

KEELERESSURSSIDE LOOMISE JA KASUTAMISEGA SEONDUVAID ISIKUANDMETE KAITSE KÜSIMUSI

Aleksei Kelli, Kadri Vider, Irene Kull, Triin Siil,
Kristen Lindén, Arvi Tavast, Age Värvi,
Carri Ginter, Einar Meister

Ülevaade. Kui keeleressursside autoriõiguslikke küsimusi on eelnevalt põhjalikult uuritud, siis isikuandmete kaitse temaatika keeleressursside kontekstis on suurelt osalt katmata. Teema on aktuaalne, sest 25. mail 2018 hakatakse kohaldama EL-i isikuandmete kaitse üldmäärust. Käesolevas artiklis analüüsitakse isikuandmete kaitse küsimusi keeleressursside loomisel ja kasutamisel teadustöö eesmärgil kahes kontekstis: 1) andmesubjekt ehk isik, kelle andmed keeleressurssis sisalduvad, on andnud oma andmete töötlemiseks nõusoleku; 2) nõusolekut ei ole ning isikuandmete töötlemine toimub teadustöö erandi alusel. Artiklis analüüsitakse keeleressursside loomisel kasutuses olevaid nõusolekuvorme ning tehakse ettepanekuid, kuidas keeleressursside kontekstis isikuandmete kaitse regulatsiooni järgida.*

Võtmesõnad: keeleressurssid, isikuandmete kaitse, isikuandmete töötlemine seaduse alusel, nõusolek isikuandmete töötlemiseks

1. Sissejuhatus

Digitaalsete keeleressursside autoriõiguslikke küsimusi on eelnevalt põhjalikult uuritud. Uuringud on katnud nii keeleressursside loomist (keeleressursside loomise autoriõiguslike aspektide kohta vt Kelli jt 2012, Tavast jt 2013) kui ka kasutamist (keeleressursside kasutamise autoriõiguslike aspektide kohta vt Kelli jt 2016a, Kelli jt 2015, Oksanen jt 2010).

Sissejuhatavalt on isikuandmete temaatikat keeleressursside valdkonnas küll markeeritud (vt Jents, Kelli 2014, Lewis jt 2017), kuid suurem töö seisab veel ees.

Euroopa Liidu põhiõiguste harta näeb ette igapäevase õiguse oma isikuandmete kaitsele (art 8). Eesti õiguses reguleerib isikuandmete töötlemist isikuandmete kaitse

* Artikli valmimist on toetatud Euroopa Regionaalarengu Fondi tegevuste "Valdkondliku teadus- ja arendustegevuse tugevdamine" (RITA programm) ja "Riikliku tähtsusega teaduse infrastruktuuri toetamine teekaardi alusel" projekti "Eesti keeleressursside keskuse (EKRK) ühendatud sisuotsing" raames.

Autorid soovivad tänada retsense ja Liisa-Maria Puuri advokaadibüroost Sorainen väärtuslike kommentaaride eest, mis aitasid parandada artikli kvaliteeti. Võimalike vigade eest on autorid siiski ise vastutavad.

seadus (edaspidi IKS), mis tugineb isikuandmete kaitse direktiivile. Isikuandmete kaitse direktiivi tunnistab kehtetuks 25. maist 2018 kohalduv EL-i isikuandmete kaitse üldmäärus¹ (edaspidi ka üldmäärus või ÜM). Ühelt poolt võib nõustuda valdkonna ekspertidega, et ÜM võib oluliselt mõjutada keeleuuringuid kogu Euroopas (Lewis jt 2017: 60). Teisalt isikuandmeid sisaldavate keeleressursside loomine ja kasutamine on allutatud isikuandmete kaitse regulatsioonile nii praegu kehtivas kui ka tulevases regulatsioonis. Artikli kirjutamisel võetakse arvesse nii kehtivat kui ka uut regulatsiooni.

Artikli eesmärgiks on sissejuhatavalt selgitada, millistele isikuandmete kaitse nõuetele peab keeleressursside loomine ja kasutamine vastama. Üldmääruse rakendamise eeldab siseriikliku rakenduseaduse vastuvõtmist.² Artikli kirjutamise hetkel on võimalik tugineda isikuandmete kaitse seaduse eelnõu 6.11.2017 versioonile (IKS uus eelnõu 2017). Kuna eelnõu ei ole lõplik ning puudub ka ÜM rakenduspraktika, siis võivad artiklis olevad seisukohad erineda tulevases praktikast.

Artikli suunitlus on praktiline ning peaks aitama nii keeleressursside loojatel kui ka kasutajatel hinnata isikuandmete kaitse raamistikku ja valida sobiv käitumismudel.

Isikuandmete kaitse keskseks mõisteks on isikuandmed, mida ÜM määratleb järgnevalt: “igasugune teave tuvastatud või tuvastatava füüsilise isiku (“andmesubjekti”) kohta; tuvastatav füüsiline isik on isik, keda saab otseselt või kaudselt tuvastada, eelkõige sellise identifitseerimistunnuse põhjal nagu nimi, isikukood, asukohateave, võrguidentifikaator või selle füüsilise isiku ühe või mitme füüsilise, füsioloogilise, geneetilise, vaimse, majandusliku, kultuurilise või sotsiaalse tunnuse põhjal” (art 4). Kuna isikuandmed on füüsilise isikuga seonduv teave, siis hakkab ÜM regulatsioon laienema mitmetele keeleressurssidele (nt keeleressursid, mis sisaldavad suulist kõnet, isikuandmeid sisaldavad blogid ja artiklid).

Teine põhimõiste isikuandmete kaitse raamistikus on isikuandmete töötlemine, mida ÜM määratleb järgmiselt: “isikuandmete või nende kogumitega tehtav automatiseeritud või automatiseerimata toiming või toimingute kogum, nagu kogumine, dokumenteerimine, korrastamine, struktureerimine, säilitamine, kohandamine ja muutmine, päringute tegemine, lugemine, kasutamine, edastamine, levitamise või muul moel kättesaadavaks tegemise teel avalikustamine, ühitamine või ühendamine, piiramine, kustutamine või hävitamine” (art 4). Juhul kui keeleressurssis sisalduvad isikuandmed, siis on selle loomine ja kasutamine käsitletav isikuandmete töötlemisena. Isikuandmete töötlemine võib toimuda nii nende loomisel kui ka kasutamisel. Lihtsustatult öeldes saab isikuandmete töötlemine toimuda

- 1) isiku nõusolekul või
- 2) õigusaktides reguleeritud juhtudel (nt isikuandmete töötlemine teaduslikul eesmärgil).³

Isikuandmete töötlemist keeleressursside loomisel ja kasutamisel vaadeldakse eelkõige teadustöö kontekstis, mis on allutatud eriregulatsioonile (nn teadustöö erand), kuid andmete töötleja peab olema valmis tõendama, et töötlemine toimub just teaduslikul, mitte mingil muul eesmärgil.

Kehtiv isikuandmete kaitse seadus ja uus isikuandmete kaitse seaduse eelnõu (2017) kasutavad oma regulatsioonis mõistet “teadusuuring” (IKS § 16, IKS uus eelnõu § 11). Kehtiva IKS eelnõu (2006) seletuskirja kohaselt tuleks seda mõistet

¹ <http://eur-lex.europa.eu/legal-content/ET/TXT/HTML/?uri=CELEX:32016R0679&from=ET> (23.3.2018).

² Oluline on meeles pidada, et riigisiseseid õigusakte tuleb tõlgendada ja rakendada kooskõlas EL-i õigusega (vt Ginter 2006: 923).

³ ÜM artikkel 6 nimetab isikuandmete töötlemise alustena nõusolekut, lepingu täitmist, juriidilise kohustuse täitmist, füüsilise isiku eluliste huvide kaitsmist, avalikes huvides oleva ülesande täitmist, õigustatud huvi.

tõlgendada laialt, viidates teadus- ja arendustegevuse korralduse seaduse (edaspidi TAKS) § 2 punktis 7 defineeritud “teadustegevuse” mõistele⁴, mitte kitsalt TAKS § 2 punktis 1 nimetatud alusuuringutele või sama paragrahvi punktis 6 defineeritud rakendusuuringutele (IKS eelnõu 2006). Üldmäärus lähtub samuti teadusuuringu laiast määratlusest ning see peaks muuhulgas katma “tehnoloogiaarendust ja tutvustamistegevust, alusuuringuid, rakendusuuringuid ja erasektori vahenditest rahastatavaid uuringuid” (põhjenduse punkt 159).

Kehtiva IKS eelnõu (2006) seletuskirja kohaselt peaks üldjuhul teadustöö erand laienema ainult teadus- ja arendustegevuse korralduse seaduse §-s 3 määratletud teadus- ja arendusasutustele. Tekib küsimus, kas taoline lähenemine on absoluutne (eriti ÜM valguses) või saab lähtuda ka laiemast isikute ringist.

Autorite hinnangul tuleks keeleressursside arendamise vajadusi silmas pidades eelistada laiemat isikute ringi, nii et ka Eesti Keeleressursside Keskusel kui Tartu Ülikooli, Tallinna Tehnikaülikooli ja Eesti Keele Instituudi poolt moodustatud konsortsiumil (seltsingul) oleks võimalik õiguspäraselt töödelda isikustatud keeleressursse IKS § 16 alusel (IKS eelnõu 2017 § 11). Erialakirjanduses on leitud, et üldmäärus ei piiritle õigustatud isikute ringi ning teadustöö erand laieneb teoreetiliselt ka ärilistele teadusuuringutele (ingl *commercial research*) (Pormeister 2017a: 138). Üldmääruse rakendamise juhtnöörides asutakse samas seisukohale, et teadustöö mõiste sisustamisel ei saa minna kaugemale selle tavatähendusest. Teadustöö (ingl *scientific research*) tähendab uurimisprojekti, mida teostatakse kooskõlas valdkondliku metodoloogia ja eetikastandardiga (WP 29 2017: 27).

Enne isikuandmete töötlemise aluste juurde minemist avatakse artiklis keeleressursside liikide üldine raamistik. Seejärel analüüsitakse isikuandmete töötlemist isiku nõusoleku alusel (täiendavaks infoks isikuandmete töötlemisest nõusoleku alusel vt Pormeister 2017). Seni on keelejuhtidelt isikuandmetega seotud nõusolekute hankimisel tähelepanu pööratud peamiselt suulise keele korpusi vm suulise keele andmestikke kogudes, sest selliste korpuste kogumine on otseselt seotud heli või video salvestamisega, mis ilma salvestatava isiku nõusolekuta on üldjuhul keelatud. Lisaks raskendab suulise keele salvestiste kogumist asjaolu, et Euroopas loetakse inimese hääl tema isikut tuvastavate biomeetriliste andmete hulka.⁵ Nõusoleku alusel toimuva töötlemise analüüsimiseks oleme palunud teadlastelt mõned näited Eestis keeleressursside kogumisel kasutatud nõusolekutest ja kokkulepetest (vt tabel 1).⁶

Artikli jaoks kogutud nõusolekuvormid jagunevad sisu poolest kahte liiki. Esiteks nõusolekuvormid, mis sisaldavad lisaks nõusolekule ka isikuandmete ankeeti (nr 2, 3, 7) ning teiseks nõusolekuvormid, mis sisaldavad ainult nõusolekut isikuandmete töötlemiseks (nr 5, 6, 9). Lisaks tuleb eristada nõusolekuvorme, milles antakse nõusolek kasutada andmesubjekti poolt õppeprotsessis loodud salvestusi ja sellega seotud materjale ning kus nõusolek on antud andmesubjekti suulise kõne salvestuste kasutamiseks. Artikli autorite kasutuses oli ka kaks kogutud andmete kasutamiseiga seotud kohustusi sisaldavat tahteavalduse vormi (nr 1 ja 8).

Artikli lõpuosas keskendutakse ilma nõusolekuta isikuandmete töötlemisele keeleressursside kontekstis ja teaduseesmärkidel.

⁴ Teadustegevus on TAKS § 2 punkti 7 kohaselt isiku loomevabadusel põhinev tegevus, mille eesmärk on teaduslike uuringute abil uute teadmiste saamine inimese, looduse ja ühiskonna ning nende vastastikuse toime kohta.

⁵ ÜM art 4 p 14 defineerib biomeetrilisi andmeid kui “konkreetse tehnilise töötlemise abil saadavad isikuandmed isiku füüsiliste, füsioloogiliste ja käitumuslike omaduste kohta, mis võimaldavad kõnealust füüsilist isikut kordumatult tuvastada või kinnitada selle füüsilise isiku tuvastamist, näiteks näokujutis ja sõrmejälgede andmed”.

⁶ Kuna keeleressursside uuringud on rahvusvahelised, siis erialakirjanduses on tehtud ka ettepanek, et Euroopa Komisjon ja ELRA (European Language Resources Association, täiendav info: <http://www.elra.info/en/about/> (23.3.2018)) võiks ÜM-i järgides töötada välja nõusoleku tüüpvormid teadusandmete taaskasutamiseks. Lõppastmes peaks see viima tähistesüsteemini, mis on sarnane Creative Commons'i logodega (Lewis jt 2017: 63–64).

Tabel 1. Nõusoleku ja konfidentsiaalsuskokkulepete näiteid⁷

Nr	Kogu või üksus	Dokumendi nimetus	Autorid	Märkused ja kommentaarid
1.	TÜ suulise ja arvutisuhtluse labori salvestus- ja tekstikogu suulise eesti keele korpus	Kasutamise konfidentsiaalsus-kokkulepe	Tiit Hennoste	Andriela Rääbis: "Eeskujuga polnud tol ajal kusagilt võtta, vormid on enda koostatud ja ülikooli jurist vaatas üle. Hiljem on seda veidi kaasajastatud."
2.	TÜ suulise ja arvutisuhtluse labori salvestus- ja tekstikogu suulise eesti keele korpus	Õiguste üleandmine õppeprotsessis loodud salvestus(t)e ja sellega seotud materjalide kohta	Tiit Hennoste	
3.	TÜ rakenduslingvistika osakonna keele-ressursside andmebaasi õppijakeele korpus	Nõusoleku-kinnitus	Kristiina Praakli, täiendanud ja parandanud Raili Pool	Varasema nõusoleku-kinnituse (aastast 2012) alus-teksti autor Kristiina Praakli, vaatas üle ka ülikooli jurist.
4.	TTÜ Kübl laste ja noorte kõnekorpus	Salvestusel osalemise nõusolek lapsele, lapsevanemale, õpetajale; õpilase ankeet		Einar Meister: "Algselt oli lastekorpuse nõusoleku tekst pisut pikem ja keerulisem, selle vaatas üle ja kiitis heaks jurist. Laste jaoks oli see liialt keerulise sõnastusega ja me tegime ise selle pisut lühemaks ja lihtsamaks."
5.	TTÜ Kübl aktsendikorpus	Nõusolek ja keelejuhi ankeet		
6.	TTÜ Kübl eesti keele foneetiline andmebaas	Tõend tekstide sisselugemise kohta		Projekti "BABEL: A Multilingual Database" raames ja selle tingimustega nõustudes.
7.	TÜ eesti keele spontaanse kõne foneetiline korpus	Keelejuhi ankeet (sisaldab nõusolekut)		
8.	TÜ eesti keele spontaanse kõne foneetiline korpus	Kasutamisingimused		Eeskujuks Hennoste tehtud konfidentsiaalsuskokkulepe (nr 1), tugevasti kärbitud
9.	TÜ uurimisprojekt "Eesti keele töötuse uurimine"	Uuritava informeeritud nõusoleku leht		Pärtel Lippus: "Ühes koos psühholoogidega läbi viidud katses tegime eraldi taustaandmete ankeedi ja põhjalikuma tekstiga nõusolekuvormi, tegelikult sellel on oma point, sest siis on katseisiku kood ja nimi ainult nõusolekulehel."

2. Keeleressursside olemus

Keeleteadlase või keeletehnoloogi vaatenurgast jagunevad digitaalsed keeleressurssid iseloomu poolest kaheks: keeleteaduslikuks andmestikuks ja tarkvaraks (vt ka Vider jt 2013).

- 1) Keeleteaduslik andmestik:
 - a) korpused ehk keelematerjali kogumid;
 - b) leksikonid ehk sõnavara kogumid;
 - c) keelemudelid (näiteks statistiline keelemudel või grammatika).
- 2) Tarkvara, mis keeleandmestikku suuremal või vähemal määral töötleb ja kasutab (näiteks annoteerib, loob sisendfaile järgmise masintöötamise etapi jaoks, loob väljundit lõppkasutaja jaoks, olgu lõppkasutajaks IT-süsteem või inimene):
 - d) iseseisvad programmid või lausa infotehnoloogilised terviksisüsteemid;
 - e) tarkvarajupid või teegid.

2.1. Tundlik keeleandmestik

Isikuandmete kaitse seisukohalt on peamiseks huviobjektiks (1a) ehk keeleline toormaterjal, mida allpool nimetame tundlikuks keeleandmestikuks.

Tundlik keeleandmestik esitab inimeste keelekasutuse juhtumeid ja koosneb kirjalikust tekstist, heli-, video- või muudest salvestistest. Sellesse kuuluvad materjalid sisaldavad sageli nii isikuandmeid kui ka autoriõigusega kaitstud teoseid. Lisaks võivad nad sisaldada konfidentsiaalsuspiirangutega dokumente ja kommertskasutuse keeluga materjali. Need materjalid on tundlikus keeleandmestikus esitatud äratundmist ja reprodutseerimist võimaldaval kujul, näiteks võib tekstikorpuses sisalduda terve romaan või oluline osa sellest.

Isikuandmeid võib keeleandmestikus sisalduda kahel viisil.

- 1) Andmed keele “tootja” ehk keelejuhi või informandi isiku kohta (peamiselt sugu ja vanus, aga ka elukoht, sünnikoht jms), mis lisatakse korpusele meta-andmetena. Kõneldud keele salvestise puhul kuuluvad isikuandmete hulka ka keelejuhi hääl jm biomeetrilised andmed. Isikuandmete kaitse kontekstis on keelejuhid ja informandid peamised andmesubjektid.
- 2) Keelelise materjali enda sisu. Näiteks vestluses mainitakse kellegi isikuandmeid või on koha, aja jm jutus leiduvate parameetrite põhjal võimalik kellegi isikut tuvastada.

Korpuse koostatakse juba olemasolevate tekstide või salvestiste kogumise või, eriti kõneldud keele puhul, uute salvestiste tegemise teel. Järgmises etapis võidakse (enamasti materjali koguja enda poolt) materjali mingil moel struktureerida (lõikudeks, lauseteks, paladeks, värssideks, kõnevoorudeks jm) ja annoteerida (märgendada lingvistilisel tasandil ehk sõnavormidena, lauseliikmetena, sõnatähendustena jne).

Korpuse koostamise käigus on osaliselt võimalik isikuandmeid eemaldada (anonüümida) või asendada teiste samasse semantilisse kategooriasse kuuluvate sõnade või sisuüksustega (pseudonüümida), näiteks Mari asemel Kadri⁸, Järvakandi

⁸ Nime eemaldamine ei ole tihti piisav meede isikuandmete kaitseks (vt Zimmer 2010).

keskkooli asemel mõne teise samaliigilise asula keskkool jne. Leidub aga ka isikuandmeid, mida ei saa kuidagi eemaldada ega pseudonüümida, näiteks kõneleja hääl.

Tundliku keeleandmestiku kasutuseesmärk on keele teaduslik uurimine, muu keeleandmestiku, leksikonide ja keelemudelite koostamine ja masinõppesüsteemide treenimine. Sealjuures võivad eriti keeleuurimiseks olla vajalikud nii isikuandmed (sellest täpsemalt allpool) kui ka olulised osad terviktekstidest (mis võib olla problemaatiline autoriõiguse või konfidentsiaalsuse seisukohalt).

2.2. Isikuandmete kaitsega mitteseonduvad tegevused

Keelemudelid ja leksikonid esitavad fakte keele kohta ja koosnevad loenditest, seaduspärasustest, kirjeldustest, üldistustest, tõenäosusjaotustest jms. Neil puudub seos ühegi konkreetse isiku konkreetse ütluse või kirjutisega, nad ei sisalda isikuandmeid ega olulisi osi teostest või dokumentidest. Näiteks sõna hääldus, käänamisviis või esinemissagedus on empiirilised faktid keele kohta, mis on küll saadud tundlike keeleandmestike uurimise teel, aga ei sisalda neid andmestikke.

Kuna muu keeleandmestik isikuandmeid ei sisalda, siis ei ole nende töötlemine isikuandmete töötlemine, mistõttu me seda siin ei käsitle. Samamoodi ei vaja nad käsitlemist ka autoriõiguse, konfidentsiaalsuse ega kommertskasutuse piirangute mõttes, kuna nad on nende koostaja algupäraseid teosed, mis ei sisalda teiste autorite teoseid, konfidentsiaalseid dokumente ega kommertskasutuse piiranguga tundlikke andmestikke.

Keelemudel või leksikon annab infot selle korpuse kohta, mille uurimise põhjal see koostatud on. Korpused on tänapäeval enamasti nii suured (vähemalt sadu miljoneid sõnu), et iga konkreetse kõneleja, autori või teose eripärad lahustuvad selles tundmatuseni.

2.3. Isikuandmete kaitsega seonduvad tegevused mittetundlike keeleressursside puhul

Samas on põhimõtteliselt võimalik koostada keelemudeleid ja leksikone ka palju väiksema korpuse, näiteks ühe autori või ühe teose kohta. Sel juhul võib ka tulemus kajastada just selle autori individuaalset keelekasutust. Ka hääldusmudel võib olla koostatud ühe kõneleja häälduse põhjal ja esitada äratuntavalt selle isiku kõnemaneeeri.

Saadaoleva arvutusvõimsuse kasvades tekib võimalus, ja tulemuse kvaliteeditaotlusest lähtuvalt ka soov, koostada keelemudeleid ja leksikone järjest pikematest tekstikatketest, näiteks lisada sõnade sagedusloendile ka sõnapaaride, sõnakolmikute jne sagedusloendid. Mida pikemad need fragmendid on, seda võimalikumaks muutub konkreetsete tekstide või isikute kohta käiva info äratundmine keeleandmestikus, eriti kui sageduse miinimumlävi on madal või puudub.

Kui uuritav tundlik andmestik on olnud väike ja ebapiisavalt anonüümitud, siis võib tulemuses sisalduda nii isikuandmete kui ka teoste ja dokumentide äratuntavaid fragmente. Sellise fragmendi maksimaalne lubatav pikkus võib osutada vaidluste teemaks (vt täpsemalt Tavast jt 2013).

Paljud tänapäevased keelemudelid on tõenäosuslikud ja üks nende rakendusi on hinnata lausete võimalikkust keeles. Teoreetiliselt võiks olla kujuteldav, et kuigi keelemudel ise ei sisalda konkreetse isiku andmeid, peab see mõnda selle isiku kohta käivat lauset tavalisest palju tõenäolisemaks, andes sel teel isiku kohta kaudselt infot.

3. Isikuandmete töötlemise nõusolek

3.1. Nõusoleku vorminõuded

Andmetöötleja jaoks on üheks esimeseks küsimuseks, kas andmesubjekti poolt antud nõusolek isikuandmete töötlemiseks peab olema mingis kindlas vormis või piisab ka näiteks e-kirjast, salvestatud suulisest loast või kirjalikult fikseeritud materjalide edastamisest töötlejale. Üldmäärus nõusolekule vorminõudeid ei kehtesta, kuid paneb andmetöötlejale kohustuse tõendada, et andmesubjekt on nõustunud oma isikuandmete töötlemisega (ÜM art 7 p 1). Üldmääruse selgituste kohaselt võib nõusoleku anda ka elektroonilisel teel või suulise avaldusena. See hõlmab näiteks vajaliku lahtri märgistamist veebisaidil, infoühiskonna teenuste tehniliste seadmete valimist või muud avaldust või käitumist, millest on näha andmesubjekti nõusolek teda puudutavate isikuandmete kavandatavaks töötlemiseks. Seega on oluline tagada võimalus nõusoleku taasesitamiseks juhul, kui nõusoleku andmist tuleb tõendada.

Kui andmesubjekt annab nõusoleku kirjaliku kinnitusena, mis puudutab ka muid küsimusi, tuleb nõusoleku taotlus esitada viisil, mis on muudest küsimustest ja taotlustest selgelt eristatav, arusaadavalt ja lihtsasti kättesaadav ning selges ja lihtsas keeles sõnastatud (ÜM art 7 p 2). Uuritud nõusolekuvormidest ainult ühes sisaldub pikem teadusuuringu kirjeldus, milles on selgelt eristatud küsimus nõusoleku kohta kasutada andmeid teadusotstarbel ja teiste sarnaste andmekogudega seostamiseks (nr 9). Ankeetides, milles sisaldub nii küsimustik kui ka nõusolek kogutud andmete töötlemise kohta, on küsimus andmete töötlemise nõusoleku kohta samuti muudest küsimustest (nagu näiteks kinnitus andmete õigsuse kohta või saladuse hoidmise kohta) selgelt eristatud (nr 2, 3, 7).

Järgnevalt analüüsime nõusolekuvormide sisulist vastavust üldmääruse ja uue IKS eelnõu (2017) nõuetele.

3.2. Nõusoleku sisulised nõuded

3.2.1. Vabatahtlikkus

Üldmääruse kohaselt tähendab andmesubjekti nõusolek vabatahtlikku, konkreetset, teadlikku ja ühemõttelist tahteavaldust, millega andmesubjekt kas avalduse vormis või selge nõusolekut väljendava tegevusega nõustub tema kohta käivate isikuandmete töötlemisega (ÜM art 4 p 11). Nõusoleku vabatahtlikkus tähendab, et andmesubjektil on tõeline ja vaba valikuvõimalus ning ta saab kahjulike tagajärgedeta nõusoleku andmisest keelduda või seda tagasi võtta (ÜM põhjenduspunkt 42). Nõusoleku tagasivõtmine peab olema lihtne ning teostatav samal viisil kui toimus nõusoleku andmine. Lisaks peab andmesubjekti teavitama nõusoleku tagasivõtmise õigusest (ÜM art 7 p 3).

Nõusoleku andmise vabatahtlikkust tuleb hinnata konkreetseid asjaolusid arvesse võttes (Pormeister 2017: 19). Näiteks ei saa eeldada vabatahtlikkust, kui andmesubjekt ja vastutav töötleja on selgelt ebavõrdses olukorras, eriti juhul kui vastutav töötleja on avaliku sektori asutus (ÜM põhjenduspunkt 43). Üldmääruse kohaselt ei pea nõusolekuvorm sisaldama viiteid poolte omavahelisele seosele. Vaidluse korral peab nõusoleku andmise vabatahtlikkust tõendama andmetöötleja (ÜM art 4 p 11).

Nõusoleku vabatahtlikkuse tagab mh võimalus anda erinevatele isikuandmete töötlemise toimingutele, nagu näiteks kogumine, korrastamine, säilitamine, muutmine, kasutamine, kustutamine (vt ÜM art 4 p 2 isikuandmete töötlemise mõiste), eraldi nõusolek. Samas võib eraldi nõusoleku andmise võimaluse puudumine olla asjakohane (ÜM põhjenduspunkt 43). Teadusuuringu eesmärgil andmete töötlemise nõusoleku vabatahtlikkuse hindamisel tuleks lähtuda põhjenduspunktis 33 antud selgitustest, mille kohaselt ei ole sageli isikuandmete kogumise ajal võimalik täielikult kindlaks määrata teadusuuringute eesmärgil toimuva andmetöötlemise eesmärki ja seega tuleks eraldi igale töötlemise toimingule nõusoleku andmise võimaluse puudumine lugeda põhjendatuks (Chassang 2017: 10). Uuritud nõusolekuvormides on kirjeldatud erinevaid toiminguid koos töötlemise eesmärkidega, mida tuleks lugeda piisavaks, et oleks tagatud nõusoleku vabatahtlikkus. Näiteks kasutamine projektis ettenähtud eesmärkidel ja eesti keele uurimiseks (nr 6), andmebaaside loomiseks ja kasutamiseks teaduslikel eesmärkidel (nr 3), kasutamine teadusotstarbel ja teiste andmekogudega seostamisel (nr 9), keele uurimine ja avalik kasutamine keeletehnoloogiliste rakenduste arendamisel (nr 7). ÜM nõuetele vastavuse tagamiseks tuleks aga töötlemise eesmärk sõnastada selliselt, et oleks võimalik hinnata, millise teadusvaldkonna või projekti raames tehtavaid toiminguid nõusolek hõlmab. Üldsõnaline teadusuuringutele viitamine ei ole ilmselt piisav (vt p 3.3.2).

Teave nõusoleku tagasivõtmise õiguse kohta on esitatud ainult kahes nõusolekuvormis (nr 9 ja 7). Samuti puudus nõusolekuvormidel teave andmesubjekti õiguse kohta nõuda kogutud andmete kustutamist või parandamist. Ka ei sisalda uuritud vormid teavet andmesubjekti muude õiguste kohta seoses kogutud andmete või antud nõusolekuga. Teave andmetele juurdepääsu kohta on siiski reeglina nõusolekuvormilt või ankeedilt tuvastatav läbi viite andmete töötlejale.⁹

Üldmäärus kehtestab ka üldpõhimõtte, mille kohaselt on andmesubjektil õigus "olla unustatud" (ÜM art 17). Kui andmesubjekt kasutab oma õigust nõusolek tagasi võtta, siis kaasneb sellega üldjuhul ka õigus nõuda andmete kustutamist. ÜM art 17 p 3d kohaselt on andmetöötlejal õigus sellest keelduda, kui andmete töötlemine toimus avalikes huvides toimuva arhiveerimise, teadus- või ajaloouringute või statistilisel eesmärgil sel määral, mil kustutamise nõudmise õigus tõenäoliselt muudab sellise töötlemise eesmärgi saavutamise võimatuks või häirib seda suurel määral. Seega võib andmetöötleja nõusolekuvormis välistada andmesubjekti õiguse nõuda andmete kustutamist nõusoleku tagasivõtmise korral (vt ÜM art 17 p 1b, p 3d).

3.2.2. Konkreetsus

Üldmääruse kohaselt peab nõusolek olema konkreetne (ÜM art 4 p 11). Seejuures eeldatakse, et isikuandmete kogumise ajal ei ole võimalik täies mahus kindlaks määrata teadusuuringute eesmärgil toimuva andmetöötlemise eesmärke. Nii piisab

⁹ Ainult ühest vormist ei ole võimalik välja lugeda, kus kogutud andmeid töödeldakse või hoitakse.

nõusolekust töödelda andmeid mingis uuringuvaldkonnas, võttes arvesse valdkonna tunnustatud eetikanorme (ÜM põhjenduspunkt 33). Nõusolekuvormis ei pea loetlema andmeid, mida teadusuuringu raames kogutakse, küll aga peavad selles sisalduma selged, õiguspärased ja konkreetsed andmete töötlemise eesmärgid andmete kogumise ajal. Kui andmesubjekt ei tea nõusolekut andes, millisel eesmärgil temalt andmeid kogutakse, siis ei saa ta anda konkreetset ja teadlikku nõusolekut andmete töötlemise kohta (Macenaite, Kosta 2017: 158). Seega ei piisa andmesubjekti informeerimisest, milliseid andmeid kogutakse (nt soo, vanuse jm demograafiliste andmete, subjekti enda ja tema vanemate keeleoskuse ja -kogemuse kohta (nr 9); nime, sünniaja, soo, sünnikoha, rahvuse, elamiskohtade ja neis elatud aastate kohta, emakeele, keeleoskuse, vanemate keele, kõne- või kuulmishäirete, käelisuse, hariduse, muusikalise hariduse, töö- või ametikoha, gümnaasiumi ja selle õppekeele, eesti keele oskuse hinnangu ning kontaktandmete kohta (nr 3 ja 7); ainult nime- ja allkirjanõue (nr 4, 5 ja 6)), vaid märkida tuleb ka töötlemise eesmärgid. ÜM kohaselt peavad kogutud isikuandmed olema asjakohased, olulised ja piiratud sellega, mis on vajalik nende töötlemise eesmärgi seisukohalt (ÜM art 5 p 1c). Uuritud nõusolekuvormides esitatud teave kogutavate andmete kohta oli küll kooskõlas “võimalikult väheste andmete kogumise” põhimõttega, kuid mitte alati ei ole selgelt eristatud eesmärgid, milleks andmeid töödeldakse.

Problemaatiline on märkida eesmärgina nii õppe- kui ka teadustööd ilma neid eristamata või teadusuuringu valdkonda nimetamata. Näiteks sõnastused nagu “kasutamine õppe- ja teadustöös ning erinevate keeletehnoloogiliste rakenduste loomisel” (nr 4); “kasutamine õppe- ja teadustöös ja taasesitamine” (nr 2); “lisamine aktsendikorpuse koosseisu, kasutamine avalikult, teadustöös ja ärirakenduste loomisel” (nr 5) ei pruugi olla kooskõlas üldmääruse nõuetega. Õiguskirjanduses on leitud, et kui nõusolek on antud isikuandmete töötlemiseks teadustöös, siis ei saa seda laiendada õppetööle (Magnusson Sjöberg 2017). Seda põhjusel, et teadustöö raames on teadlaste tegevus isikuandmete kaitse aspektist küllaltki põhjalikult reguleeritud ning kontrollitav, õppetöös osalevate õppijate puhul see aga nii ei ole. Samuti peaksid teadusvälised eesmärgid olema esitatud viisil, mis võimaldaks andmesubjektil valida, kas anda oma nõusolek või mitte (vt nt Lewis jt 2017: 63).

3.2.3. Teadlik tahteavaldus

Üldmääruse põhjenduspunkti 42 kohaselt tähendab teadlik nõusolek, et andmesubjekt peab olema teadlik vähemalt sellest, kes on vastutav töötleja ja milleks kavatsetakse isikuandmeid töödelda. Üldnõuded andmesubjektile edastatava teabe kohta sätestab ÜM art 13. Uue IKS eelnõu (2017) § 25 kohaselt tuleb lisaks isikuandmete töötlemise kavandatud eesmärgile (vt p 3.3.2) teavitada ka õigusest tutvuda oma andmetega, nende parandamise, kustutamise või piiramise õigusest ja õiguste teostamise korrast ning õigusest esitada kaebus Andmekaitse Inspeksioonile. Lisaks tuleb teatada vastutava töötleja ning andmekaitseametniku nimi ja kontaktandmed ning Andmekaitse Inspeksiooni kontaktandmed. Teavitamiskohustuse täitmiseks piisab ka vastava teabe avalikustamisest vastutava töötleja veebilehel või muus andmesubjektile kergesti juurdepääsetavas asukohas.

Andmetöötleja kontaktandmete esitamine tagab andmesubjektile tegeliku võimaluse oma õigusi teostada. Kõik uuritud nõusolekuvormid sisaldavad teavet

andmetöötleva kohta, kuigi enamikus neist puuduvad andmekoguja kontaktandmed (v.a nr 7 ja 9). Juhul kui andmetöötleva on määranud andmekaitseametniku, peab nõusolekuvormil ära märkima ka andmekaitseametniku nime ning kontaktandmed. Mitte kõigist uuritud nõusolekuvormidest ei olnud võimalik teada saada, kas kogutud materjale (tekstikorpuse salvestised) töödeldakse isikustatud viisil või anonüümitult.

3.2.4. Nõusoleku selgus, kättesaadavus ja õiglus

Üldmääruse põhjenduspunkti 42 kohaselt peab nõusolek olema kooskõlas direktiiviga ebaõiglaste tingimuste kohta tarbijalepingutes, st olema arusaadav ja lihtsasti kättesaadav, selles tuleks kasutada selget ja lihtsat keelt ning selles ei tohiks olla ebaõiglasi tingimusi. Eesti õiguses on üldmääruses viidatud direktiivi ebaõiglaste tingimuste kohta tarbijalepingutes nõuded sätestatud VÕS § 37 (Varul jt 2016).

Artikli koostamisel uuritud nõusolekud olid väljaspool seda lepingulist suhet, mis aga üldmääruse kohaselt ei välista tüüpitingimuste regulatsiooni kohaldamist. Nii peavad nõusolekus sisalduvad tingimused VÕS § 37 lg 1 kohaselt olema sellised, et neist on võimalik aru saada. Kui tingimuste sisu, väljendusviis või esituslaad on niivõrd ebataoline või arusaamatu, et teine pool ei võinud tingimust mõistlikkuse põhimõttest lähtudes oodata või seda tingimust olulise pingutuseta mõista, ei ole need tingimused andmesubjektile siduvad (VÕS § 37 lg 3, ÜM art 7 p 2).

Uuritud nõusolekuvormide keelekasutus on üldjuhul suhteliselt lihtne ja selge. Siiski on ka selliseid nõusolekuvorme, kus nõusolek andmete töötlemiseks ei ole piisavalt selgelt eristatud küsimustest, mis on esitatud isikuandmete kogumise eesmärgil (nt nr 3), või kus toimingute kirjeldusest ei ole võimalik üheselt aru saada (nt "seostamine teiste andmekogudega", nr 9).

Põhimõtteliselt peaks alluma õigluskontrollile nõusolekuvormides sisalduvad vastutust välistavad tingimused, eelkõige kui need on vastuolus üldmääruse või liikmesriigi õigusega. Nt on vaieldav, kas on kehtiv andmetöötleva vastutuse välistamine pärast salvestise eemaldamist korpusest kolmanda isiku valduses olevate salvestiste koopiate või nende osade kasutuse ja leviku eest (nr 9). Nõusolekuvormis sisalduva autoritasust loobumise võib lugeda õiglaseks tingimuseks, kui autoritasust loobumine konkreetsel juhtumil on seadusega lubatud¹⁰ (vt nr 6). Seega sõltub konkreetse nõusolekuvormi sisust ja andmete töötlemise eesmärgist, millises ulatuses on võimalik selle sisu õiglust kontrollida.

4. Isikuandmete töötlemine ilma nõusolekuta teadustöö eesmärkidel (teadustöö erand)

Üldmääruse kohaldamine algab 25.5.2018 ja enne seda on vaja siseriiklik õigus viia kooskõlla üldmäärusega, kuid lõplik lahendus ei ole selge. Vastavalt AKI seisukohale tuleks praeguse teadustöö regulatsiooni paindlikkus seoses loamenetlusega säilitada. Uuringuluba annab tuvastatavate isikuandmetega uuringu tegemiseks õigusliku aluse. Umbisikustatud (anonüümitud) andmetega uuringud ei käi loakohustuse alla, sest isikuandmeid ei töödelda (AKI seisukoht 2017: 1). Sama lähenemist toetab

¹⁰ Autoriõiguse seaduse kohaselt on autori varalised õigused üleantavad tasu eest või tasuta (§ 11 (3)).

ka isikuandmete kaitse uus kontseptsioon, esitades järgmise seisukoha: “säilitada saab teadusuuringute loamenetluse (IKS § 16) kui kolmanda alternatiivi seadusekohase ja nõusolekupõhise teadusuuringu kõrval. Topeltprotseduuride vältimiseks saab võrdsustada loamenetluse läbimise üldmääruse kohase eelkonsulteerimisega (ÜM art 36).” (Isikuandmete kaitse uus kontseptsiooni 2017: 25).

Euroopa andmekaitseasutuste töörühm on seisukohal, et isikuandmete töötlemise aluseks teadustöö eesmärgil on õigustatud huvi (ingl *legitimate interest*) (WP 29 2014: 24–25). Isikuandmete töötlemisele õigustatud huvi alusel peab see kaaluma üles andmesubjekti huvid või põhiõigused ja -vabadused (ÜM art 6 p 1f).

Konkreetsed nõuded isikuandmete töötlemisele teadustöö kontekstis seab ÜM art 89.¹¹ Sellest tuleneb kohustus kohaldada isikuandmete töötlemisel asjakohaseid kaitsemeetmeid andmesubjekti õiguste ja vabaduste kaitseks. Need meetmed võivad hõlmata pseudonüümimist või anonüümimist, mida saab täita täiendava töötlemisega, mis võimaldavad või ei võimalda enam andmesubjektide tuvastamist. ÜM annab liikmesriikidele õiguse näha ette oma täpsustav regulatsioon seoses teadustegevusega. Nii võivad liikmesriigid teha erandeid järgmistes andmesubjekti õigustes:

- 1) andmesubjekti õigus tutvuda andmetega (art 15);
- 2) õigus andmete parandamisele (art 16);
- 3) õigus isikuandmete töötlemise piiramisele (art 18);
- 4) õigus esitada vastuväiteid (art 21).

IKS uues eelnõus (2017) reguleerib isikuandmete töötlemist teadusuuringus § 11. Üldpõhimõtteks on, et isiku nõusolekuta võib teadusuuringu vajadusteks töödelda isiku kohta käivaid andmeid eelkõige pseudonüümitud¹² või samaväärset andmekaitse taset võimaldaval kujul. Enne isikuandmete üleandmist teadusuuringu või riikliku statistika vajadustel töötlemiseks asendatakse isiku tuvastamist võimaldavad andmed isikut mittetuvastavate andmetega (IKS uus eelnõu § 11 (1)).

Andmesubjekti tuvastamist võimaldaval kujul on andmeid teadustöö eesmärkidel nõusolekuta töödelda lubatud üksnes juhul, kui on täidetud järgmised tingimused (IKS uus eelnõu § 11 (3)):

- 1) pärast tuvastamist võimaldavate andmete eemaldamist ei oleks andmetöötlemise eesmärgid enam saavutatavad või oleks neid ebamõistlikult raske saavutada (vältimatu vajadus);
- 2) kui selleks on teadusuuringu teostaja hinnangul ülekaalukas avalik huvi;
- 3) töödeldavate isikuandmete põhjal ei muudeta andmesubjekti kohustuste mahtu ega kahjustata muul viisil ülemäära andmesubjekti õigusi;
- 4) kui isikuandmete töötleja on võtnud kasutusele kaitsemeetmed.

Isikuandmete töötlemine keeleandmetike kontekstis on vältimatult vajalik (IKS uus eelnõu § 11 lg 3 p 1; kehtiv IKS § 16 lg 2). Keeleteaduses võib mingit isikuandmete alamhulka tarvis olla näiteks järgmistel juhtudel:

- a) kõneldud (helisalvestatud) keele uurimisel on keelejuhi kui andmesubjekti häääl vältimatult kaasas;
- b) longituuduuringus on vaja omavahel seostada sama isiku keelekasutust eri aegadel, seega on uurijal vaja kindlasti ka kontaktandmeid, et kokku leppida intervjuud regulaarsete ajaintervallide järel;

¹¹ ÜM lubab isikuandmeid töödelda ka akadeemilise eneseväljenduse tarbeks (art 85). Eriliiki isikuandmeid (delikaatseid isikuandmeid) lubab töödelda ÜM art 9 p 2j, mis sätestab, et isikuandmete eriliikide töötlemise keeld ei kehti, kui töötlemine on vajalik avalikes huvides toimuva arhiveerimise, teadus- või ajaloouringute või statistilisel eesmärgil ning on proportsionaalne saavutatava eesmärgiga, austab isikuandmete kaitse õiguse olemust ning tagatud on sobivad ja konkreetsed meetmed andmesubjekti põhiõiguste ja huvide kaitseks.

¹² Üldmääruse kohaselt tuleb pseudonüümitud andmeid, mida saaks füüsilise isikuga seostada täiendava teabe abil, käsitada isikuandmetena (põhjenduspunkt 26).

- c) sotsiolingvistilistes uuringutes on vaja teada kõneleja demograafilisi andmeid, nt murdetausta, lisaks ka vanemate emakeelt, elukohta (murdeala) lapseas, elamist välisriikides, milliseid keeli kasutatakse koduses ja tööalases suhtluses jne;
- d) kombineeritud uuringumetoodika korral võib olla tarvis keelejuhi kontaktandmeid, et temaga intervjuu läbi viia.

Kui teadusuuringu teostaja hinnangul esineb ülekaalukas avalik huvi, on põhjendatud töödelda isikuandmeid andmesubjekti nõusolekuta (IKS uus eelnõu (2017) § 11 lg 3 punkt 2; kehtiv IKS § 16 lg 2). Kehtiva IKS eelnõu (2006) seletuskirja kohaselt on "avalik huvi" määratlemata õiguspõhiste, millele tuleb sisu anda iga üksikjuhtumit eraldi silmas pidades. Andmesubjekti nõusolekuta tema kohta käivate andmete teaduslikul eesmärgil töötlemiseks pole vajalik pelgalt avaliku huvi kui sellise olemasolu, vaid avalik huvi peab lisaks kaaluma üles vastanduvad huvid, eelkõige andmesubjekti huvi tema eraelu puutumatuses suhtes (IKS eelnõu 2006; vt ka AKI teadustöö juhend 2015). Ülekaalukaks avalikuks huviks võib selles kontekstis pidada eesti keele kaitsmise vajadust (vt Jents, Kelli 2014, Kelli jt 2012). Ilma digitaliseerimata võib eesti keel hääbuda (Kelli jt 2016b). Eesti keele kui riigikeele staatust rõhutab ka Eesti Vabariigi põhiseadus (§ 6).

Isikuandmete töötlemine teadusuuringus ei tohi kahjustada andmesubjekti (IKS uus eelnõu § 11 lg 3 p 3; kehtiv IKS § 16 lg 2). Andmesubjekti kohustuste mahtu ja õigusi võidakse muuta rakendustes, kus isikuandmete põhjal tehakse järeldusi isiku kohta (tööl sobivus, laenukõlblikkus vms). Keeleteadlased ja keeletarkvara arendajad tegelevad keelega, mitte isikutega, ja keeleuurimisel puudub enamasti igasugune seos andmesubjekti kohustuste ja õigustega. Praegu on teada ainult üks erand: isiku keelekasutuse järgi on võimalik ennustada tema isikuomadusi ja tulevast käitumist (sh nt suitsidaalsust või soodumust ühiskonnaohtlikuks käitumiseks). Tegelikult keeleteadlasi konkreetse keelejuhi isikuomadused ei huvita ja andmete levitamise täiendav piiramine võimalikke kahjustavaid järeldusi ära ei hoiaks, sest soovi korral on iga aktiivse inimese keelekasutuse näidiseid niikuinii võimalik leida avalikest allikatest.

Isikuandmete kaitseks on isikuandmete töötlemise kohustatud kasutusele võtma piisavad organisatsioonilised, füüsilised ja infotehnilised turvameetmed (IKS § 16 lg 3). Uus IKS eelnõu § 8 loetleb meetmetena muuhulgas:

- a) tehnilisi ja korralduslikke kaitsemeetmeid;
- b) meetmeid, mis võimaldavad hiljem kontrollida ja kindlaks teha, kas isikuandmeid on lisatud, muudetud, kustutatud või muul viisil töödeldud ja kes seda on teinud;
- c) andmekaitseametniku määramist;
- d) isikuandmete pseudonüümimist;
- e) isikuandmete krüptimist;
- f) konkreetseid juhiseid, mis tagavad isikuandmete edastamise või muul eesmärgil töötlemise.

Isikuandmete töötlemine teadusuuringus ilma andmesubjekti nõusolekuta tuleb IKS uue eelnõu (2017) kohaselt kooskõlastada Andmekaitse Inspektsiooniga (§ 11 lg 4).

Kokkuvõtlikult võib öelda, et isikuandmete töötlemine teadustöö eesmärgil on paindlikum võrreldes isikuandmete töötlemisega muudel eesmärkidel. Näiteks näeb

Üldmäärus (art 5) ette põhimõtte, mille kohaselt isikuandmeid kogutakse täpselt ja selgelt kindlaksmääratud ning õiguspärasel eesmärkil ning neid ei töödelda hiljem viisil, mis on nende eesmärkidega vastuolus (eesmärgi piirang). Üldmäärus sätestab ka säilitamise piirangu, mille kohaselt isikuandmeid säilitatakse kujul, mis võimaldab andmesubjekte tuvastada ainult seni, kuni see on vajalik selle eesmärgi täitmiseks (art 5). Teadustöö kontekstis ei ole eesmärgi ja säilitamise piirang nii range. Isikuandmete edasist töötlemist avalikes huvides toimuvate teadusuuringute eesmärgil ei loeta üldmääruse kohaselt algsete eesmärkidega vastuolus olevaks. Selles kontekstis võib ka isikuandmeid kauem säilitada eeldusel, et rakendatakse asjakohaseid tehnilisi ja korralduslikke meetmeid (art 5 p 1b ja 1e).

Keeleuuringu kontekstis on tekkinud ka küsimus isikuandmete kaitse ajalisest kehtivusest. Üldmäärus jätab selle küsimuse liikmesriikide otsustada: “Käesolevat määrust ei kohaldata surnuid puudutavate isikuandmete suhtes. Liikmesriikidel peaks olema võimalik ette näha surnuid puudutavate isikuandmete töötlemise eeskirjad” (ÜM p 27). Kehtiva IKS kohaselt on isikuandmed kaitstavad 30 aastat pärast andmesubjekti surma (IKS § 13 lg 1). Uus IKS eelnõu (2017) lähtub samast põhimõttest (§ 14 lg 1).

5. Järeldused ja ettepanekud

Isikuandmete kaitse regulatsioon omab tähendust üksnes keeleressursside puhul, mis sisaldavad isikuandmeid. Isikuandmete kaitse nõudeid ei pea järgima ka siis, kui isikuandmed on keeleandmestikust lõplikult eemaldatud. Osas keeleressurssides ei ole isikuandmete täielik ja lõplik eemaldamine võimalik (nt keeleressursid, mis sisaldavad suulist kõnet). Soovitav on ka sel juhul isikuandmeid keeleressurssidest eemaldada nii palju kui võimalik. Isegi kui kõiki isikuandmeid ei saa eemaldada, on see üheks meetmeks isikuandmete kaitsel. Meetmeks isikuandmete kaitsel on ka võimalikult väheste andmete kogumine.

Üldmäärus lähtub teadusuuringu laiast tõlgendusest, mis hõlmab ka erasektori rahastatavaid uuringuid (põhjenduste punkt 159). Keeleteaduslikuks uurimistööks võib kasutada ka muud isikuandmeid sisaldavat materjali, mis algselt ei ole loodud teadustööks. Üldmääruse kohaselt on isikuandmete teaduslikul eesmärgil töötlemine (isegi, kui neid andmeid selleks ei kogutud) kooskõlas isikuandmete kaitse põhimõtetega (ÜM art 5 p 1b). Teadustöö kontekstis on andmete säilitamise piirang samuti paindlik (ÜM art 5 p 1e).

Kui isikuandmed ei ole keeleressurssidest lahutatavad, siis saab neid teaduslikul eesmärgil töödelda peamiselt kahel alusel: 1) andmesubjekti nõusoleku ja 2) teadustöö erandi alusel.

Tuginedes olemasolevate nõusolekuvormide ja regulatsiooni analüüsile jõuavad autorid järgmiste järeldusteni.

- 1) Nõusolekuvormid, mida kasutatakse keeleressursside kogumisel, ei pea sisaldama loetelu toimingutest, milleks nõusolekut soovitakse ning lisaks ka võimalust anda nõusolek igäihe suhtes eraldi. Kuivõrd tegemist on teaduseesmärgil toimuva isikuandmete töötlemisega, siis piisab, kui piiritleda teadusvaldkond, mille raames isikuandmete töötlemine toimub. Andmesubjektil peaks olema eraldi võimalus nõustuda oma andmete kasutamisega muul kui teaduslikul eesmärgil (nt ärirakenduste loomiseks).

- 2) Nõusolekuvormides tuleb esitada teave viisil, mis on selge ja andmesubjektile arusaadav. Kasutuses olevad nõusolekuvormid ei sisalda kõiki kehtiva IKS-i ja uue üldmäärusega ettenähtud teavet, et oleks tagatud teadlik nõusolek. Kasuks tuleb teksti parem struktureerimine ning süsteemsem ülesehitus. Mitte iga vorm ei sisalda võimalust valida, kas anda nõusolek andmete töötlemiseks ka muul kui teaduslikul eesmärgil, ning teavet andmesubjekti õiguste kohta.
- 3) Soovitatav on välja töötada tüüp-nõusolekuvormid, mida saaks eri teadusvaldkonna vajadustele vastavalt kohandada. Keeleressursside nõusolekuvormid peaksid olema CLARIN-i konsortsiumis ühtlustatud, mis aitaks kaasa EL-i sisesele teadustööle.
- 4) Kui keeleressursse tahetakse kasutada õppetöös, siis peaks see eesmärk kajastuma ka nõusolekuvormides.

Keeleressurssides, mille loomisel ei ole andmesubjektilt nõusolekut küsitud, on oluline töödelda isikuandmeid õigustatud huvist tuleneva teadustöö erandi alusel. Üldistatult öeldes jätab regulatsioon teadustöö erandi alusel toimuva uurimistöö samaks. Ilma isikuandmete töötlemiseta ei pruugi nt suulist kõnet sisaldavate keeleressursside loomine olla võimalik. Kuna keeleressurssidel on oluline roll eesti keele digitaliseerimisel ja säilitamisel, siis saab jaatada kaaluka õigustatud huvi olemasolu isikuandmete töötlemisel. Keeleteaduslik uurimistöö ei mõjuta reeglina andmesubjekti õigusi või kohustusi, kuid keeleressursside loomisel ja kasutamisel tuleb isikuandmete kaitseks rakendada organisatsioonilisi, füüsilisi ja infotehnilisi turvameetmeid.

Viidatud kirjandus

- Chassang, Gauthier 2017. The impact of the EU general data protection regulation on scientific research. – *Ecancer*, 11, 709. <https://doi.org/10.3332/ecancer.2017.709>
- Ginter, Carri 2006. Constitutional review and EC law in Estonia. – *European Law Review*, 31 (6), 912–923.
- Jents, Liina; Kelli, Aleksei 2014. Legal aspects of processing personal data in development and use of digital language resources: the Estonian Perspective. – *Jurisprudence*, 21 (1), 164–184. <https://doi.org/10.13165/JUR-14-21-1-08>
- Kelli, Aleksei; Vider, Kadri; Pisuke, Heiki; Lindén, Krister 2016a. Digitaalsete keeleressursside loomist ja kasutamist määrav õiguslik raamistik Eestis ja selle ühildumine CLARIN-i infrastruktuuriga [‘Regulatory framework determining the development and utilization of digital language resources in Estonia and its compatibility with CLARIN infrastructure’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 12, 81–98. <https://doi.org/10.5128/ERYa12.05>
- Kelli, Aleksei; Vider, Kadri; Pisuke, Heiki; Siil Triin 2016b. Constitutional Values as a basis for the limitation of copyright within the context of digitalization of the Estonian language. – *Collection of Research Papers in Conjunction with the 6th International Scientific Conference of the Faculty of Law of the University of Latvia “Constitutional Values in Contemporary Legal Space”*, Riga, 16–17 November, 2016, 126–139. https://www.lu.lv/fileadmin/user_upload/lu_portal/apgads/PDF/Book_Juristu_6_konf_II_dalja.pdf#page=126 (23.10.2017).
- Kelli, Aleksei; Vider, Kadri; Lindén, Krister 2015. The regulatory and contractual framework as an integral part of the CLARIN infrastructure. – 123: *Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wrocław, Poland*. Koenraad De Smedt (Ed.), Linköping Electronic Conference Proceedings. Linköping: Linköping University

- Electronic Press, 13–24. <http://www.ep.liu.se/ecp/article.asp?issue=123&article=002> (3.10.2017).
- Kelli, Aleksei; Tavast, Arvi; Pisuke, Heiki 2012. Copyright and constitutional aspects of digital language resources: The Estonian approach. – *Juridica International*, 19, 40–48. http://www.juridicainternational.eu/public/pdf/ji_2012_1_40.pdf (3.10.2017).
- Lewis, David; Moorkens, Joss; Fatema, Kaniz 2017. Integrating the management of personal data protection and open science with research ethics. – *Proceedings of the First Workshop on Ethics in Natural Language Processing*, 60–65. <http://www.ethicsinnlp.org/workshop/pdf/EthNLP07.pdf> (15.11.2017).
- Macenaite, Milda; Kosta, Eleni 2017. Consent for processing children’s personal data in the EU: following in US footsteps? – *Information & Communications Technology Law*, 26 (2), 146–197. <https://doi.org/10.1080/13600834.2017.1321096>
- Magnusson Sjöberg, Cecilia 2017. Scientific research and academic e-learning in light of the EU’s legal framework for data protection. – Marcelo Corralse, Mark Fenwick, Nikolaus Forgó (Eds.), *New Technology, Big Data and the Law*. Singapore: Springer Nature, 43–63. <https://doi.org/10.1007/978-981-10-5038-1>
- Oksanen, Ville; Lindén, Krister; Westerlund, Hanna 2010. Laundry Symbols and License Management: Practical Considerations for the Distribution of LRs Based on Experiences from CLARIN. <http://hdl.handle.net/10138/29359>
- Pormeister, Kärt 2017a. Genetic data and the research exemption: is the GDPR going too far? – *International Data Privacy Law*, 7 (2), 137–146. <https://doi.org/10.1093/idpl/ix006>
- Pormeister, Kärt 2017. Informed consent to sensitive personal data processing for the performance of digital consumer contracts on the example of “23andMe”. – *Journal of European Consumer and Market Law*, 1, 17–23.
- Tavast, Arvi; Pisuke, Heiki; Kelli, Aleksei 2013. Õiguslikud väljakutsed ja võimalikud lahendused keeleressursside arendamisel [‘Legal challenges and possible solutions in developing language resources’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 9, 317–332. <https://doi.org/10.5128/ERYa9.20>
- Varul, Paul; Kull, Irene; Kõve, Villu; Käerdi, Martin; Sein, Karin 2016. *Võlaõigusseadus I. Üldosa* (§§ 1-207). Komm vlj. Tallinn: Juura.
- Vider, Kadri; Kelli, Aleksei; Pisuke, Heiki; Mõtsküla, Peeter P.; Tuulik, Triin 2013. Keeleressursside litsentsimise õiguslike küsimusi [‘Legal considerations in licensing language resources’]. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 9, 333–347. <https://doi.org/10.5128/ERYa9.21>
- Zimmer, Michael 2010. “But the data is already public”: on the ethics of research in Facebook. – *Ethics and Information Technology*, 12 (4), 313–325. <https://doi.org/10.1007/s10676-010-9227-5>

Võrguviited

- AKI seisukoht = Andmekaitse Inspektsiooni täiendavad seisukohad uue andmekaitseõiguse kontseptsiooni asjus, 22.06.2017. http://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/andmekaitse_inspektsiooni_taiendavad_seisukohad_22062017.pdf (13.11.2017).
- AKI teadustöö juhend = Andmekaitse Inspektsioon. Isikuandmete töötlemine teadusuuringus (juhend), 12.01.2015. http://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/Juhised/20150917%20-%20Isikuandmete%20t%C3%B6tlemine%20teadusuuringus.pdf (14.11.2017).
- Autoriõiguse seadus. RT I 1992, 49, 615 ... RT I, 16.06.2017, 1. <https://www.riigiteataja.ee/akt/116062017008> (14.11.2017).
- CLARIN = Common Language Resources and Technology Infrastructure. <https://www.clarin.eu/content/clarin-in-a-nutshell> (3.10.2017).

Eesti Vabariigi põhiseadus. RT 1992, 26, 349 ... RT I, 15.05.2015, 1. <https://www.riigiteataja.ee/akt/115052015002> (7.1.2018).

Euroopa Liidu põhiõiguste harta. OJ C 326, 26.10.2012, 391–407. <http://eur-lex.europa.eu/legal-content/ET/TXT/HTML/?uri=CELEX:12012P/TXT&from=ET> (3.10.2017).

Hea teadustava. Tartu, 2017. Eetikaveeb: http://www.eetika.ee/sites/default/files/www_ut/hea_teadustava_o.pdf (13.11.2017).

IKS = Isikuandmete kaitse seadus. RT I 2007, 68, 421 ... RT I, 06.01.2016, 10. <https://www.riigiteataja.ee/akt/106012016010?leiaKehtiv> (1.10.2017).

IKS eelnõu 2006 = Isikuandmete kaitse seaduse eelnõu nr 1026. Algtekst ja seletuskiri SE I, 08.11.2006. <https://www.riigikogu.ee/tegevus/eelnoud/eelnou/adf18590-e176-3c6f-91co-6ffa8f501cbo/Isikuandmete%20kaitse%20seadus%C2%B9%20> (23.10.2017).

IKS uus eelnõu 2017 = Isikuandmete kaitse seaduse eelnõu, 06.11.2017. <http://eelnoud.valitsus.ee/main#s9Cm1EME> (22.3.2018).

Isikuandmete kaitse direktiiv = Euroopa Parlamendi ja nõukogu direktiiv 95/46/EÜ, 24. oktoober 1995, üksikisikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise kohta. Euroopa Liidu Teataja L 281, 23/11/1995, 31–50. <http://data.europa.eu/eli/dir/1995/46/oj> (1.10.2017).

Isikuandmete kaitse uus kontseptsioon = Isikuandmete kaitse uue õigusliku raamistiku kontseptsioon, 18.04.2017. http://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/andmekaitse_kontseptsioon_11.04.2017.pdf (13.11.2017).

Isikuandmete kaitse üldmäärus (ÜM) = Euroopa Parlamendi ja nõukogu määrus (EL) 2016/679, 27. aprill 2016, füüsiliste isikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise ning direktiivi 95/46/EÜ kehtetuks tunnistamise kohta (isikuandmete kaitse üldmäärus) (EMPs kohaldatav tekst) ELT L 119, 4.5.2016, 1–88. <http://eur-lex.europa.eu/legal-content/ET/TXT/HTML/?uri=CELEX:32016R0679&rid=1> (3.10.2017).

Keeleressursside kogumise nõusolekuvormid. Eesti Keeleressursside Keskuse repositoorium. <https://entu.keeleressursid.ee/public-document/entity-7062> (23.3.2018).

Nõukogu direktiiv 93/13/EMÜ, 5. aprill 1993, ebaõiglaste tingimuste kohta tarbijalepingutes EÜT L 95, 21.4.1993, 29–34. <http://eur-lex.europa.eu/legal-content/ET/TXT/?qid=1515535550379&uri=CELEX:31993L0013> (10.1.2018).

TAKS = Teadus- ja arendustegevuse korralduse seadus. RT I 1998, 111, 1832 ... RT I, 04.12.2014, 4. <https://www.riigiteataja.ee/akt/104122014014?leiaKehtiv> (23.10.2017).

VÕS = Võlaõigusseadus. RT I 2001, 81, 487 ... RT I, 01.07.2017, 1. <https://www.riigiteataja.ee/akt/101072017020> (13.11.2017).

WP29 2017 = Article 29 Working Party (WP29). Guidelines on Consent under Regulation 2016/679. http://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/wp29_enpdf.pdf (28.11.2017).

WP29 2014 = Article 29 Working Party (WP29). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. http://collections.internetmemory.org/haeu/20171122154227/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (28.11.2017).

Aleksei Kelli on intellektuaalse omandi õiguse professor Tartu Ülikoolis. Ta tegeleb digitaalsete keeleressursside õiguslike küsimustega TÜ-s ning Eesti Keele Instituudis ning on CLARIN-i õigusasjade komitee esimees. Teadustöös on tema peamised huvid avatud teadus, teadmussiire ja keeletehnoloogia õiguslikud küsimused.

Kaarli pst 3, 10119 Tallinn, Estonia

aleksei.kelli@ut.ee

Kadri Vider on Eesti Keeleressursside Keskuse (EKRK) tegevjuht, CLARIN-i riiklik koordinaator Eestis ja Tartu Ülikooli arvutiteaduse instituudi keeletehnoloogia teadur. EKRR ülesanded on mh riiklike keeletehnoloogia teadus- ja arendusprogrammide tulemusena loodud ressursside ja tarkvara haldamine, kättesaadavaks tegemine ning litsentsimine või sellealane konsulteerimine. J. Liivi 2, 50409 Tartu, Estonia

kadri.vider@ut.ee

Irene Kull on tsiviilõiguse professor Tartu Ülikoolis. Ta tegeleb Eesti ja Euroopa Liidu lepinguõigusega. Teadustöö peamised huvid on EL-i lepinguõiguse ühtlustamine, tsiviilõiguse üldprintsiipide kohaldamine ning lepinguliste kohustuste tasakaalu õiguslikud probleemid.

Näituse 13a, 50409 Tartu, Estonia

irene.kull@ut.ee

Triin Siil on Cybernetica AS-i jurist. Nii oma varasemas töös advokaadina kui ka praeguses ametis tegeleb Triin peamiselt tehnoloogiate litsentsimise ja teadmussiirde õiguslike küsimustega.

Ta kirjutab Tartu Ülikooli õigusteaduskonnas doktoritööd, mis keskendub intellektuaalse omandi ja andmekaitseõiguse kokkupuutepunktilede.

triin.siil@gmail.com

Krister Lindén on teadusdirektor Helsingi Ülikoolis ning FIN-CLARIN-i juht, Soome CLARIN-i riiklik koordinaator ja CLARIN-i õiguskomitee liige. 2008–2011 juhatas ta õiguslike küsimustega tegelemist ESFRI projektis ja CLARIN ERIC ettevõtmisfaasis. Teadustöös on tema peamised huvid soome- ja mitmekeelsed ressursid.

krister.linden@helsinki.fi

Arvi Tavast on Qlaara Labsi asutaja. Tegelenud sõnastike ja keelemudelite koostamisega tundlike keeleandmetike põhjal nii era- kui ka akadeemilises sektoris (EKI, TÜ).

arvi@qlaara.com

Age Värv on võlaõiguse dotsent Tartu Ülikoolis. Ta tegeleb lepinguväliste võlasuhete ning võlaõiguse üldosa ning võrdleva õiguse küsimustega. Teadustöös keskendub ta deliktiõigusele ja alusetu rikastumise õigusele ning intellektuaalomandi õiguse ja võlaõiguse kokkupuutepunktilede.

age.varv@ut.ee

Carri Ginter on Tartu Ülikooli Euroopa õiguse dotsent, kes tegeleb Euroopa Liidu õiguse rakendamise küsimustega. Teadustöös on tema peamised huvid EL-i ja liikmesriikide õigustike omavahelised kokkupuutepunktid.

carri.ginter@ut.ee

Einar Meister on Tallinna Tehnikaülikooli tarkvarateaduse instituudi vanemteadur, omab pikaajalist kogemust kõnekorpusete loomisel ja kasutamisel eksperimentaalfoneetilistes uuringutes ning kõnetehnoloogia rakendustes.

einar.meister@ttu.ee

DATA PROTECTION ISSUES RELATING TO THE DEVELOPMENT AND UTILISATION OF LANGUAGE RESOURCES

**Aleksei Kelli¹, Kadri Vider^{1,2}, Irene Kull¹, Triin Siil^{1,3},
Kristen Lindén⁴, Arvi Tavast⁵, Age Värvi¹,
Carri Ginter¹, Einar Meister⁶**

University of Tartu¹, Center of Estonian Language Resources², Cybernetica AS³,
University of Helsinki⁴, Qlaara Labs⁵, Tallinn University of Technology⁶

Copyright issues related to language resources have received considerable attention. Personal data protection in the field of language research is an area which is not thoroughly explored yet. This topic is very relevant since the General Data Protection Regulation is applicable from 25 May 2018. The authors analyse the issues relating to personal data protection in the context of the development and utilization of language resources. Firstly, the authors analyse the cases where there is a consent which allows processing of personal data to develop and use language resources for research purposes. Secondly, the processing of personal data without consent for language research purposes is explored. The article is meant to provide practical information to language research community and create a foundation for further discussion.

Keywords: language resources, personal data protection, processing personal data for research purposes, consent for processing personal data