

Dokumentide aastakasv Eestis – mõõtmisest ja eriti mõõdupuust

Priit Pirsko

Arhiiviteooria on ideestik, mida vähesed tunnevad või tunnistavad ning üksikud viljelevad. Arusaadavalt on ka elav arhiiviteoreetiline dispuut siinmail üsna raskesti saavutatav – asjatundjate vähesuse tõttu, aga samuti praktiliste probleemide domineerimise tõttu meie igapäevatoos. Ometi on üsna paradoksaalsel kombel just praktilise arhiivinduse ette kerkimas aina enam keerulisi küsimusi, millele vastamine pelgalt senise kogemuse najal ei ole võimalik.

Sestap võibki nentida, et 20. ja 21. sajandi vahetusel sündinud eestikeelne arhiiviteoreetiline mõttevahetus on praeguseks leidnud küll kitsa, kuid samas kindla voolusängi, mis rikastab siinset arhiivindusmaastikku uute ideede ja mõttekäikudega nii terminoloogia, hindamise, juurdepääsu kui ka digitaalse arhiivihalduse vallas. Nii mõnigi autor on maakeelselt infoväljalt suundunud edasi indogermaani maastikule.¹ Katrin Roosilehe artikkel, mis keskendub dokumentide aastakasvu mõõtmismetoodikale, on seega üle kinnitamas arhiiviteoreetilise mõtte võimalikkust Läänemere ja Peipsi vahelisel lausmaal. Minu alljärgnev vaatlus on pigem dialoogi- kui dispuudikatse.

K. Roosileht annab ülevaate uurimisprobleemist, mis keskendus katsele luua selgus sel-

les, kui palju on meie ümber arhiivinduslikku hindamist vajavad infot – kui palju tekib Eesti riigiasutuste koostöös dokumente ühe ajaühiku (aasta) jooksul. Tõepoolest, sellist meetodikat mujal maailmas teadaolevalt välja töötatud pole. Selle üks põhjuseid võib olla mõõtmise ülim keerukus, sest dokumente luuakse ja saadakse pidevalt ja kõikjal, ning nende olek on nii sisulises kui ka füüsilises (teabekandja) tähenduses muutuv ja mitmekesine. Lõpuks pole ühist meelt selleski, missugune info on dokument ja missugune kindlasti ei ole.

Dokumentide nn. ülelugemise üldtunnustatud metoodika puudumise teine põhjus võib seisneda selle küsimuse püstituse tähtsusetuses või liigsuses. Juurdlemine selle üle, kui palju on liivateri mererannal või vihmapiisku rajahoos, võib osutada mõtteharjutuseks ilma kasuliku tulemuseta, võiksid väita John Robertsi laadi arhiiviteooria paadunuimad vastased.²

Siinkirjutaja läheb viimasest küsimusest teemateadlikult üle ja keskendub alljärgnevalt kahele aspektile K. Roosilehe töös – projekti sisuliste tulemuste vaagimisele ja teiseks, väljapakutud mõõtmismeetodi hindamisele.

Uurimuse tulemused tekitavad kahetisi tundeid. Ühest küljest on valgustav teada saada, et avaliku sektori dokumentide juurdekasv võis 2005. aastal olla umbkaudu 22,3 miljonit. Teisalt, esitatud numbrite võrdlemine kergitab nii mõnegi küsimuse või isegi kahtluse. Teatud määral kumab saadud andmete lahtiseletamisel ebakindlust autori enda sõnakasutusestki, näiteks: “tulemusi peab käsitlema teatava üldistusastmega”. K. Roosileht küll kinnitab, et andmete analüüsimine pole tema käsitlemise eesmärk, kuid ilma hinnanguta analüüsitulemuste vettpidavusele pole ju õigupoolest võimalik otsustada ka meetodi kohasuse üle.

Põhimõtteliselt võib uurimuse läbiviijate poolt langetatud otsust – kasutada asutuste paljususe tingimusi statistiliselt esindusliku valimi meetodit – pidada ainuõigeks. Järgnevalt uurimuses kasutatud tüpoloogiline ehk strati-

¹ Kõige värskema näitena: J. Valge, B. Kibal. Restrictions on Access to Archives and Records in Europe: A History and the Current Situation. – Journal of the Society of Archivists 2007, vol. 28, nr. 2, lk. 193–214.

² J. Roberts. Archival Theory – Much Ado about Shelving. – American Archivist 50 (Winter 1987), lk. 66–74; J. Roberts. Archival Theory: Myth or Banality? – American Archivist 53 (Winter 1990), lk. 110–120.

fitseeritud valimi meetod on esmapilgul samuti loogiline, kuid seda eeldusel, et asutuste tüübiti grupeerimisel ei tehta olulisi vigu. Samuti on kriitilise tähtsusega, et iga grupi (kokku 22) ja rühma (kokku 41) valim võrduks tõesti selle grupi (rühma) statistilise keskmise väärtusega. Mida suurema ja eritüübilisema grupiga on tegu, seda suurem on ka hälbe risk.

Uurimuse läbiviijate tunnustuseks tuleb öelda, et neid riske on nad ka tajunud ja jõudumööda maandanud. Näiteks on projekti otsustavas faasis loobunud asutuste grupist, kuhu algse kava järgi oleksid pidanud kuuluma niinimetatud ülejäänud asutused (sh. Eesti Haigekassa ja Hiiumaa Turismiinfokeskus, mis on tõesti täiesti võrreldamatud).

Samas ei saa märkimata jätta, et mõne rühma puhul tekkis siinkirjutajal ka kahtlusi väljavalitud nn. keskmise asutuse keskmiseks olemisses. Nõnda näiteks ei tundu Tallinna Tööstushariduskeskus 71 kõrgkooli ja kutseõppeasutuse esindusliku keskmisena ega Kuressaare Haigla AS 50 haigla parima keskmise näitena. Teisisõnu, isegi siis, kui need ja teisedki valitud asutused sobiksid aritmeetiliseks keskmiseks oma töötajate või õpilaste või patsientide või mingi muu näitaja poolest, ei pruugi nende asjaajamises ringlev dokumentatsioon peegeldada rühma kogudokumentatsiooni keskmist. Nõnda võib eeldada, et kümnekonnas Eesti suurimas haiglas või pooltosinas suurimas kõrgkoolis ringleb üldse suurim osa sarnaste asutuste dokumentidest. Või vähemasti väga suur osa meditsiini- ja haridusalasest infost. Kordi väiksema dokumentide hulgaga Tööstushariduskeskuse või Kuressaare Haigla tulemuse korrutamine vastava grupi asutuste arvuga võib seega jääda vägagi kaugele grupi tegelikust dokumentide üldarvust. Kõrgkoolide puhul ületab ilmselt Tartu Ülikooli ja Tallinna suuremate kõrgkoolide dokumendiringluse ulatus kogu ülejäänud kõrgharidusealase infovahetuse – seega aritmeetiline keskmine võib selliste väga erineva suurusega asutustega gruppide puhul osutada eksitavaks.

Niisuguse kahtluse hajutamiseks oleks olnud võimalik toimida kahel moel. Esiteks, ja keerulisemalt, oleks aritmeetiliste keskmiste tunnuste asemel olnud ehk võimalik kasutada

mediaankeskmiestele tunnustele vastavaid asutusi. Ja kui see oleks olnud liiga komplitseeritud, siis oleks kindlasti kaalumist väärinud rühmade arvu suurendamine. Hiigelülikoolid või -haiglad oleksid võinud moodustada omaette grupi ning pisikõrgkoolidest või maakonna-haiglatest tuletada eraldi keskmise väärtuse.

Teine suurem küsitavus saadud andmete usaldusväärsuse kohta tekib valimisse valitud asutuste dokumentide arvu kõrvutamisel. Näiteks on täiesti uskumatu, et ühe suurima valitsusasutuse, majandus- ja kommunikatsiooniministeeriumi asjaajamises tekib aasta jooksul ainult 932 alatise säilitusväärtusega dokumenti, s. o. kõigest 0,31% teabe kogumahust. Samal ajal ladestuvat näiteks Rahvusarhiivis 2,7 korda ja Riigikantselei toimetamistes 5,2 korda rohkem alatiseks säilitamiseks mõeldut.

Nii see olla ei saa. K. Roosileht jõuab ise järeldusele, et uurimuse tulemuste järgi on 6% dokumentidest alatise säilitusväärtusega, ja soovib dokumentide loeteludes teha eelkõige kirjavahetuse sarjade puhul korrektiive, et alatise säilitusväärtuse (ja hiljem eeldatava arhiiviväärtusega) saaks keskmiselt 10% dokumentidest. See soovitus on siinkirjutaja meelest arukas.

Ometi on raske ette kujutada, et mõne ministeeriumi või muu valitsusasutuse dokumentidest on alatiselt säilitatav vaid murdosa, alla 1% koguhulgast. Küsitluse valimist torakab peale majandus- ja kommunikatsiooniministeeriumi silma veel Keskkonnainspeksioon, kus 2005. aastal olevat tekkinud vaid 70 arhiivikõlblikku dokumenti (0,16% koguhulgast). Muudest asutustest näib näiteks Tartu Ülikooli Tehnoloogiainstituudi 22 alatise säilitusväärtusega dokumenti (0,11% koguhulgast) pehmelt öeldes tagasihoidlikuna, arvestades tõsiasjaga, et tegemist on väga olulise teadus- ja arendusasutusega, kuhu kuulub mitu arenduskeskust ja tuumiklaborit.

Arvan, et siin on tegemist elementaarse veaga statistilistes uuringus – suur osa alusandmetest (antud juhul dokumentidest, ja nende aluseks olevatest funktsioonidest) on jäänud küsitlusfookusest mingil põhjusel välja. Neid ei ole hõlmatud valimisse, mis peaks peegeldama keskmist. Küllap on selline kahe silmavahele

jätmine kõige hõlpsamalt tekkiv kõige keerulisemate organisatsioonide puhul, kus põhitegevust kajastav ja seega paljuki alatist säilitusväärtust eviv dokumentatsioon ladestub hajali asuvas struktuuriüksusis. Ja vastupidi, mida tsentraliseeritum organisatsioon, seda väiksem on dokumentide mittemärkamise oht.

Käesoleva käsitluse kontekstis tuleb aga paraku eeldada, et vähemasti mõne asutuse rühma puhul jääb keskmiseks valitud asutuse näit ülelugemisvea tõttu eksitavaks. Stratifitseeritud valimi meetodi puhul mõjutab keskmise näidu viga kogu rühma ja grupi arvestuslikku tulemust. Kui suur selline viga võib olla, seda ei sõanda siinkirjutaja küsitluse läbiviimise detaile tundmata oletada.

Uurimuse numbrilistesse tulemustesse süvenemisel torkab lisaks ülaltoodule silma veel üks põhimõtteline kitsaskoht. See puudutab selliseid asutusi, mille põhiülesanded seonduvad teatavate inimgruppide teenindamise, hoolduse või kohtlemisega (õppeasutused, haiglad, vanglad) ja mille dokumentide hulk on niisuguste isikupõhiste avalike ülesannetega otseselt seotud. Siin peaksime nägema numbreid, mis seisavad omavahel võrreldavas loogikas, aga tegelikult on täiesti ebaloogilised, et mitte öelda valed.

Vaadeldgem esmalt haridusasutuste gruppide dokumentide arvestuslikku koguhulka, nii nagu see küsitluse tulemusena on esitatud. Koolieelsetes õppeasutustes (lasteaiad) tekib aasta jooksul väidetavalt 1,55 miljonit dokumenti. Võime arvestada, et tegemist on vanuselise grupiga 1–7 eluaastat ning pidada silmas, et kaugeltki mitte kõik lapsed ei käi lasteaias. Seega saame eeldada koolieelsete õppeasutuste dokumendihalduse kogusubjektina umbes 50 000 last (sündimuse juures 11–14 tuhat last aastas).

Koolikohustus on teatavasti üldine, seega saame oletada, et suurem osa vanusegrupist 7–18 eluaastat on seotud põhikoolide ja internaatkoolide (496 000 dokumenti aastas), keskkoolide ja gümnaasiumide (369 000 dokumenti) ning algkoolide ja lasteaed-algkoolide (57 000 dokumenti) asjaajamise ja dokumendiringlusega. Laias laastus võib väita, et erinevates koolides on registreeritud

kolm korda rohkem lapsi kui lasteaedades ning seega peaks sellistes asutustes tekkima ka umbes kolm korda rohkem dokumente kui koolieelsetes asutustes. Usun, et see vahe peaks olema suuremgi kui kolm korda, sest lasteaiad on siiski lihtsama struktuuri ja funktsioonidega kui koolid.

Niisiis, eeldusel, et lasteaedade dokumentide koguhulk on õige, peaks koolides aasta vältel ladestuma umbes 4,65 miljonit dokumenti. K. Roosilehe poolt toodud tabelleist aga nähtub, et alg-, põhi- ja keskkoolides, gümnaasiumides ja internaatkoolides kokku luuakse ja saadakse kõigest 0,92 miljonit dokumenti aastas. Erinevus loogilise ootuse ja näiva tegelikkuse vahel on viiekordne!

Arvan, et siin on tegemist suure veaga kas lasteaedade või koolide dokumendiarvestuses. Iseenesest ei saa välistada ka eksimist mõlema rühma puhul. Väga teoreetiliselt pinnal saab viimasest tulenevalt ka osundada tillukesele võimalusele, et üks suur viga on taandatud teise suure veaga ja lasteaedade-koolide dokumentide väidetav koguhulk – 1,55 miljonit pluss 0,92 miljonit, kokku 2,47 miljonit – on tõene või tõenäoline. Kuid see ei oleks enam teaduslikule meetodile põhinev analüütiline järeldus, vaid juhuslik numbriline kokkulangevus.

Siinkirjutaja kaldub siiski arvama, et dokumentide kokkuarvestamisel on mitmel juhtumil sisse lipsanud põhimõttelisi eksimusi. Lasteaedade dokumendihulk ei klapi ju kuidagi ka kõrgkoolide ja kutseõppeasutuste näitajaga, ületades viimase pea kahekordselt. Samal ajal teame, et ainuüksi kolmes suuremas ülikoolis õpib umbes sama palju tudengeid, kui on lapsi kokku Eesti lasteaedades, ülejäänud 68-st sama grupi kõrg- ja kutseõppeasutusest rääkimata.

Jätkakem samalaadsete näidetega. Vaatamata digitaalse dokumendihalduse võidukäigule meditsiinasutuses ei usu siinkirjutaja seda, et haiglates aasta jooksul tekkiv dokumendihulk ületab ainult kahekordselt üsna marginaalse elanikkonna osaga tegelevat vanglate dokumendiringlust. Samuti ei ole ilmselt alust arvata, et 2005. aastal majandusbuumi kogenud Eesti notarite tegevuses tekkis vähem dokumente kui mainitud kinnipidamisasutustes.

Õigupoolest võiks niisuguste kahtlustega veel edasigi minna, kuid palju asjakohasem on siinkohal küsida: millest võisid sellised ebakohad tekkida? Kus on vigade või veidruse algpõhjus ja kas saame siin piirduda vaid võimaliku rehkendusvea konstateerimisega üksikute valikusse hõlmatud asutuste dokumendi arvestuses?

K. Roosileht keskendub oma artiklis just uurimuse metoodika tutvustamisele ja mina arvan, et vigade peamine selgitus peitubki valitud meetodi ülevaatamises. Ebausutavad tulemused viitavad samuti võimalikele nõrkustele meetodi rakendamisel, mitte ainult labastele arvestusvigadele. Kõrvaltvaatajana tundub mulle, et tegemist võib olla kaht sorti libastumisega stratifitseeritud valimivaatluses.

Esiteks. Tüpoloogiline ehk stratifitseeritud valim eeldab, et üldkogumi jagamine osadeks toimub moel, mis tagab osade (asutuste rühmade) seesmise ühetüübilisuse. Eelnevas arutluses selgus, et teatud juhtumel ei pruugi nominaalse tunnuse järgi ühesuguste asutuste (kõrgkoolid, haiglad) aritmeetilise keskmisena käsitletav asutus anda adekvaatset tulemust rühma koguarvestuses – juhul, kui äärmuste hulk on suur või isegi otsustav. Seega, ilmselt oleks mõistlik olnud rühmade ühetüübilisuse huvides nende arvu suurendada.

Teiseks. Nimetatud meetodi puhul on tähtis, et valim oleks piisavalt suur, et väljendada esinduslikku ja seega ka üldkogumile teisendatavat väärtust. Seda on ka uurimuse autorid teadnud ja seadnud eesmärgiks ca 2% suuruse valimi avaliku sektori asutuste seast. Tundub, et järgneva käigus on see siht ununenud või on seda asjaolude muutumisel korrigeeritud. Tegelikult on küsitletud vaid 2/3 algul kavatsatud asutustest ehk valimi moodustab 1,35, mitte 2 protsenti. Tulemusi analüüsides tundub, et siin võibki olla üks paljude ebaloogilisuste põhjus – valimisse sattunud asutused ei peegeldanud tegelikult kogumit statistiliselt representatiivsel viisil.

Lisaks tuleb nii mõnegi rühma puhul tähelepanu juhtida statistikateooria mõistes ilmselgelt liiga väikesele väljavõtule. Näiteks vallavalitsuste puhul on see 0,5 ja eelpool pikalt analüüsitud lasteaedade puhul ainult

0,2% üldkogumist. Kuna tegemist on väga suurte rühmadega, siis on võimaliku vea võimendumisefekt ka suhteliselt suurem. Seda riski oleks saanud maandada, kui suurimate rühmade puhul oleks silme eest hoitud vähemalt 2% raja valimi moodustamisel.

Valimi suurendamine oleks arvatavasti aidanud taandada ka neid erisusi, mis dokumendiloome puhul on subjektiivsed ega allu statistilisele töötlusele. Pean eelkõige silmas seda, et isegi sarnase töötajate arvu, ühesuguste ülesannete ja võrreldava eelarve korral võib asutustes ringlev dokumentide hulk erineda päris suurel määral. Nimelt on formaliseeritud informatsiooni osakaal asutuse asjaajamises paljuskii sõltuv kehtestatud asjaajamiskorrast, juhtimisstiilist (nt. suuline korraldus *versus* kirjalik) ja lõpuks ka dokumendihalduse kvaliteedist.

Nende märkustega lõpetades ei taha ma aga väita, et uurimus oleks olnud kasutu või üdini eksitav. Mitmed K. Roosilehe artiklis esile toodud järeldused on väga asjalikud ja tähelepanuväärsed. Raamatupidamisdokumentide puhul kirjendite ja alusdokumentide suhte määramine ning alatise, pikaajalise ja lühiajalise ainese vahekorra selgitamine dokumendihalduses on kindlasti abiks teema käsitlemisel tulevikus. Autori eesmärgipüstis – välja töötada usaldusväärne metoodika avaliku sektori asutuste dokumendiloome mõõtmiseks – on samuti vähemalt osaliselt kasutatav. Selle eelduseks on aga ilmnenud ebakohtade silumine – seda nii metoodika rakendamisel kui ka tulemuste vettpidavuse kontrollimise kaudu.



Priit Pirsko

(1964)

Lõpetanud Tartu ülikooli, filosoofiamagister ajaloo erialal aastast 1992. Töötanud Tartu Ülikoolis õppejõu ja teadurina, aastast 1993 erinevatel ametikohtadel arhiivinduses. Aastast 2000 riigiarhiivaar. Tartu Ülikooli külalisprofessor. Peamised uurimisvaldkonnad: arhiiviteooria ja -metodoloogia, arhiivinduse ajalugu. Eesti rahvastiku, sotsiaal- ja majandusajalugu 19. sajandil.