

DIGITEERIMISE JUHIS ASUTUSTELE

Sisukord

1.	Üldpõhimõtted	2
2.	Digiteerimisele eelnevad tegevused.....	2
3.	Failide nimetamine	3
3.1	Failide nimetamine dokumendihaldusesse hõlmatava ainese digiteerimisel.....	3
3.2	Failide nimetamine arhiiviainese massdigiteerimisel.....	3
4.	Metaandmed	4
4.1	Tehnilised metaandmed.....	4
4.2	Kirjeldavad metaandmed	4
5.	Tekstidokumentide digiteerimine.....	4
6.	Fotode digiteerimine	5
7.	Digiteerimise kvaliteedi kontroll.....	6
	Lisamaterjalid	7

1. Üldpõhimõtted

Juhis on mõeldud arhiivimoodustajatele ja siin esitatud nõuded puudutavad eelkõige arhiiviväärtuslike dokumentide (arhivaalide) digiteerimist. Samas võib allolevaid nõudeid rakendada ka iga teine asutus oma mitte-arhiiviväärtuslike dokumentide digiteerimisel (nt 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 standardi“ nõuetest (vt Lisamaterjalid).

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 enne kooskõlastada Rahvusarhiiviga ning tugineda „Arhiivieeskirja“ nõuetele (ArhE § 13 lõige 3).

Kui digiteerimise tulemusena tekkivad failid on plaanis üle anda Rahvusarhiivile, tuleb enne digiteerima asumist Rahvusarhiiviga kokku leppida, kuidas failid asutuses korrastakse ja mil viisil need üle antakse. Korrastamise ja üleandmise võimalusteks on näiteks eksport elektroonilisest dokumendihaldussüsteemist (EDHS), universaalse arhiveerimismooduli (UAM) kasutamine või tegevused peagi valmiva ASTRA vastuvõtumooduli kaudu.

2. Digiteerimisele eelnevad tegevused

1. Välja selgitada, kas eesmärgiks on paberdokumentide viimine EDHSi dokumendihalduse osana (vältimaks hübriidseid dokumente) või toimub dokumentide digiteerimine suurte kogudena ehk massdigiteerimine.
2. Analüüsida, milline on digiteeritavate dokumentide tüüp (tekst, foto), konstruktsioon (köites või lahtised lehed, suureformaadilised lehed) ja füüsiline seisund.
1. Eelnevate sammude põhjal otsustada, millist tüüpi skannerit on vaja kasutada või milliseid ettevalmistavaid töid tuleb teostada (nt lehtede köitest lahti võtmine).
3. Digiteerimiseks kasutatavad seadmed ja tarkvara peavad tagama käesolevas juhises esitatud tehniliste nõuete täieliku ja kvaliteetse täitmise.

Järgnev loetelu annab lähtekohad skanneri valimiseks ja seadistamiseks:

- tasa- ehk lameskanner (fotode ja üksiklehtede digiteerimiseks),
 - planetaar- ehk raamatuskanner (köites ja üksiklehtede digiteerimiseks),
 - laiformaatskanner (suureformaadiliste dokumentide digiteerimiseks),
 - sööturiga skanner (mitmelehelise dokumendi, sh kartoteegikaartide digiteerimiseks),
 - DSLR- ehk digitaalne peegelkaamera (vastavalt vajadusele, sh kaartide digiteerimiseks).
4. Kalibreerida skanner (või fotokaamera) ning seadistada tarkvara selliselt, et tehnilised metaandmed lisatakse automaatselt skaneeritavale dokumendile.

5. Määrata sobiv skanneri väljundvorming. Vormingu valimisel tuleb lähtuda allpool nimetatud failivormingutest ning skanneri ja/või selle tarkvara võimalustest.
6. Kavandada failide nimetamine vastavalt sellele, kas tegemist on massdigiteerimisega või hübriidse ainese viimisega digitaalsesse dokumendihaldusesse. Täpsemad juhised on antud vastavates alampunktides.
7. Dokumenteerida digiteerimisprotsess.

3. Failide nimetamine

Korrektne ja informatiivne failinimi võimaldab faili hiljem hõlpsamini ülesse leida ning aitab vältida ühilduvusprobleeme, mis võivad tekkida ühest infosüsteemist teise liikumisel.

3.1 Failide nimetamine dokumendihaldusesse hõlmatava ainese digiteerimisel

- Failide nimetamisel tuleb asutusel lähtuda oma üldisest dokumentide haldamise korrast. Kui failide nimetamine ei ole reguleeritud, on soovitatav see enne digiteerimist fikseerida ja dokumenteerida.
- Failinimi peab olema lihtne ja arusaadav ning edastama viitamisinfort dokumendi kohta (pealkiri, sari, kuupäev [aaaakkpp] vms). Vajaduse korral tuleb nimes kirjeldada ka digiteeritud dokumendi osa identifitseerivad metaandmed (nt Lisa1, Joonis64, foto4, lehekülg jne).
- Tuleb veenduda, et failil oleks seos originaaldokumendi kirjeldusega EDHSis, UAMis vms metainfosüsteemis.
- Failinime pikkus võib olla kuni 30 tähemärki.
- Failinimes ei ole soovitatav kasutada lühendeid. Kui see on vältimatu, tuleb lühendid dokumenteerida (vt Digiteerimiseelsed tegevused).

3.2 Failide nimetamine arhiiviainese massdigiteerimisel

Dokumentide massdigiteerimisel tuleb failide nimetamisel lähtuda üldistest nõuetest.

Arhiiviväärtuslike dokumentide massdigiteerimisel soovitage:

- Moodustada failinimed Rahvusarhiivist saadud fond-nimistu-säilik (FNS) koodi alusel, lähtudes „Rahvusarhiivi digiteerimise standardis“ esitatud failide nimetamise reeglitest.
- Seadistada digiteerimise tarkvaras nii sisestatavad kui ka automaatselt genereeritavad failinime osad.
- Luua seosed digiteeritud faili ja originaaldokumendi vahel.

4. Metaandmed

4.1 Tehnilised metaandmed

Reeglina lisatakse digiteerimisega seotud tehnilised metaandmed faili päisesse. Samuti võib neid aga hallata ka eraldiseisvas süsteemis. Juhul, kui tehnilised metaandmed asuvad eraldi infosüsteemis, peab olema loodud seos dokumentide ja metaandmeid sisaldavate failide vahel.

Tehnilised metaandmed lisatakse failile vastavalt digiteerimiseks kasutatavas tarkvaras või väljundvormingus toetatud standardile (EXIF, XMP, IPTC). Metaandmete olemasolu tuleb kontrollida vahetult pärast digiteerimise lõppu (vt pt 7. Digiteerimise kvaliteedi kontroll) .

Nõutavad on vähemalt järgmised tehnilised metaandmed:

- faili looja (asutuse või digiteerija nimi);
- 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).

4.2 Kirjeldavad metaandmed

Dokumentide kirjeldavad metaandmed tuleb luua vastavalt asutusesisese töökorralduse reeglitele teenuste korraldamise ja teabehalduse alustele. Rahvusarhiivi üleandmisel tuleb tugineda arhiivivaalide kirjelduselementidele, mille näidistabel on kättesaadav juhise lisamaterjalides (Exceli tabelfail).

5. Tekstidokumentide digiteerimine

1. Dokument digiteeritakse originaalsuuruses (1:1), resolutsiooniga 300 dpi.
2. Kui dokument on mingil põhjusel halvasti loetav (nt väikeses kirjas või kustuva tekstiga), võib digiteerimiseks kasutatavat resolutsiooni suurendada.
3. Värviline tekstidokument digiteeritakse värviliselt (24 bit). Tekstidokumenti, mis ei sisalda värvidega edasiantavat infot (on täies mahus mustvalge), võib digiteerida ka halltoonides (8 bit).
4. Dokumendi leheküljed, mis sisaldavad teksti või muud infot (sh lehekülje number), digiteeritakse terviklikult. Tühje lehekülgi ei digiteerita.
5. Digiteerimisel tuleb dokument asetada seadmesse nii, et hiljem ei peaks faile töötleva. 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.

Kasutatavad failivormingud

Tekstidokumentide digiteerimisel tuleb kasutada alljärgnevat failivorminguid.

- **TIFF** (kompresserimata) – kui tekstidokument on üheleheline ega vaja OCRi ehk teksti otsitavust.

NB! Kui mitmeleheline tekstidokument sisaldab fotosid (nt köidetud fotoalbum), siis on soovitatav luua igast dokumendi lehest tervikvaade vormingus **PDF/A** ning digiteerida iga foto eraldi **TIFF**-vormingusse, lähtudes fotode digiteerimise nõudest.

- **PDF/A** – kui dokumendi tekst peab olema otsitav st. on vaja rakendada OCRi.

Parim võimalik digiteerimisel kasutatav PDF/A versioon 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.

- **TXT** – vormingut kasutatakse lihtsa struktuuriga dokumentide puhul, kui ei ole vaja säilitada kujundust. Fail tekib OCRi rakendamisel.

NB! Kuna tekstifail tekib OCRi rakendamisel, siis tuleb veenduda tulemuse vastavuses originaaldokumendiga ning testida, kas otsing OCRitud tekstist toimub veavabalt.

Kui on oluline tagada digiteeritud dokumendi visuaalne samasus originaaliga, on soovitatav salvestada fail digiteerimisel vormingusse **PDF/A** või **TIFF**.

NB! Fotod ja muud mittetekstilised objektid (nt kirjamargid) tekstidokumendi peal või vahel, digiteeritakse fotode digiteerimise nõudeid järgides.

6. Fotode digiteerimine

- 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 veel kord eraldi fotode digiteerimise nõuete järgi.
- Fotode skaneerimise resolutsioon sõltub originaali mõõtmetest, mis on nähtavad järgnevas tabelis:

Skaneeritava foto mõõdud	Skaneerimise resolutsioon
Pikema külje pikkus on kuni 4 cm	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

Kasutatavad failivormingud

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

- **TIFF** (komprimeerimata kujul),
- **JP2** (JPEG2000 Part 1, kadudeta pakkimine).

Fotode digiteerimisel pikaajaliseks säilitamiseks on infokvaliteet paremini tagatud **JP2**- ja **TIFF**-vormingute korral. Tihedamat pakkimist võimaldab aga **JP2**-vorming.

7. Digiteerimise kvaliteedi kontroll

Igapäevasel digiteerimisel (nt ainese dokumendihaldusesse hõlmamisel) peab saadud tulemi üle vaatama ja seda originaaliga võrdlema vahetult pärast dokumendi digiteerimist.

Kontrolli käigus tuleb:

- veenduda, et saadud fail avaneb nõuetekohaselt;
- olla kindel, et sisu on loetav ja arusaadav;
- veenduda, et digiteeritud koopia on visuaalsel vaatlusel originaaldokumendiga vastavuses;
- veenduda, et see vastab vormingule esitatud nõuetele (nt vormingu versioon);
- kindlustada, et digiteerimise tulem **on** seostatud metaandmetega EDHSis, UAMis või vormistatud vastavasse Exceli tabelifaili (vt Lisamaterjalid).

Massdigiteerimisel tuleb kõiki digiteerimisel saadud koopiaid originaaldokumendiga võrrelda peale digiteerimise lõppu või vastavalt kokku lepitud ajaraamile. Juhul, kui massdigiteerimine tellitakse teenustööna, on mõistlik enne tööde algust kõigi osapoolte vahel kokku leppida, kes ja millal digiteeritud kujutisi kontrollib.

Lisaks ülal nimetatud punktidele tuleb massdigiteerimisel üle vaadata ka:

- metaandmete olemasolu ja vastavus reeglitele (kasutada näiteks tarkvara [ExifTool](#));
- digiteeritud kujutiste tehniline kvaliteet,
- failis sisalduva teabe terviklikkus,
- failinimedede õigsus ja vastavus nõuetele.

Lisamaterjalid

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