

DIGITEERIMISE JUHIS ASUTUSTELE

Sisukord

1.	Üldpõhimõtted	1
2.	Digiteerimisele eelnevad tegevused.....	1
3.	Failide nimetamine	2
3.1	Failide nimetamine arhiiviainese massdigiteerimisel	3
3.2	Failide nimetamine dokumendihaldusesse hõlmatava ainese digiteerimisel	3
4.	Metaandmed	4
4.1	Tehnilised metaandmed	4
4.2	Kirjeldavad metaandmed	4
5.	Tekstidokumentide digiteerimine	5
5.1	Digiteerimise üldised nõuded	5
5.2	Kasutatavad failivormingud	5
6.	Fotode digiteerimine	6
6.1	Üldnõuded.....	6
6.2	Kasutatavad failivormingud	7
7.	Digiteerimise kvaliteedi kontroll.....	7
7.1	Manuaalselt tehtavad kontrollid	8
7.2	Automatiseeritavad kontrollid	8
	Lisamaterjalid	9

1. Üldpõhimõtted

Selles juhises esitatud nõuded on kohustuslikud, kui digiteeritakse arhiiviväärtuslikke dokumente ehk arhivaale. Kuna allolevad nõudeid on kirjeldatud parimate praktikate eeskujul, siis soovime neid järgida ka teistel asutustel mitte-arhiiviväärtuslike dokumentide digiteerimisel. Seda eriti juhul, kui asutus plaanib pärast digiteerimist analoogkandjal teabe hävitada.

Juhises käsitletakse ainult teksti- ja fotodokumente. Teiste analoogkandjatel dokumentide digiteerimise põhimõtted leiab juhise lõpus viidatud lisamaterjalidest. Audio- ja videoainese digiteerimisel tuleb lähtuda Rahvusarhiivi digiteerimise standardis¹ olevatest nõuetest.

Digiteeritud arhivaalid võetakse Rahvusarhiivi vastu juhul, kui digiteerimise protsess ja digiteerimisel saadud fail(id) vastavad selles juhises esitatud nõuetele. Juhul, kui asutus soovib pärast arhivaalide digiteerimist analoogkandjal originaalid hävitada, tuleb see vastavalt arhiivieeskirja § 13 lõikele 3 kooskõlastada Rahvusarhiiviga.

Kui digiteerimise tulemusena tekkinud failid on plaanis üle anda Rahvusarhiivile, tuleb enne digiteerima asumist Rahvusarhiiviga kokku leppida, kuidas arhivaalid asutuses korrastakse ja mil viisil need üle antakse (nt eksport elektroonilisest dokumendihaldussüsteemist [EDHS], korrastamine universaalses arhiveerimismoodulis [UAM]). Arhivaalide üleandmise kohta on samm-sammuline juhised nähtav RA kodulehel².

2. Digiteerimisele eelnevad tegevused

Enne digiteerima asumist peab asutus välja selgitama kellele on digiteerimise tulemus suunatud (nt digiteerimine enda tarbeks, avalikku arhiivi edastamiseks jne) ja mis on digiteerimise eesmärgiks (nt hübriidarhiivi ühtlustamine, paberdokumentide hävitamine jne). Samas peab asutus analüüsima ka digiteeritavate dokumentide seisukorda, nende eripärasid ning soetama (ostma või rentima) ja seadistama vajaliku riist- ja tarkvara. Lühidalt öeldes on digiteerimist mõistlik planeerida siis, kui sellest tuleneb mingi kindel kasu või asutusel on konkreetne vajadus kindla eesmärgi saavutamiseks.

Täpsemalt peab asutus:

1. Läbi mõtlema, kas eesmärgiks on tagada jooksvalt, üksikult saabuvate ning asutusse pöörduvate kasutajate poolt tellitud dokumentide digiteerimine või kindla dokumentide sarja/koguse massdigiteerimine.
2. Välja selgitama, kas eesmärgiks on hübriidse sarja/toimiku ühtlustamine ja hõlmamine EDHSi.
3. Läbi mõtlema, kus hoiustatakse digiteeritud dokumendid (nt kas on piisavalt kettamahtu) ja kuidas tagada neile juurdepääs (nt millistesse süsteemidesse on vaja lisada failid või nende viited).

¹ Rahvusarhiivi digiteerimise standard, https://www.ra.ee/wp-content/uploads/2020/03/Digistandard_2019-2.pdf

² Arhivaalide üleandmine, <https://www.ra.ee/arhiivihaldus/juhised/#6>

4. Analüüsima, milline on digiteeritavate dokumentide tüüp (nt. tekst, foto, joonis, kaart), milline on dokumentide konstruktsioon (köites või lahtised lehed), milliste mõõtmetega on digiteeritav materjal (nt suureformaadilised lehed) ning milline on füüsiline seisund.
5. Eelnevate sammude põhjal otsustada, kuidas digiteerimisprotsess läbi viia sh. valida sobivad failivormingud, hankida sobiv riistvara ning sellega ühilduv sobiv tarkvara.

Lähtekohad riistvara valimiseks

- tasa- ehk lameskanner (kui soovid digiteerida fotosid või üksiklehti),
 - planetaar- ehk raamatuskanner (köites ja üksiklehtede digiteerimiseks),
 - laiformaatskanner (suureformaadiliste dokumentide digiteerimiseks),
 - sööturiga dokumendiskanner (mitmiklehtede, sh kartoteegikaartide digiteerimiseks),
 - DSLR- ehk peegelkaamera (vastavalt vajadusele, sh kaartide digiteerimiseks)
6. Kalibreerida riistvara (skanner või fotokaamera) ning seadistada tarkvara selliselt, et nõutavad tehnilised metaandmed (vt pt 4.1) lisanduksid automaatselt skaneerimisel tekkinud failile).
 7. Määrata sobiv skanneri väljundvorming. Vormingu valimisel tuleb lähtuda allpool käsitletud nõutavatest failivormingutest peatükkidest (pt 5.2 ja pt 6.2) ning skanneri ja/või selle tarkvara võimalustest.
 8. Kavandada ja seadistada failide nimetamine. Täpsemad juhised on antud peatükis 3.
 9. Dokumenteerida digiteerimis-eelsed otsused ja riistvara ning tarkvara seadistused. (Kirjeldada lühidalt dokumenteerimise vajadus. Konkreetseid juhiseid dokumenteerimiseks ei ole aga võimalik kirja panna vabas vormis.)

NB! Kõik digiteerimiseks kasutatavad seadmed ja tarkvaralised lahendused peavad tagama käesolevas juhises esitatud nõuete täitmise täielikult ja kvaliteetselt.

3. Failide nimetamine

Järjepideva ja kindla failinime formaadi kasutamine aitab tagada failide leidmist, avamist ning säilitamist nii kaua, kui vaja. Informatiivne failinimi võimaldab faili hiljem hõlpsamini üles leida ning korrektne nimeformaad ning nimetamisega seotud põhireeglite järgimine aitab vältida ühilduvusprobleeme, mis võivad tekkida failide liikumisel ühest infosüsteemist teise (nt dokumendihaldussüsteemide versioonide uuendamisel või süsteemide-vahelisel migreerimisel).

Nimetamise juures on oluline tagada, et:

- 1) failinime arusaadavus ja mõistetavus (dokumendi pealkiri, nimi jne) süsteemide ja/või inimeste poolt (FNS, muu viide või identifikaator, EDHSi viide);
- 2) failide terviklik liikumine läbi erinevate failisüsteemide. Selle kindlustamiseks tuleb vältida allpool välja toodud keelatud tähemärke.

Failinime koostamise üldnõuded on järgmised:

- Failinimes ja laiendis on soovitatav kasutada ainult väiketähti.
- Failinime pikkus (laiendita) ei tohiks ületada 30 tähemärki.
- Kasutada ei tohiks täpitähti (õ, ä, ö, ü) ega järgmisi tähemärke: \ / : * ? " < > | [] & \$, .
- Vältida tühikute kasutamist. Tühiku asendajana kasutada tähemärkide eraldamiseks alakriipsu (_) või sidekriipsu (-) ning sõnade eristamiseks või rõhutamiseks kasutada vajadusel suurtähti.
- Kuupäeva kirjeldamisel failinimes tuleks lähtuda standardist ISO 8601:2004³. Sellest tulenevalt on soovitatav kirjutada kuupäev formaadis „YYYYMMDD“, ehk siis eesti keelde pannes „AAAAKKPP“.

3.1 Failide nimetamine arhiiviainese massdigiteerimisel

Massdigiteerimisel tuleks failide nimetamine automatiseerida ehk seadistada see juba digiteerimistarkvaras. Selleks tuleb:

- Eelnevalt kavandada ja seadistada nii sisestatavad kui ka automaatselt genereeritavad failinime osad. Seda tasuks teha nii massdigiteerimise kui ka üksikdokumentide digiteerimise korral (nt EDHSi kaudu), kuna sellega hoitakse ära vigade tekkimine failinimes.
- Luua seosed digiteeritud dokumendi ja originaaldokumendi vahel kasutades dokumendi viita (funktsiooni, sarja, toimiku ja dokumendi numbrit). Lisaks on võimalik kasutada ka selle juhise lisamaterjalides näidatud Exceli tabelit⁴. Juhul, kui digiteeritud dokumendid hõlmatakse elektroonilisse dokumendihaldussüsteemi, tuleb luua ka seosed EDHSis.
- Välja selgitada, kas on tegemist arhiiviväärtuslike dokumentide digiteerimisega. Kui on, tuleb failinimed moodustada Rahvusarhiivist saadud fond-nimistu-säilik (FNS) koodi alusel, tuginedes Rahvusarhiivi digiteerimise standardis⁵ esitatud failide nimetamise reeglitele.

NB! Eelolevat nõuet ei pea järgima, kui digiteeritakse asutuse enda vajadusest lähtuvalt. Sel juhul otsustab asutus ise, millise loogika alusel toimub failinimede kirjeldamine, tuginedes failide nimetamise üldistele nõuetele.

3.2 Failide nimetamine dokumendihaldusesse hõlmatava ainese digiteerimisel

Failide nimetamisel tuleb asutusel lähtuda oma üldisest dokumentide haldamise korrast. Kui failide nimetamine ei ole reguleeritud, on soovitatav need luua allolevate nõuete alusel.

- Failinimi peab olema lihtne ja arusaadav ning edastama viitamisinfot dokumendi kohta, milleks on dokumendi viit. Lisaks võib failinimesse kirjeldada ka dokumendi pealkirja, sarja ja kuupäeva. Vajaduse korral võib lisada ka digiteeritud dokumendi osa identifitseerivad

³ ISO 8601:2004, <https://www.iso.org/standard/40874.html> (08.08.2021)

⁴ [MS Exceli tabel](#)

⁵ Rahvusarhiivi digiteerimise standard, https://www.ra.ee/wp-content/uploads/2020/03/Digistandard_2019-2.pdf

metaandmed (nt Lisa1, Joonis64, foto4, lehekülg jne). Kindlasti tuleb veenduda, et faili nimes oleksid seosed originaaldokumendi kirjeldusega EDHSis või mõnes teises dokumendihaldussüsteemis.

- Failinimes ei ole soovitatav kasutada lühendeid. Kui see on vältimatu, tuleb lühendid dokumenteerida (vt Digiteerimiseelsed tegevused).
- Soovitatav on koostada failinimi nii, et see eristuks teistest samanimelistest failidest (nt „kaskkiri.pdf“).

4. Metaandmed

Erinevad metaandmete kategooriad täidavad digiteeritud dokumentide juures erinevaid funktsioone. Pikaajalise digitaalse säilitamise seisukohast on digiteerimisel oluline kirjeldada kindlad, allpool ära toodud metaandmete tüübid.

4.1 Tehnilised metaandmed

Digiteerimisel kasutatav riistvara (skanner või fotokaamera) lisab reeglina tehnilised metaandmed automaatselt faili päisesse.

Tehnilised metaandmed lisatakse failile vastavalt digiteerimiseks kasutatavas tarkvaras või väljundvormingus toetatud standardile. Erinevates vormingutes on kasutusel erinevad standardid (nt EXIF standard on kasutusel vormingutes TIFF ja JPEG, standard XMP on aga kasutusel PDF vormingus), seepärast tuleb ka metaandmete kontrollimisel vaadata vastava standardi osa. Seda peab jälgima, kui kasutatakse tarkvaralist tööriista, mis vaikimisi kuvab erinevate standardite infot. Metaandmete olemasolu tuleb kontrollida vahetult pärast digiteerimise lõppu (vt pt 7.1 Manuaalselt tehtavad kontrollid).

Nõutavad on vähemalt järgmised tehnilised metaandmed:

- faili looja / digiteerija (autor);
- andmed tarkvara kohta (nimetus, versioon, tootja), millega fail on loodud või seda on töödeldud;
- andmed seadme kohta, millega on fail loodud (skanneri või kaamera tootja, mark, mudel).

PS! Metaandmeid võib talletada ja hallata eraldiseisvas infosüsteemis. Juhul, kui asutusel on vajadus tehnilised metaandmeid infosüsteemi kanda, peab olema infosüsteem vastavalt seadistatud (nt tehniliste metaandmete lugemise faili päisest).

4.2 Kirjeldavad metaandmed

Dokumente kirjeldavad metaandmed tuleb luua vastavalt asutusesisese töökorralduse reeglitele,

„Teenuste korraldamise ja teabealduse alustele“⁶ või Rahvusarhiivi arhiivikirjelduse elementide loendile. Viimane on kättesaadav juhise lisamaterjalides (Exceli tabelfail⁷).

5. Tekstidokumentide digiteerimine

Tekstidokumentide digiteerimisel peab olema tagatud digiteeritud kujutisel oleva teksti loetavus ning kujunduselementide selgus ja arusaadavus.

5.1 Digiteerimise üldised nõuded

Selleks, et oleks garanteeritud parim võimalik kujutise kvaliteet, tuleb aluseks võtta allolevad nõuded:

- Dokument digiteerida originaalsuuruses või resolutsiooniga vähemalt 300 dpi.
- Kui dokument sisaldab detailset infot (näiteks väikest teksti), võib resolutsiooni suurendada.
- Värviline tekstidokument digiteeritakse värviliselt (24 bit). Tekstidokument, mis ei sisalda värvidega edasiantavat infot (nt. on täies mahus mustvalge), võib digiteerida ka halltoonides (8 bit).
- Dokumendi leheküljed, mis sisaldavad teksti või muud infot (sh lehekülje number), digiteeritakse terviklikult. Tühje lehekülgi ei digiteerita.
- Digiteerimisel tuleb dokument asetada seadmesse nii, et hiljem ei peaks faile töötlema. Faile on lubatud töödelda vaid viisil, mis ei moonuta digiteeritud dokumenti: pöörata kujutist 90 või 180 kraadi ning lõigata ära servi, mis ei sisalda digiteeritud kujutist. Keelatud on tarkvara abil parandada kujutise kvaliteeti.

5.2 Kasutatavad failivormingud

Tekstidokumentide digiteerimisel soovitame kasutada alljärgnevat failivorminguid.

Vorming	Millal sobib kasutada?	Täpsustav info
PDF/A	Dokumendi tekst peab olema otsitav, seega on vajadus rakendada OCRi.	Parim võimalik vorming on PDF/A-2u . Kui skanneri tarkvara seda versiooni ei toeta, on lubatud ka teised PDF/A versioonid (järjestus paremuse järgi: PDF/A-2b, PDF/A-1u, PDF/A-1b). NB! Ei ole soovitatav kasutada vorminguid PDF/A-1a või PDF/A-2a , sest need on eelkõige mõeldud „digitaalselt sündinud“ failide vorminguks.
TIFF (kompresserimata)	Juhul, kui tekstidokument on üheleheline ega vaja	Kui dokument on köidetud ja sisaldab fotosid (nt fotoalbum), siis on soovitatav luua lisaks ka

⁶ TKTA, <https://www.riigiteataja.ee/akt/131052017007>

⁷ MS Exceli tabel

	OCRi ehk teksti otsitavust.	tervikvaade ning kasutada selleks vormingut PDF/A. Igast fotost eraldi tuleb luua ka täiendav TIFF.
TXT (lihtvorming)	Kui on oluline ainult teksti sisu ja hilisem otsitavus	Kuna fail tekib OCRi rakendamisel, siis tuleb veenduda, et tulemus oleks vastavuses originaaliga. Kindlasti tuleks testida ka otsingut üle teksti.

6. Fotode digiteerimine

Fotode digiteerimisel tuleb eelkõige lähtuda fotomaterjali tüübist, omadustest, vormistusest ja seisukorrast. Nendest sõltuvalt tuleb valida digiteerimise viisiks kas skaneerimine või pildistamine fotokaameraga.

6.1 Üldnõuded

Alljärgnevalt on ülesse loetud üldised nõuded, mida tuleb fotode digiteerimisel arvesse võtta.

- Kõik fotod (ka mustvalged) digiteeritakse värviliselt (24 bit), RGB värviruumis.
- Paberfotodel digiteeritakse ka pööre, kui sellel on informatsiooni foto sisu kohta.
- Kui fotod asuvad albumis, siis tervikvaate säilitamiseks digiteeritakse terve albumi lehekülg. Albumi lehel olevad fotod digiteeritakse veelkord igaüks eraldi fotode digiteerimise nõudeid järgides.
- Kui foto on tekstidokumendi osa, siis digiteeritakse lehekülg tekstidokumendi digiteerimise nõudeid järgides ning selle lehel asuv foto eraldi veelkord fotode digiteerimise nõudeid arvestades.
- Fotode digiteerimise resolutsioon sõltub originaali mõõtmetest ning resolutsioon määratakse järgmise tabeli alusel:

Originaalsäiliku mõõdud (nii negatiivid kui ka positiivid)	Säilitusfaili skaneerimise resolutsioon
Fotod, mille pikema külje pikkus on kuni 4 cm, ja fotonegatiivid	3200 dpi
6 x 6 cm	2400 dpi
6 x 9 cm	2400 dpi
9 x 12 cm	1200 dpi
10 x 15 cm	1200 dpi
13 x 18 cm	720 dpi
18 x 24 cm	600 dpi

24 x 30 cm	600 dpi
30 cm ja suuremad	600 dpi

6.2 Kasutatavad failivormingud

Fotode digiteerimise väljundvorming sõltub eelkõige digiteerimise seadmest ja kasutatava tarkvara võimalustest. Fotode, piltide, jooniste jne digiteerimisel tuleks kasutada järgmisi vorminguid:

Vorming	Millal sobib kasutada?	Täpsustav info
TIFF (kompresserimata)	Pikaajaliseks säilitamiseks parim, kuna võimaldab hilisemat töötlemist, liigutamist ja salvestamist säilitades esialgset kvaliteeti.	Arvestama peab sellega, et faili maht võib paisuda suureks, olenevalt digiteeritava kujutise keerukusest. Seega nõuab suuremaid salvestusmeediume.
DNG	Kaameraga digiteerimisel parim soovituslik vorming juhul, kui soovitakse säilitada ka toorvormingut.	Kaamera tootja-poolsed toorvormingud (RAW) ei ole sobilikud pikaajaliseks säilitamiseks, kuna on avatavad ainult kindla tootja tarkvaraga. Kui toorvormingut säilitada ei soovita, on mõistlik digiteerimisel tekkinud fail kohealt TIFF vormingusse konverteerida.
JP2 (JPEG2000, kadudeta pakkimine).	Kuna puudub lai tarkvaraline tugi, siis on võimalus kasutada ainult juhul, kui kaamera või skanneri tarkvara seda toetab.	Sobib hästi kasutamiseks massdigiteerimisel, kuna faili suurus on oluliselt väiksemad TIF-iga võrreldes.

7. Digiteerimise kvaliteedi kontroll

Kvaliteedikontroll on protsess, mida kasutatakse digiteerimisel kvaliteedi, täpsuse ja järjepidevuse kontrollimiseks. Süsteemne kvaliteedikontroll võimaldab tuvastada faile, mis ei vasta ette antud digiteerimise standardile. Pärast vigade tuvastamist on võimalik töövoogusid jooksvalt kohandada selleks, et edaspidiseid vigasid vältida. Ilma kontrollita võivad vead jääda tähelepanuta, mille tulemusena võib olla vajalik hilisem korduv digiteerimine, kuid see võib kaasa tuua juba täiendavaid kulusid.

7.1 Manuaalselt tehtavad kontrollid

Järgnevalt on kirjeldatud tegevused, mida tuleb teha käsitsi selleks, et veenduda digiteerimise tulemuse vastavuses originaaliga. Selleks tuleb:

- veenduda, et KÕIK dokumendi lehed on digiteeritud;
- kontrollida, et lehekülgede järjestus vastaks originaalile;
- testida digiteerimisel tekkinud väljundfaili avanemist;
- olla kindel, et sisu on loetav ja arusaadav, terviklik;
- tuvastada, et kujutis ei oleks liiga hele või tume;
- veenduda, et digiteeritud koopia on visuaalsel vaatlusel originaaldokumentiga vastavuses;
- veenduda, et digiteerimise tulemina saadud fail on varustatud korrektsete tehniliste metaandmetega. Selleks võib kasutada näiteks tarkvaralisi tööriistasid ExifTool⁸ või Metadata++⁹.

7.2 Automatiseeritavad kontrollid

Automaatkontrolle tuleks digiteerimise töövoogudesse sisse planeerida eelkõige massdigiteerimise läbiviimisel. Kontrollid peaksid käivituma näiteks failide laadimisel infosüsteemi, mille korral valideerimise tööriistad võiksid olla koheselt integreeritud vastavatesse digiteerimise või faili laadimise töövoogudesse.

Massdigiteerimisel tuleb kõiki digiteerimisel saadud koopiaid originaaldokumentidega võrrelda peale digiteerimise lõppu või vastavalt kokku lepitud ajale nii, et vajadusel saab ebakvaliteetse tulemi asendada uuesti originaalset digiteeritud väljundfailiga.

Automaatselt teostatavad kontrollid on:

- valideerimine, mille korral saab veenduda, et fail vastab vormingule esitatud nõuetele (nt PDF/A vormingu valideerimine VeraPDF¹⁰ tarkvaralise tööriistaga);
- tõestusväärtuse säilitamine juhul, kui on oluline tagada digiteeritud dokumendi tõestusväärtus ka tulevikus. Sellisel juhul tuleks rakendada digitembeldamist (massdigiteerimise korral) või digitaalne allkirjastamine (üksikdokumendi digiteerimisel).

NB! Tõestusväärtuse kandmine digiteeritud koopiale kehtib ka juhtudel, kui soovitakse hiljem analoogkandjal originaal hävitada.

Lisaks ülal mainitud tegevustele tuleb üle vaadata ja veenduda:

- metaandmete olemasolus ja nende vastavus reeglitele,

⁸ <https://exiftool.org/> (15.12.2020)

⁹ <https://www.logipole.com/metadata++-en.htm> (18.01.2021)

¹⁰ <https://docs.verapdf.org/install/> (05.11.2020)

- digiteeritud kujutiste tehnilises kvaliteedis,
- objektis sisalduva teabe tervikkuses,
- failinimedede õigsuses ja nõuetele vastavuses.

Lisamaterjalid

- [Rahvusarhiivi digiteerimise standard](#)
- [Arhiivieeskiri](#) (ArhE)
- [Arhiivivormingud](#) (ArhE Lisa 1)
- [ISO Juhised dokumentide digiteerimise korraldamiseks](#) (ISO/TR 13028:2010)
- [FADGI digiteerimise juhise](#)
- [Arhiivikirjelduse elementide loetelu](#)
- [MS Exceli tabel](#) digiteeritud failide arhiivikirjelduste vormistamiseks
- [DPC käsiraamat](#) digiteerimisprojekti sammude ülevaate ja linkidega eri riikides asutustele antud digiteerimisjuhiste