

Digitaalse arhiveerimise suundumused Eestis

Lauri Leht

Rahvusarhiivi digitaal-arhiivi direktor

Kui digitaalseid dokumente õigel ajal ei arhiveerita, ei ole praeguste riigiasutuste dokumentatsioonist tulevikus suurt midagi alles.

Peamine Eesti ajalugu, kultuuri, riiklust ja ühiskondlikke olusid dokumenteerivaid arhivaale koguv ja säilitav institutsioon on Rahvusarhiiv. Arhivaalid tuleb vastu võtta nende loomise ajast, kohast ja teabekandja iseloomust sõltumata. Traditsioonilisi käegakatsutavaid arhivaale hoitakse Eestis riiklikes arhiivides juba 1921. aastast, kui arhiivikomisjoni töö tulemusel alustasid tegevust ajalooline Riigi Keskarhiiv ja Riigiarhiiv, praeguse Rahvusarhiivi eelkäijad. Eesti riigile üleilmsest kuulsust toova ja ühiskonnale juba harjumuspäraseks saanud digitaalse asjaajamise tulemusena tekkivate digitaaldokumentide arhiveerimine on kesksesse arhiivi alles jõudnud.

Digitaalsed dokumendid

Paberarhivaalide üleandmine riigiasutustest Rahvusarhiivi on viimasel ajal olnud problemaatiline arhiivihoidlate ruumipuuduse tõttu, seepärast oodatakse suuremate koguste vastuvõtmiseks pikisilmi Rahvusarhiivi uue hoone valmimist. Paberarhiivi üleandmise venimine asutusest keskarhiivi paari aastakümne võrra ei ole igavikulises vaates siiski midagi katastroofilist, sest arhivaalid on igal juhul alles, kuigi asutustes laiali ja üpris keeruline on neile juurde pääseda. Digitaalselt sündinud dokumentide puhul on aga päris kindel, et mitte midagi tehes ei ole meil tulevikus 21. sajandi teise kümnendi riigiasutuste dokumentatsioonist eriti midagi alles.

Rahvusarhiiv on viimastel aastatel tegutsenud aktiivselt, et olla nii sisuliselt kui ka tehniliselt valmis digitaalsete arhiivimaterjalide tulvaks, mida prognoosime alates 2015. aastast. Rahvusarhiiviga teevad uute lahenduste loomisel tihedat koostööd ka kohalike omavalitsuste alluvuses olevad Tallinna Linnaarhiiv ja Narva Linnaarhiiv. Vaid kolmest asutusest koosnev Eesti arhiivindussektor on Euroopas konsolideeritumaid ja efektiivsemaid, mis annab eriti selge tulemise infotehnoloogia lahenduste ühisel juurutamisel.

Senised digitaalsete materjalide üleandmised on olnud seotud eelkõige asutuste või andmekogude likvideerimise või ümberkorraldamisega (hooneregistri muutumine ehitisregistriks, rahvastikuministri büroo likvideerimine, omandireformi andmekogude sulgemine, riikliku eksamikeskuse funktsioonide üleviimine sihtasutusse jms). Kuid peatselt on täis saamas esimene dekaad digitaalse allkirjastamise massilisest kasutuselevõtust riigiasutustes ning „varase perioodi” digitaalseid dokumente enam aktiivselt igapäevatoos ei vajata. Praktikas pööratakse digitaalarhiivi poole siis, kui asutusel on vaja muuta dokumentide haldamiseks kasutatavat tarkvara ning tekib küsimus vanade dokumentide saatusest – koos kõigi kirjeldusandmetega võib nende uude süsteemi konverteerimine olla üpris keeruline ja mõistlikum on need üle anda Rahvusarhiivi.

Uue arhiiviseadusega, mis jõustus 1. jaanuaril 2012, ulatas Rahvusarhiivile abikäe ka seadusandja, sest arhivaalide varasema 20-aastase avalikku arhiivi üleandmise tähtaja asemel sätestati üleandmistähtjaks 10 aastat, mõeldes eelkõige digitaalarhivaalidele. Nõudeid detailsemalt avav arhiivieskiri soovitab -arhivaale üle anda viieaastaste kogumitena, kuid digitaalarhivaale sagedamini. On arusaadav, et praktikas tähistab 10-aastane tähtaeg eelkõige suurusjärku, sest ühelt poolt tuleb arhivaalid üle anda niipea kui asutus neid enam ei vaja, teiselt poolt võib riigiarhivaar põhjendatud juhtudel tähtaega pikendada. Asutused peaksid suhtlema Rahvusarhiivi spetsialistidega, et optimaalne lahendus leida.

Migreerimine või emuleerimine?

Pikaajalist digitaalarhiivi pidades tuleb laias laastus valida kahe strateegia, migreerimise või emuleerimise vahel. Rahvusarhiiv on valinud migreerimise strateegia, see tähendab, me tagame arhivaalidele pikaajalise juurdepääsu neis sisalduvat infot uutesse failiformaatidesse migreerides ehk konverteerides. Kui praegu on põhiline vorming PDF, siis kaugemas tulevikus PDF-i tehnoloogia kadudes võtame kasutusele vastava uue vormingu ja kanname info sellesse edasi. Emuleerimisel failivorminguid ei muudeta, vaid keskendutakse vastava tehnoloogia pikaajalisele töökorras hoidmisele. Emuleerimise valdkonnas on tugevaimad Hollandi rahvusarhiivi spetsialistid ning see on asendamatu näiteks arvutimängude pikaajalisel säilitamisel.

Migreerimisstrateegiat kasutades on ilmne, et ka digitaalse allkirja vormingut ei saa pikaajaliselt säilitada, sest Eesti digitaalallkirja vormingud DDOC või BDOC on oma iseloomult infokapslid, mis sisaldavad teisi faile. Kui selle teise, digiallkirja sees oleva faili vorming on vananenud ja info mittekasutatav, siis ei ole palju kasu ka sellest, et digiallkiri võib olla hästi loetav ja kontrollitav. Seepärast võetakse digitaalselt allkirjastatud dokumentide arhiivi üleandmisel failist välja nii seal sisalduvad teised failid kui ka digiallkirja andmed. Tekstifailid peaksid olema PDF-id, kuid kui need on mõnes firmapõhises vormingus (tüüpiliselt DOC, DOCX või RTF), siis nad konverteeritakse PDF-iks. Digiallkirja andmed kirjutatakse digitaalarhiivi infosüsteemi andmebaasi ning esitatakse edaspidi arhivaali kirjeldusandmetes failist eraldi. Kõik failidega tehtud toimingud logitakse ehk registreeritakse. Seetõttu on digitaalarhiiv pikaajaliselt üles ehitatud usaldusele, et failidega tehtud töökorralduslikud protseduurid ja tehnilised toimingud on olnud korrektsed ning ajast aega kanduv info autentne. Usaldusväarsuse kinnituseks on digitaalarhiividel tavaks lasta end aeg-ajalt auditeerida. Kui tuleviku kohtuasutus, näiteks aastal 2075, ei usalda siiski digitaalarhiivist väljastatavat mitmekordselt migreeritud infot, siis alati hoitakse alles ka originaalfail ehk DDOC, mille sees on PDF, ning kohtul on võimalus DDOC-i iseseisvalt inspekteerida.

Arhiveerimine dokumendihaldus-süsteemidest

Riigiasutuste digitaalseid dokumente arhiveerides oleks ideaalne lahendus, kui kõigis asutustes oleks kasutusel üks dokumendihaldussüsteem, mis on liidestatud keskse digitaalarhiiviga, ja vajalikud andmed kantaks üle paari nupulevajutusega. Paraku nii lihtne see siiski ei ole. Nimelt on Eesti riigiasutustes kasutusel üle kümne dokumendihaldustarkvara ning kohati on ka sama tarkvara eri installatsioonides märkimisväärsed erinevusi. Rahvusarhiivi digitaalarhiivile tähendab see, et iga asutus tuleb digitaalarhiivi süsteemiga eraldi liidestada.

Euroopa riikide üldlevinud praktika kohaselt kehtestab keskarhiiv digitaalsete dokumentide üleandmiseks täpsed ja ranged tehnilised nõuded ning arhiivi ei huvita, kuidas riigiasutused oma süsteemid nende nõuetega vastavusse viivad. Seetõttu sageli asutused ei annagi digitaaldokumente üle, sest neil pole ressursse ega pädevust vajalikuks nõuetele vastavaks konverteerimiseks.

Eesti Rahvusarhiiv mõtles digitaalarhiivi arendama asudes kohe algul veidi laiemalt. Kõigepealt loodi riigiasutuste arhivaaridele tarkvaraline abivahend, et asutustel oleks üldse tehniliselt võimalik digitaaldokumente üle anda ja iga asutus ei peaks kulutama -märkimisväärselt ressursse andmete konverteerimisele Rahvusarhiivi nõuetele vastavaks. Tarkvara, mille nimeks sai Universaalne Arhiveerimismoodul (lühidalt UAM), muutis seni abstraktse digitaalarhiivinduse teema nii riigiasutuste kui ka Rahvusarhiivi arhivaaride ja dokumendihaldurite jaoks võib-olla esialgu keerukaks, kuid siiski praktiliselt arusaadavaks tööülesandeks. UAM-i esimene versioon valmis 2008. aasta lõpus ning seda on igapäevasest kasutamisest tulenevast tagasisidest johtuvalt mitu korda täiendatud.

Selleks et UAM asutuse dokumendihaldussüsteemiga liidestada, on vaja, et asutuse süsteem oskaks andmeid mingilgi kujul XML-is eksportida. Praktiliselt kõik süsteemid seda ka võimaldavad. Sellest tulenevalt ei ole liidestamine mitte niivõrd tehniline, kuivõrd pigem intellektuaalne ülesanne. Tuleb koostada andmekirjelduste vastavustabel, et näiteks väli „Document Title” dokumendihaldussüsteemis vastaks Rahvusarhiivi kirjeldusskeemis väljale „Dokumendi pealkiri”. Ülesanne ei ole muidugi nii triviaalne ja eeldab kohati põhjalikku süvenemist andmekirjeldustesse, aga kui see on koos Rahvusarhiivi spetsialistidega kord läbi tehtud, siis edaspidine dokumentide eksport UAM-i on juba latus. Eksportitud dokumente on arhivaaril võimalus iseseisvalt, ilma infotehnoloogia spetsialistide abita täiendavalt kirjeldada ja korrastada ning kontrollitud kujul X-tee kaudu Rahvusarhiivi digitaalarhiivi vastuvõtumoodulisse edastada.

Digitaalarhiivi infosüsteemi pikaajalise säilitamisega tegeleva osa aluseks on rahvusvaheliselt kasutatav tarkvara Tessella SDB4, mis algselt töötati välja Briti rahvusarhiivile. Kuna failide pikaajalisele säilitamisele enamasti ei rakendu lokaalsetest keelelistest eripäradest, arhiivikirjelduste tavadest, juurdepääsureeglitest või seadusandlusest tulenevad erinevused, on Eesti arhiivil selles osas võimalik kasutada ühist tarkvara ja teha koostööd Briti, Hollandi, Soome, Ungari, Läti, Austria, USA, Malaisia jt riikide arhiivide ning raamatukogudega.

Kohalikud vajadused ja tavad on aga olulised arhivaalidele juurdepääsu tagamisel. Rahvusarhiiv lõi oma soovidele vastava uue veebipõhise arhiivi infosüsteemi tarkvara, mis 2013. aasta lõpuks on jõudnud valmimisjärku. Selle põhiline ülesanne on ühendada ühtsete arhiivikirjelduste kaudu juurdepääs paber-, foto-, filmi- ja heliarhivaalidele, nende digiteeritud koopiatele ning samuti digitaalselt sündinud arhivaalidele.

Arhiivimaterjali digiteerimine

Kui digitaalselt sündinud materjali üleandmine võtab veel hoogu, siis arhiivis käib juba aastaid arhiivimaterjalide digiteerimine. Suuremad projektid euroraha eest, keskmised ja väiksemad projektid kodumaiste ja välismaiste tellijate toetusel, jooksev töö võimaluse piirides riigieelarve raames. Kõigepealt suuremat uurijahuvi pakkuvad või halvemas seisukorras materjalid, aja jooksul kõik ülejäänud. Kui praeguseks on digiteeritud mahu järgi ehk ligikaudu kolm protsenti Rahvusarhiivi arhivaalidest (ligi 12 mln kaadrit), siis tõenäoliselt kestab digiteerimine veel aastakümneid. Digiteerimise tempo on otseses seoses kasutada olevate eelarvevahendite hulga, sest tegevus on intellektuaalses mõttes üpris lihtne. Digiteeritud faile tuleb samuti pikaajaliselt säilitada, kuid erinevalt digitaalselt sündinud dokumentidest ei ole neil originaali staatust ega tõestusväärtust, mistõttu on säilitusprotseduurid lihtsamad.

Rahvusarhiivi struktuuri kuulub ka filmiarhiiv, mis säilitab suurt hulka filmi-, video-, heli- ja fotoarhivaale. Digiteerimisprojektide raames tekkinud andmete seas moodustavad mahult olulise hulga just digiteeritud filmid ja videod. Filmiarhiivil on kohustus arhiveerida kõik Eestis avalike vahendite eest tehtavad filmid, mis juba mitmeid aastaid valmivadki ainult digitaalselt. Digitaalselt sündivate filmide failivormingud on jätkuvalt pidevas arenguprotsessis, seetõttu on ühtse arhiveerimisvormingu kokkuleppimine seni olnud komplitseeritud. Sageli tuleb kultuuriväärtuse säilimise tagamiseks leppida suvalise firmapõhise vorminguga failide vastuvõtmisega ning tegelda nende pikaajaliselt säilivasse kesta valamisega edaspidi.

Probleemid ja ülesanded

Kuigi kõrvalt vaadates tunduvad eri tüüpi mäluasutused üpris sarnased, on tegelikult arhiivide, raamatukogude ja muuseumide ülesehitus üpris erinev, mis kõige selgemini hakkab silma infosüsteeme luues. Võrreldav on ainult digiteeritud failide pikaajaline säilitamine, aga niipea kui on tegu andmete kirjeldustega, ilmnevad süsteemsed erinevused. Nimelt koosnevad arhiivikirjeldused mitmetasandilistest hierarhiatest (arhivaal kuulub sarja, sarjad kuuluvad arhiivi, vahel on kasutusel ka allarhivaalid ja allsarjad), raamatukogudes ja muuseumides on aga valdavalt üksikobjektid. Seetõttu on keeruline luua ühiseid infosüsteeme, isegi sedavõrd keeruline, et Euroopa ühtse digitaalse kultuuripärandi portaaliga Europeana liitumiseks loovad rahvusarhiivid esmalt Euroopa ühise arhiiviportaaali, mis siis kõikide arhiivide andmetega korruga liitub Europeanaga. Euroopa arhiiviportaaali loomise juures on ka Eesti Rahvusarhiivil kandev roll valmiva tarkvara testimisel ja veebilahenduste uuenduslikkuse tagamis el.

Kõige rohkem digitaalset materjali vastu võtnud rahvusarhiiv Euroopas on Taanis, kus sellega alustati juba 1980. aastate lõpus. Taanis kasutatav digitaalse materjali arhiveerimise meetod põhineb andmebaaside väljavõtetel, see on nii ka dokumendihaldussüsteemidest digitaaldokumentide arhiveerimisel. Seetõttu on võimalik failid ladusalt vastu võtta, kuid arhivaalide edaspidine kasutamine ning otsingute tegemine on problemaatiline. Taanlastel on seetõttu keeruline üle minna nüüdisaegsemale XML-il põhinevale dokumendipõhisele arhiveerimisele.

Kui digitaaldokumentide arhiveerimine on Eestis teoorias ja varsti praktikaski hästi paigas, siis märksa keerukam on mõne andmekogu või registri arhiveerimine. Andmebaaside arhiveerimiseks on välja töötatud üldised juhised ja nõuded, keskendudes taas andmete pikaajalisele säilivusele. Kuid pikaajalise säilivuse huvides lähevad sageli kaotsi andmebaasi mugava kasutamise seotud funktsionaalsus ja standardpäringud, mis tulenevad konkreetse andmebaasi tarkvarast ega ole aastate pärast enam kasutatavad. See on kõigi arhiivide ühine probleem ning Euroopa arhiivid, sealhulgas Eesti Rahvusarhiiv, on alustamas andmebaaside arhiveerimise ühiste lahenduste väljatöötamist Euroopa Komisjoni raamprogrammi euroraha toel. Sama projekti raames püütakse ühtlustada ka digitaaldokumentide arhiveerimise praktikat, kus Eestil on pakkuda paindliku UAM-i loomise positiivne kogemus.