

Noorte kohta kogutavate andmete taaskasutusprojekt Statistikaametis

Ülevaade projekti esimese etapi raames valminud andmeallikate kaardistusest

Tellijaja: Eesti Noorsootöö Keskus (kontaktisik: Anne Kivimäe Anne.Kivimae@entk.ee)

Töö teostaja: Statistikaamet (kontaktisik: Eve Telpt eve.telpt@stat.ee)

Projekti esimese etapi eesmärk: Koostada ülevaade andmetest, mida noorte (vanusegrupp 7–26) kohta regulaarselt kogutakse riiklikes andmekogudes, suurtes sotsiaaluuringutes jt kättesaadavates kvaliteetsetes allikates.

Infoallikad: Andmeallikate kaardistus on põhiliselt koostatud internetis avaldatud informatsiooni põhjal.

Riiklike andmekogude ehk registrite kaardistamisel on peamisteks infoallikateks [Riigi infosüsteemi haldussüsteem](#) (RIHA) ja andmekogude põhimäärused. RIHA-st on võimalik registreid nime, lühinime, seotud asutuse jt tunnuste alusel otsinguga otsida (vt Riigi infosüsteem – Infosüsteemid). Üldjuhul on RIHA-s alamjaotuses „Alusdokumendid“ toodud ka viide registri põhimäärusele.

Uuringute kaardistamisel on peamiselt kasutatud ministeeriumide, riigiametite ja teiste asutuste veebilehti. Enamikul asutustest on veebilehel eraldi uuringute ja statistika rubriigid, kus on aastate või teemade kaupa esitatud tellitud uuringute aruanded, statistika ülevaated, valdkondlikud analüüsid jms.

Kaardistatavate allikate valimine: Nii registrite kui ka uuringute valimisel on lähtutud põhimõttest, et kaardistatav allikas peab andma informatsiooni noorte kohta. Seetõttu on kaardistamisel jälgitud, et allikast kindlasti selguks sihtrühma vanus. Valdkondi üldiselt iseloomustavad allikad, kus ei ole võimalik noori eristada, on jäetud kaardistamata.

Kaardistusest on välja jäetud ka väga spetsiifilisel teemal või konkreetse asutuse kohta teavet koguvad andmeallikad.

Kaardistuse jaoks on valitud välja ainult korduvad uuringud (st toimunud mitmel korral ja/või teada, et toimub ka tulevikus). Kuna registreid täiendatakse pidevalt, siis kasutusel olevate registrite puhul ei ole regulaarsuse põhimõtet eraldi jälgitud.

Kaardistatud allikate loetelu ei ole kindlasti kõikehõlmav ega lõplik.

Töö teostamise aeg: Andmeallikaid kaardistati 2016. aasta novembrist 2017. aasta juunini, koostades esmalt üldisem loetelu allikatest ja hiljem täiendades seda täpsema informatsiooniga. Kaardistustabelis kajastub informatsioon seisuga mai–juuni 2017.

Kaardistustabelit kasutades on vaja meeles pidada, et internetist leitav informatsioon ei ole täielik ning registrite ja uuringute info muutub jooksvalt. Seetõttu peab arvestama võimalusega, et tabelis kirjas olev teave võib olla ebatäpne, puudulik või vananenud (nt registri põhimäärust on muudetud, register ei kogu tegelikult kõiki põhimääruses kirjas olevaid andmeid, uuringu uues laines on muudetud sihtrühma või meetodikat). Mõne konkreetse allika vastu täpsemat huvi tundes tasub alati külastada tabelisse lisatud veebilinke, et leida sealt uusimat infot, vahepeal avaldatud andmeid, ilmunud aruandeid vms. Paraku ei ole ka veebilingid igavesed, sest seoses veebilehtede uuendamisega võivad need muutuda. Näiteks, hetkel valmistatakse RIHA uue versiooni kasutusele võtmiseks 2017. aasta lõpus. Kuivõrd see mõjutab kaardistustabelisse lisatud veebilinkide toimimist, ei ole teada.

Andmeallikate kaardistustabeli sisu ja ülesehitus

Andmeallikad on kaardistatud Exceli-tabelina. Uuringute ja registrite kohta on tabelisse kantud mõnevõrra erinev info, sest tegemist on oma olemuselt erinevate andmeallikatega. Täpsemad selgitused leiab järgnevalt loetelust, kus on tabeli iga veeru sisu põhjalikumalt kirjeldatud. Samalaadsed, aga lühemad selgitavad kommentaarid on lisatud ka Exceli-faili päises asuvate veeru nimetuse lahtrite juurde.

Andmeallika liik: Eristatud kolme liiki allikaid: uuring (valikuuring, küsitlus), register (RIHA-s kirjas olev riiklik andmekogu) ja muu allikas. Soovi korral saab välja filtreerida ainult endale huvipakkuvat liiki allikad.

Andmeallika nimetus: Täisnimetus, kasutusel olev lühend, rahvusvaheliste uuringute korral ka inglisekeelne nimetus ja lühend.

Uuringu tellija/teostaja, registri haldaja/pidaja: Asutus, kes selle andmeallika taga seisab ja kelle poole saab lisainformatsiooni saamiseks pöörduda. Uuringu korral tellija ja teostaja – kui teostajad muutuvad (nt erinevatel aastatel erinevad uuringufirmad), on kirjas ainult tellija. Registri korral haldaja (vastutav töötaja) ja pidaja (volitatud töötaja) – mõne registri puhul on tegemist sama asutusega.

Valdkond: Valdkonna üldnimetus, mis aitab sama valdkonna allikaid välja filtreerida ja koos vaadata. Paljude allikate puhul on neid valdkondi mitu, sest allikat ei ole võimalik kirjeldada ainult ühe märksõna kaudu.

Teemad: Loetelu teemadest, mida allikas kajastab. Registrite puhul annab see veerg ülevaate ka sihtrühmast, kelle andmeid register koondab.

Uuringu sihtrühm, vastanud: Vastajate arv viimases uuringulaines ja sihtrühma kirjeldus (vastajate vanusevahemik või muud tunnused, mis uuringu sihtrühma määratlevad). Registrite kohta ei ole seda lahtrit täidetud, sest need on üldjuhul kõiksed (st hõlmavad üldkogumit tervikuna).

Kaetud noorte vanus: Noorte (7–26 a) vanusevahemik, mida allikas katab. Kui tegemist on andmetega, mille puhul ei ole vanusepiirid teada, on selle asemel märgitud kolm punkti (nt ... –26).

Moodustatud noorte vanuserühmad: Noorte vanuserühmad, mille kohta on andmed avaldatud, näiteks avalikus andmebaasis, aruandes vm. Kui ei ole vanuserühmi moodustatud (nt olemas vaid vanuse/sünniaasta/isikukoodi tunnused), on lahtrisse märgitud "Puuduvad".

Senine periood: Seniste uuringute toimumisaeg – üksikutel kordadel või harva toimunud uuringute korral konkreetset aastat, pikema perioodi korral ajavahemik. Registrite kohta on RIHA põhjal märgitud, kui kaua on andmeid kogutud (st mis ajast pärinevad vanimad andmed).

Perioodilisus: Veeru eesmärgiks on anda infot, kui sageli tekib uusi andmeid juurde. Uuringu kohta on märgitud toimumise intervall, mõnel juhul ka järgmise uuringu toimumise aeg. Registritel, kus andmeid uuendatakse pidevalt, on lahtris vastav märg ("Pidev").

Metoodika: Uuringu küsitlusmeetod ja vajadusel ka muud selgitavad märkused uuringu kohta. Registrite kohta ei ole seda lahtrit täidetud, aga lisatud veebilinkide abil on võimalik põhimäärusest ja/või RIHA-st vaadata, kes on andmeesitajad, milline on andmete esitamise kord jne.

Isik tuvastatav: Jah/ei vastus, kas selles allikas on isik otseselt tuvastatav, st andmed on isikukoodi (või nime, soo ja sünniaja) kaudu konkreetse isikuga kokku viidavad. Isikukoodi olemasolu andmebaasis võimaldab andmeid linkida teise allikaga, kus isikud on samuti tuvastatavad. Kuigi isikukoodi andmeid kogutakse, ei tähenda see, et seda infot väljastatakse kolmandatele osapooltele. Näiteks, Statistikaamet pakub võimalust oma allikaid teiste allikatega siduda, aga linkimist teostab ameti vastavaid õiguseid omav töötaja, ning andmete taotleja saab juurdepääsu andmebaasile, kust on eemaldatud otsest tuvastamist võimaldavad tunnused.

Võrreldav teiste riikidega: Jah/ei vastus, kas on tegemist rahvusvahelise uuringuga, kus on võimalik Eesti andmeid võrrelda teiste riikidega. Võimalusel lisatud ka täpsustus, milliste riikidega. Registri kohta ei ole seda lahtrit täidetud, aga paljud kõiksed näitajad, mida registritest saab, kindlasti on võrreldavad teiste riikide registrite andmetega.

Avaldatud andmed: Veebilingid, kus on avaldatud tabelid uuringu andmetega, uuringu aruanded või registri põhjal koostatud statistika. Aruande või analüüsina avaldatud andmete korral on lisatud viimati avaldatud ülevaate pealkiri, vajadusel ka avaldamisaeg ja aruande koostanud asutuse nimi.

Algandmete kättesaadavus: Võimalusel lisatud veebilink, kust on võimalik leida andmed või info nende saamise kohta (nt andmete saamise tingimused, juhised andmete taotlemiseks). Kui internetis ei ole informatsiooni algandmete saamise võimaluse kohta, tuleb soovi korral võtta ühendust uuringu tellija või registri haldajaga.

Andmete kättesaadavus ja kehtestatud tingimused võivad erinevatel allikatel olla väga erinevad. Mõne uuringu (valdavalt tegemist rahvusvaheliste uuringutega) andmestikud on veebist vabalt kättesaadavad. Paljude allikate andmetele aga pääsevad ligi ainult teatud ametite esindajad ja teatud tingimustel, näiteks avalike ülesannete täitmiseks, ametialaseks kasutamiseks, teadustöö tegemiseks. Registriandmete puhul on üldreegliks, et juurdepääs andmetele võimaldatakse kooskõlas avaliku teabe seaduse ja isikuandmete kaitse seadusega. Konkreetse registri andmete kasutamise võimalused on kirjas põhimääruses, teatud juhtudel ka RIHA-s. Kaardistustabelisse on lisatud viide põhimääruse vastavale peatükile või paragrahvile. Statistikaamet lähtub konfidentsiaalsete andmete kättesaadavaks tegemisel riikliku statistika seadusest.

Allika veebilink, kontaktid: Veebilingid, mis andmeallikat tutvustavad. Võimalusel lisatud ka kontaktisiku nimi (nt rahvusvahelise uuringu koordinaator Eestis, tellijapoolne kontaktisik, aruande koostaja).