

Digihumanitaaria Eestis

PEETER TINITS, JANE KLAVAN, LIINA LINDSTRÖM

Keele ja Kirjanduse teemanumbri mõte tekkis seoses digikultuuri aasta toimumisega 2020. aastal. Numbri toimetajad on eest vedanud erinevaid digihumanitaaria-alaseid tegevusi Eestis ning pidevalt pidanud vastama küsimustele: mis on digihumanitaaria ja miks seda vaja on? Teemanumbri eesmärk ongi tutvustada eestikeelsele lugejaskonnale, mis on digihumanitaaria, milline on digihumanitaaria hetkeseis Eestis ja mida selle raames tehakse, ning püüda selle kaudu vastata, mida annab *digi* juurde humanitaariale. Siinsed artiklid kajastavad üsna hästi Eesti digihumanitaaria üldpilti: leidub mitmeid häid näiteid digihumanitaaria meetodite kasutamisest eri distsipliinide sees ja autorite seas on nii üksikuurijaid kui ka tööühmi, kes valdkonna arendamisega tegelevad. Loodetavasti annab käesolev artiklivalik parema arusaama digihumanitaaria meetodite kasutusvõimalustest.

Teemanumbrisse oleme koondanud mitmekesise valiku artikleid keeleteadusest kliimaajalooni. Tutvustatakse mõne eriala arengut digitaliseerivas maailmas (Ibrus jt; Langemets jt; Ojamaa jt; Plath jt), teisel on fookuses konkreetse uurimisküsimuse lahendamine mõne digihumanitaaria meetodiga ja selle meetodi üldisem tutvustus (Meriste; Pilvik jt; Puss; Šeļa). Tegemist ei ole siiski esimese digihumanitaariale pühendatud väljaandega Eestis: 2020. aasta lõpus ilmus ajakirja *Methis* teemanumber, mis põhines 2019. aasta Eesti digihumanitaaria konverentsi ettekannetel. Koos võiksid need eesti lugejale pakkuda mitmekülgset teadmist digihumanitaaria võimalustest koos kohalike näidetega.

Mis on digihumanitaaria?

Kuigi *digihumanitaaria* tundub võrdlemisi uus termin, on sel tegelikult pikk ja auväärne ajalugu: see viiakse tagasi esimeste arvutite loomiseni 1950. aastatel või lausa arvutuslike meetodite kasutuselevõtuni XIX sajandi lõpul. Paljud kvantitatiivsed meetodid humanitaarias on välja töötatud esmalt just paberi ja pliiatsiga arvutamiseks, moodne arvutusvõimsus võimaldab lihtsalt seda teha tõhusamalt (vt nt Šeļa siin numbris). Juba 1966. aastal hakkas ilmuma pikaajaline ajakiri *Computers and the Humanities* (kuni aastani 2004); 1984 alustas teine mõjukas ajakiri *Literary and Linguistic Computing*, mis kannab praegu nime *Digital Scholarship in the Humanities*. Seda vedav Euroopa Digihumanitaaria Assotsiatsioon sai alguse, ehkki teise nime all, 1973. aastal.

Märksõna *digihumanitaaria* tekkis 2000. aastate algul distsipliinide vastastikusel defineerimises, omamoodi värskenduskuurina ja programmilise liikumisena. Digirevolutsiooni teema kerkis ühiskonnas laiemalt esile ning ilmselt seetõttu avaldas

termin mõju ka digihuvilistele humanitaarteadlastele. Algusest peale on seda kasutanud ja defineerinud väga erineva taustaga uurijad, kes on sellele ka erinevaid lootusi pannud. Definiitsioone on nii palju, et need on kogutud veebileheküljele, mis sisaldab 817 digihumanitaaria definiitsiooni aastatest 2009–2014 (vt *What Is Digital Humanities?*). Defineerimise käigus on kirjutatud ideaalpilti visandavaid manifeste (nt Schnapp, Presner 2009), aga teistele uurijatele on see lihtsalt pragmaatiline märksõna, mille abil leida ühisosa teiste valdkondadega või selgitada rahastajatele töö eesmäärke (nt Kirschenbaum 2012). Pikemalt on digihumanitaaria defineerimisest, selle valdkonna kasvuraskustest ja enesemääratlusest juttu teemanumbri esimeses artiklis (Ibrus jt).

Kui püüame leida digihumanitaaria definiitsioonides ühisosa, siis enamasti on selleks digitehnoloogia rakendamine, mis eeldab infotehnoloogiaalaste teadmiste- oskuste olemasolu. Digitehnoloogia pakub uuenduslikku võimalust töödelda ja analüüsida suures koguses andmeid – seepärast hõlmab suurem osa digihumanitaaria all tehtavast teadustööst mingil viisil suurte andmehulkade töötlust, eri tüüpi andmete kombineerimist ning väga mitmekesiseid andmete analüüsi ja esitamise viise. Seega võiks digihumanitaaria tuumosa nimetada ka andmepõhiseks humanitaariaks (vt Laak jt 2020). Loomulikult on muidki lähenemisi, mis rakendavad digivahendeid pigem kvalitatiivses analüüsis või mõtestavad laiemalt digikultuuri arengut ja näiteks õppimist digiajastul (vt Ojamaa jt siin numbris).

Digihumanitaaria defineerimine või pigem defineerimatus on toonud kindlasti palju positiivset. Rõhutatud on digihumanitaaria suunatust ka väljapoole akadeemilist maailma, koostöö, avatuse ja mittehierarhiliste suhete väärtustamist ning teadlaste hulgas väljakujunenud normide kritiseerimist (nt retsenseerimis- ja publitseerimispraktikad) (Kirschenbaum 2010: 59; Bode, Arthur 2014: 2). Digihumanitaaria kui valdkonnaga kaasnevate metaküsimuste hulka, mis ei ole otseselt seotud teadusliku uurimistöö sisulise poolega, kuuluvad näiteks andmekaitse ja avatud teadus (vt nt Berry, Fagerjord 2017: 114–135; Gardiner, Musto 2015: 146–165). Need on olulised teemad, mis väärivad pikemat käsitlust, ning just digihumanitaaria andmepõhisus on aidanud algatada vastava arutelu. Kindlasti propageerib digihumanitaaria valjuhäälselt avatud ligipääsuga teadust, nii publitseerimisel kui ka andmete ja andmeanalüüsi skriptide kättesaadavaks tegemisel. Artikli aluseks olevad andmed ja skriptid on teinud andmerepositooriumis kättesaadavaks nt Pilvik jt. Avatud teaduse printsiip püüab tuua teadust lähemale ideaalile, kus kõik uurimistöö sammud on läbipaistvalt teostatud ja korratavad. Avatud teaduse põhimõtete ja vahenditega on võimalik teadustööd teha mitmel viisil paremini korratavaks: näiteks eelregistreeritud uuringute, avatud laborimärkmike, avatud lähtekoodi või avaandmete kaudu. Avaandmete loomine teadustöös või ka näiteks riigiametites on saamas peaaegu normiks, ilma milleta on raske analüüsi tõsiselt võtta. Just digitaal- sed vahendid teevad selle võimalikuks.

Eestisse ilmus digihumanitaaria (ka digitaalhumanitaaria) temaatiliste konverentside kaudu, mille algatasid Eesti Kirjandusmuuseumi uurijad aastal 2013.¹ Siiski

¹ Konverentside sisu on leitav Eesti Digitaalhumanitaaria Seltsi lehelt (vt DH Estonia).

on ka Eesti digihumanitaarial pikem ajalugu. Arvutitele toetuvast uurimistööst meenub näiteks formaalset grammatikat luua püüdnud Generatiivse Grammatika Grupp (1965–1972), mis viis välja keelekorpuste koostamiseni 1980. aastate lõpus ja tänapäeva arvutuslingvistikani; juba 1960-ndatel tehti Juhan Kahki eestvõttel katsetusi arvutite kasutamiseks ajalooteaduses (vt Plath jt siin numbris) jne. Erinevate valdkondade andmebaaside ja kogude loomise või analüüsiga tõusid esile mitmed uurijad, näiteks Aadu Must ajalooteaduses (vt ka Puss siin numbris) või Arvo Krikmann folkloristikas, keda võib õigusega pidada tänapäeva digihumanitaaria pioneerideks Eestis.

Eesti digihumanitaaria kogukonna hetkeseisust saab hea ülevaate konverentsikavade esinejaid ja teemasid vaadates. Digihumanitaaria Eestis hõlmab üpris erinevaid uurijaid, keda ühendab humanitaarteaduste taust või uurimistemaatika. Juba esimesel konverentsil olid esindatud dialektoloogia, folkloristika, arheoloogiamalestised, tekstitöötlus, geoinformaatika, interaktiivsed keskkonnad ja andmebaaside loomine. Märksõna *digihumanitaaria* on aidanud end seostada rahvusvaheliste traditsioonidega ning leida eeskujusid ja kaasamõtlejaid teistelt uurimisväljadelt. Võib ilmselt väita, et kui esimestel konverentsidel jäid kõlama andmebaaside ja kogude ülesehitamisega seotud küsimused, siis üha enam on laienenud teadlaste ring, kes digihumanitaaria meetodeid teadustöös rakendavad.

Andmepõhine humanitaaria

Igasugune teadustöö toetub üldiselt andmetele ehk millelegi antule (lad *data* 'asjad, mis on antud'). Humanitaarteadustes ei ole tihti seda mõistet tähtsustatud, kuna kasutusel on juba olnud üksikasjalikumad terminid, näiteks *primaar- ja sekundaarallikad* ajalooteaduses, samuti ei ole selge, kuivõrd vaadeldav on „antud” või on see pigem vaatleja loodud (vt Harrower jt 2020). Andmeteks saab miski vaadeldav, kui seda kasutatakse analüüsi alusena (olgu selleks fototöötlus, tekstikogu või lohk rändrahnus). Digitaalsed vahendid võimaldavad asuda seda vaadeldavat koondama suuremateks organiseeritud andmekogudeks, mis säilitaks vajaliku info ja mida oleks kerge kasutada edasistes uuringutes. Andmepõhises humanitaarias saab töö andmestikega – nende loomine, analüüs ja kriitika – loomuldasu uurimise põhiosaks.

Andmete hulk, mis on hästi ligipääsetav just digitaalselt, kasvab aina kiiremas tempos. Igas minutis kirjutatakse maailmas 500 000 avalikku lühikommentaari, saadetakse 16 miljonit tekstisõnumit, 156 miljonit e-kirja. Suhtlus on kolinud digitaalsesse kanalisse, trükimeedia ja raamatud valmivad esmalt digitaalsel kujul. Kogu seda materjali on võimalik traditsiooniliselt lugeda või luua sellest väiksemaid ja suuremaid andmekogusid, kus võiks saada täielikuma ülevaate kirjutatust. Suhteliselt hea pildi eesti kultuuriajaloo kohta võiks anda näiteks Eesti Rahvusringhäälingu arhiiv, kultuurilehtede arhiivid või ka Keeles ja Kirjanduses ilmunud artiklid.

Viimased paarkümmend aastat on aktiivselt tegeletud maailma kultuuri-pärandi ülespildistamise ja digiteerimisega. Praeguseks on mõningate hinnangute

järgi ilmunud maailmas üle 130 miljoni raamatu (vt Taycher 2010), millest näiteks Google Books on üles pildistanud üle 40 miljoni nimetuse (vt Lee 2019). Nii-samuti kui mujalgi tegeletakse Eestis aktiivselt digiteerimisega: trükipärandist on digiteeritud rohkem kui 25%, alustades tähtsamatest, kanoonilisematest autoritest ja allikatest. Pildistatud allikates tuvastatakse tekst, klassifitseeritakse artikliteks ja tekstiosadeks, lisatakse metainformatsioon kus võimalik: ilmumise aeg, autor, väljaanne jne. Selle tulemusena on võimalik koondada suur hulk tekste ülevaatile uurimuste tarbeks. Kõik digiteeritud eksemplarid ja kogud ei ole küll kergesti kättesaadavad, aga nende ligipääsetavuse nimel töötavad mitmed asutused. Näiteks Rahvusraamatukogu perioodikakogust on võimalik kohaliku ajakirjanduse vahendusel saada hea pildi siinse ühiskonna arengutest aastatel 1860–1940.² Suured kogud on Eesti Kirjandusmuuseumil (sh Eesti Rahvaluule Arhiivil), Rahvusarhiivil jt mäluasutustel. Massdigiteerimine ning kultuuripärandi laiem kättesaadavus on ka Kultuuriministeeriumi „Kultuuripärandi digiteerimise tegevuskava” eesmärk. Kui kogude kättesaadavus pole enam probleemiks, on probleemiks siiski see, millisel viisil kogud on kättesaadavad: kas tekstiandmed on tärgtuvastatud (OCR, *Optical Character Recognition*), kas tärgtuvastus on kvaliteetne, millises formaadis ja kodeeringus on tekstiandmed esitatud ning millised on piirangud nende kasutamisel. Uurija seisukohast on suurim väljakutse oskus varieeruva ja mitmekesise andmestikuga efektiivselt toime tulla, kasutades erinevaid digimeetodeid ja -vahendeid. Võib üsna julgelt öelda, et mõlemal poolel – nii mäluasutustel kui ka teadlastel – on selles osas veel väga palju õppida.

Kui massdigiteerimine võimaldab andmeid üles pildistada ning tärgtuvastusega on võimalik trükitud teksti tuvastada, siis keerukam on käsikirjaliste tekstidega, mida arhiivides on väga palju. Masinõppe meetoditel põhinev käekirjatuvastus, mida digihumanitaaria algatuste käigus on loodud (HTR, *Handwritten Text Recognition*; vt Transkribus), ei ole leidnud laialdaselt teed arhiividesse. Rahvusarhiiv on kasutanud sel eesmärgil üsna edukalt vabatahtlike abi ühisloomeks (ingl *crowdsourcing*) – arhiiv on loonud platvormi, kus vabatahtlike töö ja taibu abil kirjutatakse arhiivimaterjalid süstemaatilisel kujul masintöödeldavaks tekstiks (Rahvusarhiivi ühisloomealgatused, nagu eestlased Esimeses maailmasõjas, Vabadussõjas osalenud, Tartu 1867, tsaariaegsete vallakohtuprotokollide sisestamine; vt Rahvusarhiiv). Ühisloome korras on vanadele fotodele geoinfo märkide lisamisega tegelenud Ajapaik.

Kui suurimad kogud on tekstidest, siis oma filmiandmestikku on üles ehitanud Eesti Filmiarhiiv, teoste ülespildistamisega tegelevad kunstimuuseumid ja arheoloogid – oluline osa on seal ka esemete või hoonete 3D-mudelite loomisel fotode või laserskannerite abil. Näiteks Rahvusraamatukogu digitaalarhiivis on üle 16 000 plakati, 4000 kaardi, 7000 postkaardi (vt Digar). Rahvusarhiivi fotoinfosüsteemi on kantud üle 876 000 kirje, millest rohkem kui 714 000 fotol on olemas ka digipilt (vt Fotis). Digiteeritud esemeid on Eesti Rahva Muuseumis, Eesti Kirjandusmuuseumis, Tallinna Ülikooli Akadeemilises Raamatukogus jm. Fotod, maalid, arheoloogilised

² Ülevaadet selle kohta vt <http://data.digar.ee/>

leiud on koondatud muuseumide infosüsteemi (vt Eesti muuseumide veebivärv), kus need on seotud hulga metainfoga. Ka nende andmetega ringikäimine nõuab spetsiifilisi teadmisi ja oskusi, mis traditsiooniliselt ei ole olnud humanitaarerialade hariduse osa.

Digiteerimise võidukäik kultuuripärandiga ümberkäimisel on toimunud. Arusaadavalt alustati materjalide säilitamisest ja arhiveerimisest ning heal juhul ka andmebaasi moodustamisest. Alles viimastel aastatel on rõhutatud ka selle digiteeritud ja digitaalselt sündinud kultuuripärandi kasutuselevõtu vajadust teadustöös ning loodud selleks ka võimalusi.³ Just siin pakub digihumanitaaria vahendeid ja oskusi materjale digimeetoditega uurida ning uurijakogukonda, kes mõistab sellise tööga seotud võimalusi, aga ka väljakutseid, mida tuleb asutustevahelises koostöös ületada.

Andmetega töötamiseks on vaja vahendeid ja meetodeid

Erinevat päritolu ning eri „valmidusastmega” andmete töötlemine ja analüüs nõuavad oskusi. Digihumanitaaria arengu käigus on tekkinud väga palju vahendeid, mis enamasti on mõeldud üht tüüpi andmete või ühe andmekogu töötlemiseks ja analüüsiks (neid on vahel püütud ka loetleda, vt nt Gardiner, Musto 2015). Eri tüüpi andmetega toimetulekuks on vaja vähemalt minimaalset programmeerimisoskust. See on kahtlemata praegu digihumanitaaria kõige suurem väljakutse: humanitaarteaduste õpetamistraditsiooni ei ole seni kuulunud programmeerimis- või kvantitatiivsed meetodid. Sellest tulenevalt on käesoleva teemanumbri üheks eesmärgiks lugejatele konkreetsete meetodite tutvustamine, millele on pandud tavapärasest rohkem rõhku (vt nt Pilvik jt ja Šeļa siin numbris).

Digimeetodid võimaldavad andmetepõhist makrotasandi vaadet, nt nn kauglugemise uurimustes (vt Meriste siin numbris). Tasub meeles pidada, et head kvantitatiivset uurimust ei ole lihtsam teha kui traditsioonilisi humanitaaria kvalitatiivseid (lähilugemis)meetodeid kasutatavat uurimust. Arvutuslikud meetodid võimaldavad saavutada suuremat täpsust ning kaasata suuremat andmehulka, aga ainult juhul kui meetodite toimimist lähedalt kontrollitakse ning andmestiku iseärasustega arvestatakse. Näiteks tekstide automaattöötamise vahendid on enamasti loodud tänapäeva keelele ja tekstižanridele. Kui nad võimaldavadki suhteliselt hästi kätte saada filmiarvustuse tonaalsust,⁴ siis ei ole üldse kindel, et nad sama hästi töötavad XIX sajandi realistliku romaani peatükkide peal. Veelgi suuremat peavalu võib valmistada tänapäeva keele jaoks tehtud vahendite rakendamine XIX sajandi vallakohtuprotokollidele, mis on vajalik Rahvusarhiivi ühisloome käigus täiendatava kollektsiooni analüüsivahendite loomiseks (vt katsetust Pilvik jt 2019). Hea uurimus peab seega kaasama korraga humanitaaria-alast ja tehnilist kompetentsi ning kindlustama, et valitud meetodid ja vahendid oleksid adekvaatsed ning andmetega sobivad. Nõnda on digihumanitaaria-alane uurimistöö juba olemuslikult valdkondadevaheline ning

³ Näiteks riiklik programm Eesti keel ja kultuur digiajastul 2019–2027.

⁴ Vt nt Sentiment Analysis on IMDb, kus võrreldakse eri meetodeid. Kaksikümmend parimat meetodit on üle 90% täpsusega.

vajab toimimiseks nii kvantitatiivseid kui ka kvalitatiivseid meetodeid, rakendagu neid üksainus mitmekülgset pädev uurija või suurem uurimisrühm (Tolonen jt 2020). Kas sellist uurimistööd tuleks teha just digihumanitaaria katusnimetuse all, on eraldi küsimus (vt Ibrus jt siin numbris).

Kui tehnilised valdkonnad püüavad pigem luua aina paremaid algoritme, siis digihumanitaaria eesmärk on leida nende abiga aina põnevamaid uurimistulemusi, mis kannataksid ka igakülgset kriitikat (Nguyen jt 2020). Seetõttu võiks digihumanitaariat pidada ka autonoomseks uurimisväljaks: naaberdistipliinid ei tee digihumanitaaride eest tööd ära ei algoritmide rakendamisel, ülesannete püstitamisel ega tulemuste kriitilisel tõlgendamisel. Tekstide lugemine toimub üha enam digitaalsete abivahenditega ning andmed tunduvad mängivat ühiskonnas üha kesksammat rolli. Raivo Ruusalepp (2020) näiteks arvab, et kultuuripärandi andmete puhul võiks tulevikus, kui digioskused muutuvad veelgi tavalisemaks, arvestada juba kodanikanalüütikutega. Humanitaarteadlastel on siinkohal võimalik valida, kui palju nad sellistes arengutes kaasa tahavad rääkida.

Digimeetodite integreerimine humanitaarteadlaste ettevalmistusse on väljakutse ennekõike ülikoolidele. Õppimisvõimalused humanitaarerialade üliõpilaste jaoks on loodud Tartu Ülikoolis, Tallinna Ülikoolis, Tallinna Tehnikaülikoolis; alguse on saanud mõned uurimisrühmad, kus digihumanitaaria-alane kompetents on keskne (nt CUDAN Tallinna Ülikoolis). Neli aastat on korraldatud kraadiõppuritele suvekoole „Digitaalsed meetodid humanitaarias ja sotsiaalteadustes”, kus umbes sada osalejat aastas on omandanud uusi digioskusi. Konkreetsete distsipliinide sees – näiteks korpuslingvistikas või arvutuslikus arheoloogias – on need meetodid olnud pikemat aega kasutusel, peamiselt tingituna uurimisrühmade või üksikuurijate huvist ja vajadusest. Teema on akuutne ka sotsiaalteadustes ning meetodite ühisosa humanitaaria ja sotsiaalteadustes on väga suur. Seetõttu sobib õpik „Kuidas mõista andmestunud maailma? Metodoloogiline teejuht” (Masso jt 2020) suurepäraselt ka humanitaarteadlastele. Tuleb küll rõhutada, et edukaks tööks digihumanitaaria vallas ei piisa esmasest tutvumisest mõne meetodiga; süstemaatiliselt on vaja arendada programmeerimis- jm tehnilisi oskusi. Teisalt võib sõltuvalt eesmärgist mõne uue meetodi või vahendi tundmine oluliselt aidata edasi andmete kogumisel, töötlemisel, analüüsil või tulemuste visualiseerimisel. Ka käesolevas ajakirjas näeme üsna erinevaid lähenemisi andmete töötlemisel.

Eesti digihumanitaaria hetkeseisu võib ehk veel kirjeldada kui ärksat algusaega. On näha esimesi märke digioskuste süstemaatilise kaasmisest ka humanitaarvaldkonnas ning edukaid üksikuurimusi sel teemal. Arhiivkogud ja muuseumid on võtnud suuna andmekogude ligipääsu parandamisele ning uurijad omandavad häid avatud teaduse praktikaid. Näha on alustavaid valdkondadevahelisi uurimisrühmi, mis püüavad keeruliste humanitaaria küsimustega tegeleda, kaasates korraga nii digioskusi kui ka oma kitsama valdkonna traditsioone (nt Plath jt; Ibrus jt; Ojamaa jt siin numbris). Selles, kas seda tööd teha just digihumanitaaria nime all, oleme teemanumbri koostajatena optimistlikumad kui Ibrus jt siin numbris ja leiame, et see märksõna alles kinnitab Eestis kanda ning me vajame seda kasvuraskuste ületamiseks. Ent ükskõik mis sildi all, on kõige tähtsam sisuline töö. Lootus ja võimalus

on, et lähematel aastatel on digioskused humanitaarias nii tavalised, et ei pea enam *digi*-t eraldi esile tooma. Meile endale tundub, et nende oskuste juurutamisest on humanitaarial palju võita.

Ülevaade artiklitest

Teemanumbri soovime eesti lugejani tuua ülevaateid digihumanitaaria rakendusvõimalustest eri distsipliinides ning näitlikustada neid võimalusi juhtumiuuringutega. Digihumanitaaria hetkeseisus Eestis iseloomustab seegi, et meie üleskutse peale esitati neli korda rohkem teese, kui number mahutab, mistõttu saime esile tuua ainult mõned Eestis viljeletud teemad. Valikutes lähtusime peamiselt sellest, et Keele ja Kirjanduse põhiteemad oleks esindatud, aga samuti tuleks esile valdkonna mitmekesisus.

Kogumikku alustab **Indrek Ibruse**, **Maximilian Schichi** ja **Marek Tamme** kaastöö „Digihumanitaariast kultuuriandmete analüüsin“, mis käsitleb digihumanitaaria valdkonna üldist problemaatikat: mis ja miks ikkagi peaks olema digihumanitaaria ja kuidas selles viljeletud teemasid võiks tulevikus uurida. Autorid kutsuvad üles valdkonnaülesele koostööle ja tõstavad esile tärgava kultuuriandmete analüüsi valdkonna, mis võiks ühendada humanitaar- ja teisi teadlasi, kes kultuurivaldkonnaga tegelevad, ning aidata kõige paremini erinevaid meetodeid ühendada.

Kogumiku teises artiklis „Erinevused, kaugused ja sõrmejäljed. Stilomeetria ja mitmemõõtmelise tekstianalüüsi alused“ annab **Artjoms Šeļa** ülevaate eesti keeles seni vähe tutvustatud valdkonnast – stilomeetriast, mis püüab kvantitatiivsete vahenditega mõista ja kirjeldada autorite stiile ning tuvastada stiilitunnuseid konkreetsetes tekstides. Autor teeb näidiskatse eestikeelses minikorpuses, tutvustades automaatse autorituvastuse põhimõtteid ja nüansse, millega peab selle juures arvestama ning kuidas mitmemõõtmelise tekstianalüüsi tulemusi lugeda. Lisaks stilomeetria arengu ajaloolisele ülevaatele näitab autor, millist kasu võib kirjandusteadlastel olla tekstide kvantitatiivsetest esitustest nii autorituvastusel kui ka üldiste kirjanduslike suhete (nt žanr, kronoloogiline muutus jne) modelleerimisel.

Kvantitatiivsete meetodite kasutamisest kirjandusteaduses räägib ka **Maarja Helena Meriste**, kes artiklis „Kirjanduskeele kauglugemine. Kirjandust väärindava metodoloogia arendamine tsiteerimisvõtte näitel“ esitab kauglugemise vallas tehtud juhtumiuuringu avatud tsitaatide esinemisest valitud eesti ilukirjandusautorite tekstides. Kauglugemine on raamistik, mis ei vaatle kirjandust eraldiseisvate üksikteoste kogumina, vaid tekstuaalse väljana. Artiklis vaadeldakse avatud tsitaati kui teksti tähendusstruktuuri tugevalt mõjutavat stiilivõtet ning leitakse, et avatud tsitaat loob ilukirjandusteksti tähendusreljeefi, mille abil tuleb esile lai seostevõrgustik.

Digitalse lugemise kui humanitaarteaduste keskse mõiste teemal jätkab teemanumbri neljas artikkel „Digitaalne lugemine ja humanitaarharidus“, kus **Maarja Ojamaa**, **Peeter Torop**, **Aleksandr Fadeev**, **Alexandra Milyakina** ja **Merit Rickberg** arutlevad nii sõnaliste tekstide lugemise kui ka tekstide vaatamise ja kuulamise olulisuse üle tänapäeva humanitaarhariduses. Autorid rõhutavad, et üleminek

raamatukeskselt kultuurilt veebikesksele kultuurile on kaasa toonud tekstide ja tähenduste pihustumise. Artiklis tutvustatakse sidususe loomise ja tervikuks seostamise mehhanisme (nt arvestades teksti/platvormi väljenduslikkuse, piiritletuse ja struktureerituse parameetreid), mis on hariduslikust vaatepunktist eriti oluline. Peale teoreetiliste lähtekohtade tutvustamise kirjeldatakse konkreetseid näited digitaalsete vahendite kasutamisest õppematerjalide loomisel, mis toetavad temaatiliste tervikukujutelmade tekkimist.

Kaks järgmist artiklit esindavad keeleteaduse valdkonda, kuid on erineva rõhuasetusega. **Margit Langemets, Kristina Koppel, Jelena Kallas** ja **Arvi Tavast** mõtestavad artiklis „Sõnastikukogust keeleportaalik” leksikoloogia kui valdkonna üldisi probleeme ja lahendusi. Autorid toovad välja, et digiajastuga on sõnaraamatutöös kaasnenud laiaulatuslikud muudatused, sh tehnilised (nt liikumine ühtse sõnastikusüsteemi ja andmebaasi poole, andmete linkimine, arenenud korpuspäringusüsteemid), leksikograafilised (sõnastike ühendamine ja koostamispõhimõtete ühtlustamine, töö infokihtidega, andmete dünaamilisus) ja töökorralduslikud (kollektiivsuse kasv, avatuse suurenemine). Peale valdkonna üldiste muutuste ja digiajastu väljakutsete kirjeldavad autorid EKI uusimat sõnakogu – EKI ühend-sõnastikku – ja selle koostamisel rakendatud Ekilexi andmemudelit.

Teises keeleteaduslikus kaastöös „Murded, varieerumine ja korpusandmed. Eitussõna paiknemine võru ja seto eituslausetes” näitlikustavad **Maarja-Liisa Pilvik, Helen Plado** ja **Liina Lindström** ühe konkreetse uurimisküsimuse lahendamist digihumanitaaria meetodiga. Autorid kirjeldavad ja seletavad keeles esinevat varieerumist (ees- ja tagaeitus lõunaeesti seto ja võru keelevariantides), kasutades meetodeid, mida on võimalik edukalt rakendada ka teistes humanitaaria suundades, nt erinevate klassifitseerimisülesannete lahendamiseks. Artiklis antakse ülevaade mitmemõõtmelisest statistilisest analüüsist – meetod, mis mängib korpuspõhises kvantitatiivses dialektoloogias kesksel rolli.

Kaks viimast kaastööd esindavad otsapidi ajaloo valdkonda. **Fred Puss** annab artiklis „Kakasonist Kaljulaks. Perekonnanimede eestistamise andmebaasi analüüs” ülevaate perekonnanimede eestistamise senisest uurimisest, kirjeldades muuhulgas probleeme nimemuutumiste andmestikega töötamisel. Autor rõhutab, et nimede eestistamise andmebaasi kasutusvõimalusi on palju ning olemasolevate naaberandmestike ning digihumanitaarse oskuste ja võimaluste arenedes tuleb neid veelgi juurde. Kvantitatiivsete uurimismeetodite kasutamise üle ajalooteaduses arutlevad ka **Ulrike Plath, Priit Raudkivi, Kaarel Vanamõlder, Krister Kruusmaa** ja **Anna Helena Liiv** artiklis „Kuidas kodeerida kliimat? Eesti ajaloolise kliimauurimise digitaalsetest pöördet”. Artiklis on juttu kliimaajaloo uurimisest ning sellest, millal, millistes tingimustes, kontekstis ja sihiseadega leidis aset (kliima)ajalooteaduses digitaalne pööre. Autorid asetavad eesti ajaloolaste ja kliimaajaloo uurimise laiemasse digiteerimise konteksti ja näitavad, kuidas on digiteeritud andmeid võimalik kasutada tormisuse mõõtmiseks XIX sajandil.

Artiklid annavad erinevaid sissevaateid digihumanitaaria hetkeolukorda Eestis, pakkudes nii üldisemaid arutlusi digiajastuga kaasnenud väljakutsetest eri valdkondades kui ka konkreetsete uurimisküsimuste lahendusi digihumanitaaria

meetoditega. Kogumikku kaasatud artiklid ei esinda kaugeltki kogu pilti digi-humanitaariast Eestis, kuid loodetavasti on meil õnnestunud digihumanitaaria mõiste paremini avada ja panna lugeja kaasa mõtlema sellisel väljakutsuval teemal nagu mida annab *digi* juurde humanitaariale ja vastupidi. Jääme ootama, et tehtav uurimistöö annaks peagi põhjust veel mõneks temaatiliseks kogumikuks.

Artikli valmimist ning teemanumbri koostamist on toetanud Euroopa Liit Euroopa Regionaalarengu Fondi kaudu (Eesti-uuringute Tippkeskus).

VEEBIVARAD

Ajapaik. <https://ajapaik.ee>

DH Estonia. Estonian Society for Digital Humanities. <https://dh.org.ee>

Digar. Pildimaterjal. <https://www.digar.ee/arhiiv/et/pildimaterjal>

Eesti keel ja kultuur digiajastul 2019–2027. [https://www.hm.ee/sites/default/files/eesti_ keel_ ja_kultuur_digiajastul_2019-2027.pdf](https://www.hm.ee/sites/default/files/eesti_keel_ ja_kultuur_digiajastul_2019-2027.pdf)

Eesti muuseumide veebivärv. <https://www.muis.ee>

Eesti Rahvusraamatukogu avaandmete portaal. <http://data.digar.ee>

Fotis. Rahvusarhiivi fotoinfosüsteem. <https://www.ra.ee/fotis>

Kultuuripärandi digiteerimise tegevuskava. <https://www.kul.ee/kultuurivaartused-ja-digi- taalne-kultuuriparand/digitaalne-kultuuriparand/kultuuriparandi>

Rahvusarhiiv. Ühisloome. <https://www.ra.ee/kulastajale/uhisloome>

Sentiment Analysis on IMDb. <https://paperswithcode.com/sota/sentiment-analysis-on-imdb>

Transkribus. <https://readcoop.eu/transkribus>

What Is Digital Humanities? <https://whatisdigitalhumanities.com>

KIRJANDUS

Berry, David M.; Fagerjord, Anders 2017. Digital Humanities: Knowledge and Critique in a Digital Age. Cambridge: Polity Press.

Bode, Katherine; Arthur, Paul L. 2014. Collecting ourselves. – Advancing Digital Humanities. Toim P. Arthur, K. Bode. London: Palgrave Macmillan, lk 1–12.

Gardiner, Eileen; Musto, Ronald G. 2015. The Digital Humanities: A Primer for Students and Scholars. Cambridge: Cambridge University Press.

Harrower, Natalie; Maryl, Maceij; Biro, Timea; Immenhauser, Beat 2020. ALLEA Working Group E-Humanities. Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. Berlin: ALLEA – All European Academies.

Kirschenbaum, Matthew G. 2010. What is digital humanities and what's it doing in English departments? – ADE Bulletin, nr 150, lk 55–61.

Kirschenbaum, Matthew G. 2012. Digital humanities as/is a tactical term. – Debates in the Digital Humanities. Toim Matthew K. Gold. Minneapolis–London: University of Minnesota Press, lk 415–428.

- Laak, Marin; Viires, Piret; Sarv, Mari 2020.** Digikultuuri väljakutseid. Kultuuripärandi kasutamise humanitaarteaduses ja hariduses. – Methis. *Studia humaniora Estonica*, nr 26, lk 5–16.
- Lee, Haimin 2019.** 15 years of Google Books. 17. X. <https://blog.google/products/search/15-years-google-books> (2. IX 2021).
- Masso, Anu; Tiidenberg, Katrin; Siibak, Andra (toim) 2020.** Kuidas mõista andmestunud maailma? Metodoloogiline teejuht. (Gigantum Humeris.) Tallinn: Tallinna Ülikooli Kirjastus.
- Nguyen, Dong; Liakata, Maria; DeDeo, Simon; Eisenstein, Jacob; Mimno, David; Tromble, Rebekah; Winters, Jane 2020.** How we do things with words: Analyzing text as social and cultural data. – *Frontiers in Artificial Intelligence*, kd 3, lk 62.
- Pilvik, Maarja-Liisa; Muischnek, Kadri; Jaanimäe, Gerth; Lindström, Liina; Lust, Kersti; Orasmaa, Siim; Tärna, Tõnis 2019.** *Mõistus sai kuulotedu*: 19. sajandi vallakohtuprotoollide tekstidest digitaalse ressursi loomine. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, kd 15, lk 139–158.
- Ruusalepp, Raivo 2020.** Andmestuv kultuuripärand. – Eesti digikultuuri manifest. Toim Indrek Ibrus, Marek Tamm, Katrin Tiidenberg. (Bibliotheca Mediorum et Communicationis.) Tallinn: Tallinn Ülikooli Kirjastus, lk 32–38.
- Schnapp, Jeffrey; Presner, Todd 2009.** Digital Humanities Manifesto 2.0. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf (1. VI 2020).
- Taycher, Leonid 2010.** Books of the world, stand up and be counted! All 129,864,880 of you. – Google Books Search. 5. VIII. <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html> (1. VII 2021).
- Tolonen, Mikko; Mäkela, Eetu; Marjanen, Jani; Tahko, Tuuli 2020.** Arvutiteaduse kaasamine humanitaarharidusse. – Methis. *Studia humaniora Estonica*, nr 26, lk 35–50.

Peeter Tinits (sünd 1986), MA, Tartu Ülikooli digihumanitaaria ja infoühiskonna keskuse digihumanitaaria spetsialist (Jakobi 2, 51005 Tartu); Tartu Ülikooli ühiskonnateaduste instituudi tekstikaave spetsialist; Eesti Rahvusraamatukogu digihumanitaaria spetsialist, peeter.tinits@ut.ee

Jane Klavan (sünd 1983), PhD, Tartu Ülikooli maailma keelte ja kultuuride kolledži inglise keele kaasprofessor (J. Liivi 4, 50409 Tartu), jane.klavan@ut.ee

Liina Lindström (sünd 1973), PhD, Tartu Ülikooli tänapäeva eesti keele professor ning digihumanitaaria ja infoühiskonna keskuse juhataja (Jakobi 2–443, 51005 Tartu), liina.lindstrom@ut.ee

Digital humanities in Estonia

Keywords: digital humanities, data in humanities, digital methods

The article serves as a preface to the special issue on digital humanities of *Keel ja Kirjandus*. The aim of the issue is to familiarize the Estonian readers with the complex and controversial notion of digital humanities by way of examples. The issue gathers overviews of themes and methods in digital humanities along with empirical case studies representing them. We provide a brief discussion on what digital humanities is and why we need it. We consider that the core of digital humanities is the integration of digital data into humanities research. A growing amount of materials of interest to the humanities are now becoming digitized or digitally born. Archival institutions that maintain it have become recently active in facilitating their use for research purposes. Digital humanities can provide the methods, tools and the research community to take advantage of these opportunities. The digitalization of society is happening on many levels and humanities have a chance to contribute to its understanding and benefit from its opportunities. The label digital humanities itself can work as a pragmatic banner at this stage of development, potentially leading to digital skills becoming the norm also in the humanities departments at some point. The articles collected in this special issue give a good overview of the status quo of digital humanities in Estonia, representing various research directions and methods. The articles offer a variety of different perspectives and practical applications in their respective fields related to digital humanities. We believe that the area of digital humanities can advance via practical examples and an understanding of the developments of digital methods in neighbouring fields. We hope that further developments in this field will provide grounds for the compilation of other such thematic issues in the very near future.

Peeter Tinits (b. 1986), MA, University of Tartu, Digital Humanities Specialist in the Centre for Digital Humanities and Text Mining Specialist at the Social Studies Institute (Jakobi 2, 51005 Tartu; National Library of Estonia, Digital Humanities Specialist, peeter.tinits@ut.ee

Jane Klavan (b. 1983), PhD, University of Tartu, College of Foreign Languages and Cultures, Associate Professor of English Language (J. Liivi 4, 50409 Tartu), jane.klavan@ut.ee

Liina Lindström (b. 1973), PhD, University of Tartu, Professor of Modern Estonian and Head of the Centre for Digital Humanities and Information Society (Jakobi 2-443, 51005 Tartu), liina.lindstrom@ut.ee