10.2020).

²³ <https://www.ria.ee/et/riigi-infosusteem/dhx/asutusele.html> (09.10.2020).

²⁴ W3C XML 1.0 standard <https://www.w3.org/TR/xml/> (09.10.2020).

²⁵ W3C XSL 1.1 standard <https://www.w3.org/TR/xsl11/> (09.10.2020).

- XMLi sisse (*embedded base64*) või lingituna XMLi kõrvale manustatud alamfailidele kehtivad vastava vorminguliigi kohta antud soovitused (vt eraldi peatükke tekstide, piltide, tabelite jne kohta).

Hüpertexti jaoks soovitame **HTML**-vormingut, kus lingitud lisafailid (pildid, helid jne) on samuti arhiivivormingus. Samuti on asjakohased **CSS** jms HTMLi loomisel kasutatavad tüüpfailivormingud. **JSON**-vorming sobib lihtsama struktuuriga andmete vahetuseks süsteemide vahel.

Märgendfailides ei ole lubatud kasutada JavaScripti koodi ega dünaamiliselt muutuvat osa.

Pildifailide vormingud

Rastervormingud

*** Eelistatavad vormingud:

- TIFF – *Tagged Image File Format* ver 6; pakkimata
- PNG – *Portable Network Graphics* ver 1.2
- JP2 – JPEG2000 standardi osa 1²⁶, kadudeta pakkimine
- DNG – *Adobe Digital Negative* ver 1.4.0.0

** Riskantsed vormingud:

- RAW (fotokaamerate toorvormingud)

* Juhtumipõhiselt otsustatavad vormingud:

- JPG, JPEG – *Joint Photographic Experts Group*
- GIF – *Graphics Interchange Format*

Digitaalselt sündinud rasterpildifailide ja fotode pikaajaliseks säilimiseks on parim valik kasutada vorminguid **JP2** ja **TIFF**. JP2 annab paremaid pakkimistulemusi fotode puhul, joonised on aga optimaalne luua PNG-vormingusse.

Digifotode toorvorminguna on soovitatav **DNG**, mis tagab detailsemate fotoparameetrite salvestumise metaandmetena. Seda vormingut toetavad rohkemad rasterpildi kasutustarkvarad võrreldes fotokaamerate tootjaspetsiifiliste RAW-vormingutega. DNG asemel võib kasutada ka vormingut **JP2**, mille eeliseks on DNGst väiksem maht. RAW-failide puhul esineb risk, et fotode arhiivile üleandmise hetkeks ei ole fail enam avatav. RAW-failide konverteerimisel DNGsse tuleb säilitada fotode metainfo. **RAW** on soovitatav vorming juhul, kui on vaja tagada rangemat tõestusväärust, autentsust või kui DNG üldvormingu metaandmed ei kajasta kõiki olulisi (RAW tootjaspetsiifilisi) metaandmeid.

Rahvusarhiiv ei soovita luua faile levinud vormingus **JPG**, sest see ei ole piisavalt veakindel²⁷ võrreldes TIFF- ja JP2-vorminguga. JPG-vorminguga kaasneb pakkimisalgoritmi tõttu suurem infokadu²⁸. Juhul, kui failid on juba loodud JPG-vormingus, siis Rahvusarhiiv ei soovita neid täiendavalt konverteerida

²⁶ JPEG2000 ISO/IES 15444-1 standard <http://www.jpeg.org/public/fcd15444-10.pdf>

²⁷ Vormingute TIFF, JP2 ja JPG veakindluse võrdlus: <http://www.dlib.org/dlib/july08/buonora/07buonora.html> (09.10.2020).

²⁸ Vormingute BMP, PNG, TIFF, JPG ja JP2 kompressioonikvaliteedi võrdlus: https://homepages.cae.wisc.edu/~ece533/project/f06/aguilera_rpt.pdf (09.10.2020).

JP2-vormingusse, sest sellega mahu kokkuhoidu ei saavutata. Samuti ei ole soovitatav luua faili kadudega pakitud JP2-vormingus, sest sellisel juhul on keerulisem tagada sisulist väärtuse säilimist²⁹. GIF-vormingu asemel soovitame kasutada PNG-vormingut, sest PNG on kaasaegsema ja efektiivsema pakkimisalgoritmiga.

Vektorvormingud ja CAD-vormingud

*** Eelistatavad vormingud:

- SVG – *Scalable vector graphics* ver 1.1, keelatud on *Java binding* ja JavaScripti kood
- DXF – *AutoCAD Drawing Interchange Format*
- DGN – *MicroStation Design*
- IFC – *Industry Foundation Classes*

** Riskantsed vormingud:

- DWG – *AutoCAD Drawing*

* Juhtumipõhiselt otsustatavad vormingud:

- STL – *3D Print Stereolithography Format* (3D-printerites kasutatav vorming)
- PDF/A-2a

Lihtsamate jooniste puhul on piisav **SVG**-vorming. SVG sobib ka veebis kuvamiseks, kuid turvariskide tõttu tuleks vältida vaikimisi toetatud Javat ja JavaScripti³⁰.

Keerukamate jooniste puhul (möödistused, plaanid ja arhitektuurijoonised) on soovitatavad CAD-vormingud. Eelistatud on tarkvara AutoCAD vorming **DXF** (avatud vorming ASCII-teksti kujul sisuga) ja **DWG** (laialt levinud, kuid suletud binaarvorming). DWG on mahult kompaktsem kui DXF. Samas on DWG-vorming riskantsem kui DXF, sest selle standardist esineb rohkem erinevaid versioone. Mõlemat eelmainitud CAD-vormingut toetab ka Eestis levinud CAD-tarkvara MicroStation³¹. Arhiivivorminguna on vastuvõetav ka MicroStationi oma vorming **DGN**, kuid sellel on mitmeid andmemahu- ja struktuuripiiranguid võrreldes *AutoCAD*-vormingutega.

Vormingut **IFC**³² (täpsemalt IFC-SPF [tekstifail], IFC-XML ja IFC-ZIP) kasutatakse ehitiste ja rajatiste info 3D/BIM³³-mudelites. Seda toetavad näiteks Majandus- ja Kommunikatsiooniministeeriumi poolt loodavad menetlussüsteemid.

²⁹ <http://www.digitizationguidelines.gov/still-image/documents/JP2LossyCompression.pdf> (09.10.2020).

³⁰ <https://kinsta.com/blog/wordpress-svg/> (09.10.2020).

<https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/svg-files-are-not-as-benign-as-it-may-seem/> (09.10.2020).

³¹ https://communities.bentley.com/products/microstation/w/microstation_wiki/8807/opening-and-saving-dwg-files-in-microstation-v8 (09.10.2020).

<https://docs.bentley.com/LiveContent/web/MicroStation%20PowerDraft%20Help-v12/en/GUID-726D3C1A-D1A3-52F6-E3F9-D749721C4F63.html> (09.10.2020).

https://communities.bentley.com/products/microstation/w/microstation_wiki/26644/which-microstation-version-supports-which-dwg-dxf-version-formats (09.10.2020).

³² https://en.wikipedia.org/wiki/Industry_Foundation_Classes#File_formats (09.10.2020).

³³ https://en.wikipedia.org/wiki/Building_information_modeling (09.10.2020).

Erijuhtudel võib kasutada trükivaate säilitamiseks vormingut PDF/A-2a³⁴, kuid andmete taaskasutuseks on see piiratum kui CAD-vormingud.

GIS-vormingud

Vektorvormingud

*** Eelistatavad vormingud:

- GML (*Geographic Markup Language* alates ver 3.0),
- OGC GeoPackage³⁵
- GeoJSON

** Riskantsed vormingud:

- SHP (ESRI Shapefile)

* Juhtumipõhiselt otsustatavad vormingud:

- LAS/LAZ (*LIDAR Laser Scanning* või *LIDAR Laser Scanning-Zipped*) – kõrgusandmete vorming.
- E57 (*LIDAR Point Cloud Data*) – 3D kaamera loodav failivorming.
- OCAD – orienteerumiskaartide vorming.

Eelistatud vorming on **GML**. Selle puhul kehtivad kõik XMLi lisasoovitused. Lihtgeomeetria (näiteks punktobjekti) WKT³⁶ atribuudina haldamiseks võib piisata **CSV** tabelvormingust.

SHP-vorming on siin juhises esitatud selle laia leviku tõttu, kuid ei ole soovitatav, sest vormingu sisu ei ole inimloetav lihttekst. Samuti esinevad piirangud andmemahule, failisuurusele ja andmekirjelduse väljanimedele jms.

Rastervormingud

*** Eelistatavad vormingud:

- GeoTIFF (TIFF+EWF.XML)
- OGC GeoPackage
- GeoPNG
- GeoJPEG2000

* Juhtumipõhiselt otsustatavad vormingud:

- ECW satelliidi- ja aerofotode puhul

³⁴

https://www.webarchive.org.uk/wayback/archive/20140616090304mp_/http://www.jisc.ac.uk/media/documents/programes/preservation/vector_images.pdf (2008) (09.10.2020).

³⁵ GeoPackage on OGC standard, mille üks eesmärk on pakkuda alternatiivi vananema hakkavale SHP-vormingule. Võimaldab samas konteineris hoida nii vektor- kui ka rasterandmeid. www.geopackage.org (09.10.2020).

³⁶ WKT (*Well Known Text*) – ruumiobjekti geomeetria esitus standardiseeritud teksti kujul, näiteks punktobjekti / pikkus- ja laiuskoordinaatide paar WGS84 referentssüsteemis.

GeoTIFF- ja **GeoJPEG2000**-vormingus arhivaalide puhul on eelistatud kompresseerimata variant.

Andmebaasis sisalduvad GIS-objektid arhiveeritakse nii üksikobjektidena (1 GML-fail = 1 objekt) kui ka kihina (1 GML-fail = 1 nähtusteklass), sest piirdudes ainult ühega neist kahest variandist, oleks nende objektide tulevane võimalik kasutus väga piiratud.

Filmi- ja videovormingud

Digitaalsete videodokumentide vormingud³⁷

- **AVI** (*Audio Video Interleaved Format*) koodek: kompresseerimata 4:2:2
- **MOV** (*QuickTime File Format*) koodek:
 - pakkimata 4:2:2
 - ProRes vähemalt 422HQ profiiliga või
 - DNxHD vähemalt 8bit 4:2:2 profiiliga
- **MP4** (MPEG-4) koodek: H.264
- **MXF** (*Material Exchange Format*) koodek:
 - MPEG-2
 - JPEG2000 (kadudeta pakkimine) või
 - DNxHD vähemalt 8bit 4:2:2 profiiliga

Professionaalse digitaalse filmi (Digital cinema) vormingud

Arhivaali tüüp	Arhiivivorming	Soovitused
Filmi pildi-master (pildikaadrite jada)	<ul style="list-style-type: none"> ● TIFF (<i>Tagged Image File Format</i>) sequence, pakkimata ● DPX (<i>Digital Picture Exchange</i>) sequence, pakkimata ● JPEG2000³⁸, kadudeta pakitud 	<ul style="list-style-type: none"> ● pildikaadrite lineaarne bitisügavus 16, 12 või 10 bitti ● pikslite proportsioon 1:1 (<i>square pixels</i>) ● resolutsioon DCI³⁹ standardi järgi ● esimene pildikaader peab olema esimene „aktiivne“ kaader, rakordkaadrid puuduvad <p><u>Värvusruumi standard (neli varianti):</u></p> <ul style="list-style-type: none"> ● <i>rec.709 inRGB full-range, optimized for display gamma of 2.4</i> ● <i>1931 CIE inXYZ, optimized for a display gamma of 2.6</i>

³⁷ Siin on mõeldud tõestusväärtusega dokumenteerimise eesmärgil loodud videoid, kus oluliseks omaduseks ei ole suurel ekraanil taasesitatavus ega kunstiline väärtus (erinevalt kinoproduksiooniks loodavatest mängu-, dokumentaal-, animajms filmidest).

³⁸ JPEG2000 ISO/IES 15444-1 standard: <http://www.jpeg.org/public/fcd15444-10.pdf>

³⁹ *Digital Cinema Initiatives*.

		<ul style="list-style-type: none"> ● DCI P3 <i>inRGB</i>, optimized for a display gamma of 2.6 ● sRGB, optimized for a display gamma of 2.2
Filmi heli-master	<ul style="list-style-type: none"> ● WAV (<i>Waveform Audio File Format</i>) ● BWF (<i>Broadcast Wave Format</i>) 	<ul style="list-style-type: none"> ● LPCM (<i>Linear Pulse Code Modulation</i>) ● resolutsioon 24 bit ● diskreetimissagedus (<i>sample rate</i>) 48 kHz või 96 kHz ● eraldi monoheli igas helikanalis ● heli peab olema pildiga sünkroniseeritud
Filmi subtiitrid	<ul style="list-style-type: none"> ● SRT (<i>SubRip Subtitle Text Format</i>) ● SUB (<i>Subtitle Text Format</i>) ● XML (<i>Extensible Markup Language</i>) 	<p>Subtiitrifailid peavad olema pildi ja heliga sünkroniseeritud.</p> <p><u>Subtiitrite standard (variandid):</u></p> <ul style="list-style-type: none"> ● CineCanvas rev C⁴⁰ või ● SMPTE 428-7⁴¹
Kinolevikoopia	<ul style="list-style-type: none"> ● DCP (<i>Digital Cinema Package</i>) – kinode projektorite jaoks ● IMF (<i>Interoperable Master Format</i>) – VOD⁴² platvormide jaoks, koodek h.264 	<ul style="list-style-type: none"> ● DCP peab olema loodud DCI standardi järgi (alates ver 1.2)⁴³ ● krüpteerimata <p><u>Subtiitrite standard:</u></p> <ul style="list-style-type: none"> ● CineCanvas rev C või ● SMPTE 428-7

⁴⁰ Subtiitrite CineCanvas standard: [http://www.dlp.com/downloads/p\[.\]CineCanvas_Rev_C.pdf](http://www.dlp.com/downloads/p[.]CineCanvas_Rev_C.pdf)

⁴¹ Subtiitrite SMPTE standard: [https://www.smpte.org/sites/de\[.\]features-20121001.pdf](https://www.smpte.org/sites/de[.]features-20121001.pdf)

⁴² *Video-on-demand*.

⁴³ <http://dcimovies.com/specification/> (09.10.2020).

Filmide ja videote kasutuskooptate vormingud

Kui filmide ja videote *master*'i kõrval on kasutuseks otstarbekas luua eraldi failiversioonid, siis Rahvusarhiiv soovib selleks järgmisi vorminguid.

Arhiivivorming	Soovitused
<ul style="list-style-type: none">● AVI (<i>Audio Video Interleaved Format</i>) koodek: Uncompressed 4:2:2● MOV (<i>QuickTime File Format</i>) koodek: ProRes● MP4 (MPEG-4) koodek: H.264● MXF (<i>Material Exchange Format</i>) koodek: MPEG-2 või J2K kadudeta kompresseeritud	<ul style="list-style-type: none">● profiil 422P@ML (<i>main profile at main level</i>)● video bitikiirus (<i>bitrate</i>) 50 Mbit/sec (CBR, <i>constant bitrate</i>)● kaadrisagedus 25 kaadrit sekundis

Helivormingud

Heliteoste ja -dokumentide loomiseks soovib Rahvusarhiiv järgmisi vorminguid.

Arhiivivorming	Soovitused
<ul style="list-style-type: none">● WAV (<i>Wave Waveform Audio</i>) koodek: <i>Linear Pulse Code Modulated Audio (LPCM)</i>● BWF (<i>Broadcast Wave ver 0, 1 & 2</i>) koodek: <i>Linear Pulse Code Modulated Audio (LPCM)</i>● FLAC (<i>Free Lossless Audio Codec ver 1.21</i>)● AIFF (<i>Audio Interchange File Format</i>)	<ul style="list-style-type: none">● resolutsioon vähemalt 16 bit● diskreetimissagedus vähemalt 48 kHz

NB! Vorming MP3 ei ole helifailide loomiseks soovitatav, sest selle pakkimisalgoritm on kadudega.

Üldsoovitused – digiallkirjastamine, pakkimine ja kaustastruktuur

Pakkimine, konteinerfailid

Arhiivinduslikult on eelistatud failide haldamine ilma pakkimiseta (lihtsaimal tehnilisel kujul ja kadudeta), sest iga lisatud keerukus muudab faili bitiriknemise suhtes haavatavamaks. Teisalt võib olla otstarbekas pakkimist kasutada erijuhtudel, kui on oluline üksikfailide asemel säilitada failikogumite terviklust ja/või failid on jaotatud kaustadesse. Failide pakkimise või mittepakkimise üle otsustamiseks soovime kindlasti konsulteerida Rahvusarhiiviga. Konteinerfailide vorminguna soovime selle laia leviku, tarkvaratoe ja kadudeta pakkimise tõttu ZIP-vormingut. Vähem levinud, kuid arhiivivorminguna vastuvõetavad on ka vormingud 7z, TAR, gzip, bzip.

Digiallkirjastamine

Siin peatükis on selgitused arhivaalide digiallkirja ja -templi failivormingute, failistruktuuri ja digiallkirja kasutamise üldiste nõuete kohta. Dokumendihalduse parimate tavade asjus tuleb lähtuda Riigi Infosüsteemi Ameti, Majandus- ja Kommunikatsiooniministeeriumi ja Riigikantselei juhistest ja normdokumentidest.⁴⁴

Digitaalselt allkirjastatud konteinerfailid luuakse DigDoc4 tarkvaraga vormingusse **ASICE**⁴⁵ (rahvusvaheliselt toetatud allkirjavorming). Ajatembeldamise rakendus TeRa⁴⁶ võimaldab luua vanadest DDOC-failidest **ASICS**⁴⁷-faile.⁴⁸

DigiDoc3 tarkvara võimaldas muu hulgas luua ka Eesti-spetsiifilises BDOC⁴⁹-vormingus faile, mida aga DigiDoc4 enam ei võimalda. Vorming DDOC⁵⁰ oli kasutusel 2014. aastani ning selles vormingus ei ole soovitatud uusi dokumente luua juba alates aastast 2015, kui võeti kasutusse turvalisemad BDOC- ja ASICE-vormingud. Dokumentide krüpteerimise vorming CDOC ei ole arhivaalide puhul lubatud ning Rahvusarhiivile üleandmisel tuleb sellised failid dekrüpteerida.

Tõestusväärtusega dokumentide puhul on oluline tagada autentsuse säilimine. See ei tähenda tingimata, et dokument peab olema digiallkirjastatud (vt allpool lõiku „Autentsuse tagamise viisid“).

Lisasoovitused

Digiallkirjastatud konteinerfail võib omakorda sisaldada teisi digiallkirjastatud faile. Tulevase arhiivikasutaja seisukohast vaadatuna on lihtsam ja mugavam, kui dokumendi sisemine failihierarhia on võimalikult lihtne. Samas peab menetluste ja üleandmise käigus säilima dokumendi ajalugu – failide seos allkirjade ja kooskõlastajatega. Digiallkirjastatud dokumentide sisuks olevat faili- ja/või kaustastruktuuri ei ole võimalik üleandmise käigus enam üle korrastada ega uuesti allkirjastada. Seetõttu kehtivad järgnevad soovitused ainult uute loodavate EDHSi-siseste või asutuse-väliste kooskõlastamisprotsesside arenduse kohta.

- EDHSi-siseseid kooskõlastamised on soovitatav luua toimingute metainfona (tegevuste logina), mitte allkirjastamisena (kui seadusaktides ei ole allkirjastamine nõutud).
- Soovitatav on vältida ebavajalikke pakkimisi (nt saab korraga allkirjastada ka mitut faili ilma neid enne üheks ZIP-failiks tegemata. Erandiks on kaustadesse jagatud failid, sest DigiDoc ei toeta kaustade allkirjastamist).
- Kui dokumendi kooskõlastajaid on mitu ja iga järgnev kooskõlastaja lisab kogumisse faile juurde, siis kehtivad eri kooskõlastused ja allkirjad erinevate failide kohta. Siinkohal on oluline selgelt eristada, milliste failide kohta millised allkirjad käivad ja milline on allkirjastamise hierarhia. Eraldi alam-allkirjafailid ühises ülem-allkirjafailis jne võivad teha selliste kogumite mõistmise lihtsamaks.
- Kui dokumenti allkirjastavad samaaegselt erinevad isikud, siis on soovitatav iga uus allkiri lisada olemasoleva allkirjafaili esimese allkirja juurde, mitte luua uue allkirja lisamiseks uut

⁴⁴ Teenuste korraldamise ja teabehalduse alused, <https://www.riigiteataja.ee/akt/131052017007> (09.10.2020)

⁴⁵ Laiendiga .ASICE ehk täpsemalt BDOC-TS ehk ASiC-E LT (ASiC-E – *Associated Signature Container Extended*) on ajatempliga (*time-stamp*) allkiri: <https://www.id.ee/?id=37026> (09.10.2020).

⁴⁶ TeRa rakendus tuleb kaasa DigiDoc4 paigalduspaketiga: <https://www.ria.ee/et/riigi-infosusteeid/digidoc-tarkvara.html> (09.10.2020).

⁴⁷ Laiendiga .ASICS ehk ASiC-S (*Associated Signature Container Simple*) allkirja- või ajatemplifail sisaldab ainult üht sisufaili: <https://github.com/open-oid/digidoc4/wiki> (09.10.2020), https://en.wikipedia.org/wiki/Associated_Signature_Containers (09.10.2020).

⁴⁸ <https://www.ria.ee/et/ametist/teated/ajatempel-ddoc-dokumentidele.html> (09.10.2020).

⁴⁹ Laiendiga .BDOC ehk täpsemalt BDOC-TM ehk ASiC-E LT-TM on ajamärgendiga (*time-mark*) allkiri: <https://www.id.ee/?id=37026> (09.10.2020).

⁵⁰ .DDOC, .BDOC ja .CDOC: <https://www.id.ee/?id=30289> (09.10.2020).

digiallkirja konteinerfaili (eelistatud on kõrvuti allkirjastajad samal tasemel nii, et uus digiallkirjastamise konteinerfail ei hõlmaks olemasolevat digiallkirjastatud konteinerfaili. Seda muidugi juhul, kui see ei ole hädavajalik, nii nagu eelmises punktis kirjeldatud juhtumil). Juba allkirjastatud dokumendile saab allkirja lisada ainult olemasoleva allkirjaga samas vormingus (ASICE).

- Kui kaustastruktuur on vajalik ainult dokumendifailide loomise etapis, siis dokumendihaldusesse hõlmamise ja/või digiallkirjastamise hetkeks on soovitatav korrastada failid kaustadest välja, et ei tekiks olukorda, kus kaustas on ainult üks fail, mis on omakorda ZIPitud.
- Kui kooskõlastamised ja digiallkirjastamised toimuvad EDHSi-väliselt, siis EDHSi hõlmamise hetkel ei ole enam võimalik digiallkirjafaili struktuuri korrastada. Küll aga saab vältida järgnevas EDHSi-siseses menetluses liigsete kaustade, ZIPide ja digiallkirjafailide loomist.
- Enne 2014. aastat digiallkirjastatud DDOC- ja BDOC1.0-vormingus dokumendid on soovitatav esimesel võimalusel uuesti digiallkirjastada, -tembeldada või ajatembeldada.
 - Lihtsaim on seda teha tarkvaraga TeRa, mis otsib arvutiga ühendatud infokandjatelt kõik vananevas allkirjavormingus failid ja ajatembeldab need krüptograafiliselt murdmiskindlamasse vormingusse ASIC-S. Selle toiminguga jäävad originaalfailid samuti alles.⁵¹
 - EDHSis või andmekogus olevate dokumentide kohta saab dokumendihaldur otsustada dokumendi kasutusaja, õigustoimingute aegumistähtaegade ja arhiiviväärtuse järgi, kas ja milliseid dokumente on vajalik üle tembeldada.

Autentsuse tagamise viisid

Dokumendi üks olulisi omadusi on autentsus (tõestusväärtus, ehtsus, usaldatavus) – dokumendi päritolu, loomise aja ning loomisjärgse muutumatuse ja terviklikkuse säilimise tõendus. ISKE definitsioon⁵²: „Andmete terviklus on andmete õigsuse, täielikkuse ja ajakohasuse tagatus ning päritolu autentsus ja volitamute muutuste puudumine.“ Autentsust saab tagada ja säilitada mitmel viisil ja sõltuvalt olukorrast saab neid ka omavahel kombineerida (vt allpool esitatud loetelu).

Loetletud viisidest kvaliteetseimalt tagab autentsuse digiallkirjastamine koos tembeldamise ja seotud protseduuridega ning transpordikihiki kaitsmisega kvalifitseeritud usaldusteenuse sertifikaatidega. Konkreetsetel juhtudel rakendatavad meetodid tuleks valida rakenduva infoturbe taseme järgi⁵³ (ISKE määrus: „Nõutav turvatase määratakse vastavalt infoturbe eesmärkidele tervikluse, konfidentsiaalsuse ja käideldavuse parameetrite kaudu.“) Loomulikult tuleb arvestada nii üldisi seadusi („Avaliku teabe seadus“, „Isikuandmete kaitse seadus“ jms) kui ka muid valdkonda, andmekogu ja dokumenteerimist puudutavaid seadusakte (vt põhjalikumalt Riigi Infosüsteemi Ameti juhendist⁵⁴).

- **Analoogkandjal dokumendi** puhul tagatakse autentsus asutuse templi ja/või isiku omakäelise allkirjaga.
- **Digiallkirjastamine ja tembeldamine kvalifitseeritud usaldusteenuse⁵⁵ osutaja vahendusel.** Dokumendi sisu fikseerimiseks allkirjastamise hetkel arvutatakse faili sisu põhjal kontrollsumma (räsi, *hash*) ja lisatakse see allkirjafaili. Peale selle lisatakse sinna usaldusteenuste abil lisaandmed isiku ja aja kohta ning registreeritakse usaldusteenuse pakkuja andmebaasis allkirjastatud dokumendi

⁵¹ <https://www.ria.ee/et/ametist/teated/ajatempel-ddoc-dokumentidele.html> (09.10.2020). Eraisikutele sobib DigiDociga kaasa tulev TeRa klient, asutustele TeRa käsurea-rakendus.

⁵² <https://www.riigiteataja.ee/akt/13125331?leiaKehtiv> (09.10.2020).

⁵³ Infosüsteemide turvameetmete süsteemi ISKE üldkirjeldus RIA lehel:

<https://www.ria.ee/et/kuberturvalisus/infosusteemide-turvameetmete-susteem-iske.html> (09.10.2020).

ISKE määrus: <https://www.riigiteataja.ee/akt/13125331?leiaKehtiv> (09.10.2020).

⁵⁴ https://www.ria.ee/sites/default/files/content-editors/publikatsioonid/andmekvaliteedi_tagamise_juhend_andmekogu_omanikele.pdf (09.10.2020).

⁵⁵ <https://www.ria.ee/et/riigi-infosusteem/usaldusteenused.html> (09.10.2020),

<https://www.riigiteataja.ee/akt/125102016001?leiaKehtiv> (09.10.2020).

kontrollsumma. Kehtivuskinnitusteenus⁵⁶ lisab .BDOC digitaalallkirjale automaatselt elektroonilise märke, mis määratleb allkirja andmise aja ja näitab, et allkirja andmiseks kasutatud isikut või asutust identifitseeriv sertifikaat⁵⁷ oli allkirjastamise hetkel kehtiv. ASiC-E-allkirjade puhul kasutatakse lisaks ka ajatempli ajatempliteenus⁵⁸. Väljastatud sertifikaatide kehtivuskinnitusi säilitatakse turvalises andmebaasis, mis võimaldab neid tõestamiseks kontrollida. Kehtivuskinnituses olev dokumendi sõnumilühend (kontrollsumma ehk räsi) ei sisalda infot allkirjastatud dokumendi sisu kohta. Sellel lahendusel on mitu alaliiki, mida saab rakendada kas eraldiseisvalt, omavahel kombineerituna või dokumendi elutsükli eri etappides.

- **Isiku digiallkiri ehk e-allkiri**⁵⁹ tõendab, et dokument pärineb selle saatnud isikult ning saadetud dokumenti ei ole pärast allkirjastamist muudetud.
- **Asutuse digitempel ehk e-tempel**⁶⁰ tõendab, et dokument pärineb selle saatnud asutuselt ning saadetud dokumenti ei ole vahepeal muudetud. E-templi saab kasutada nii koos digitaalallkirjaga kui ka ilma.
- **Ajatempel** tõendab, et teatud andmed eksisteerisid vastaval ajahetkel. Vana DDOC-faili üle tembeldades (vt eespool digiallkirjastamise ja TeRa kohta antud lisasoovitusi) kindlustatakse tehniliselt nõrgaks jääma hakkava vorminguga digiallkirja kehtimise jätkumist tehniliselt uuema vorminguga ja selles kasutatava tugevama kontrollsumma algoritmiga.
- Vormingu lahtimuukimiskindlus on seotud tõenäosusega, et ei eksisteeri arvutit, mis suudaks luua originaalfailiga samasuguse kontrollsummaga, kuid teistsuguse sisuga faili. Selline kontrollsummade juhusliku kollisiooni tõenäosus on väga väike.
- **Isikutuvastuse vahend** tagab elektrooniliste allkirjade kvaliteeditaseme⁶¹. ID-kaardi⁶², Mobiil-ID⁶³, Digi-ID (digitaalse isikutunnistuse)⁶⁴ või Smart-ID⁶⁵ abil loodud allkirja loetakse võrdseks omakäelise allkirjaga.
- **Autentsuse tagamine infosüsteemi/andmevahetuskanali turbe (ISKE) meetmetega.**
 - **Protseduurilised reeglid** (nt saab volitusi süsteemselt piirata, et dokumendi/info loomise järel ei oleks see järgmistes menetlusetappides enam muudetav).
 - **Toimingute metainfo logimine** (automaatse, kasutajast sõltumatu logimisega saab tagada toimingute läbipaistvuse ja auditeeritavuse).
 - **Süsteemile ligipääsevate isikute autentimine.** Isikud, kes dokumenti menetlevad, saab automaatselt logida või isikute protseduurilisi volitusi süsteemis reguleerida. Selleks tuleb isik tuvastada, mille rangeim viis on isikutuvastusvahendi (ID-kaart, Mobiil-ID, Digi-ID, Smart-ID) ja kvalifitseeritud usaldusteenuse osutaja (isikutuvastussertifikaadi kehtivuskinnitusteenuse)⁶⁶ rakendamine. Lihtsaim, kuid nõrgim viis on rakendada kasutajatunnuse ja parooli abil autentimist.
 - **Andmevahetuskanalite tehnilised turvaprotokollid**⁶⁷. Andmevahetuse rangemaks turvamiseks saab kasutada info saatja ja vastuvõtja autentimise sertifikaate⁶⁸.

⁵⁶ <https://www.sk.ee/teenused/kehtivuskinnituse-teenus/> (09.10.2020), <https://www.id.ee/index.php?id=30271> (09.10.2020).

⁵⁷ <https://www.sk.ee/teenused/digitaalne-identiteet> (09.10.2020).

⁵⁸ <https://www.sk.ee/teenused/ajatempliteenus/> (09.10.2020).

⁵⁹ Uue seaduse järgi kasutatakse „digiallkirja“ ja „digitempli“ asemel mõisteid „e-allkiri“ ja „e-tempel“ <https://www.riigiteataja.ee/akt/125102016001?leiaKehtiv> (09.10.2020).

⁶⁰ Digitembeldamine: <https://www.id.ee/index.php?id=30282> (09.10.2020). <https://www.sk.ee/teenused/digitempliteenus/> (09.10.2020), <https://www.id.ee/index.php?id=30282> (09.10.2020).

⁶¹ <https://www.id.ee/artikkel/digitaalne-allkirjastamine-ja-elektroonilised-allkirjad/> (09.10.2020).

⁶² <https://www.sk.ee/teenused/digitaalne-identiteet/id-kaart> (09.10.2020).

⁶³ <https://www.sk.ee/teenused/digitaalne-identiteet/mobiil-id/> (09.10.2020).

⁶⁴ <https://www.sk.ee/teenused/digitaalne-identiteet/digi-id/> (09.10.2020).

⁶⁵ Smart-ID puhul peab omakäelise allkirjaga võrdväärse allkirja andmiseks olema konto tasemeks „Smart-ID kvalifitseeritud elektrooniline allkiri“, vt lähemalt: <https://www.smart-id.com/et/abi/kkk/uldine/kas-smart-id-sobib-elektrooniliste-dokumentide-allkirjastamiseks/> (09.10.2020).

⁶⁶ <https://www.sk.ee/teenused/kehtivuskinnituse-teenus/> (09.10.2020).

⁶⁷ https://et.wikipedia.org/wiki/Transpordikihi_turbeprotokoll (09.10.2020).

⁶⁸ <https://www.sk.ee/teenused/autentsertifikaat/> (09.10.2020).

- **Autentsuse tagamine välisriikidega suheldes.**
 - **(E-)Japostillimine**⁶⁹. Notari poolt väljastatava tunnistuse abil dokumendi allkirjastaja pädevuse tõendamine.
 - **Dokumendi legaliseerimine**⁷⁰. Välisministeeriumi poolt väljastatava tunnistuse abil dokumendi allkirjastaja pädevuse tõendamine.

⁶⁹ <https://www.notar.ee/19886>

⁷⁰ <https://vm.ee/et/avaliku-dokumendi-legaliseerimine> (09.10.2020), <https://www.notar.ee/14254>