

Andmepäringute koosseis ja andmete liikumine

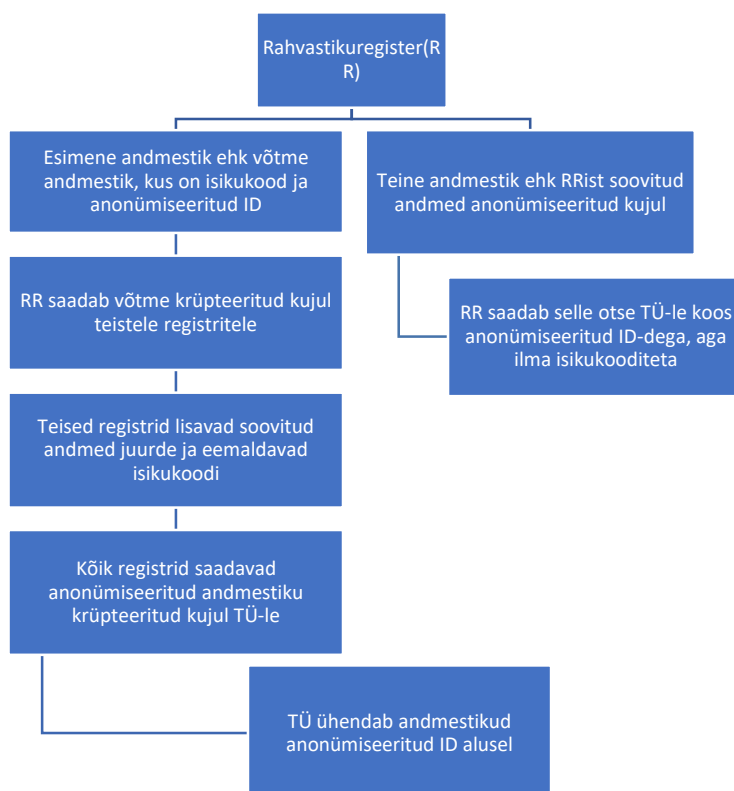
Uuringus on vaja pärida erinevatest registritest andmeid ja ühendada need isikupõhiselt. Selleks esitab uuringu läbiviija (Tartu Ülikool) koostöös tellijaga (poliitikaväljundi tellija Rahandusministeerium, lepingu järgne tellija SA Eesti Teadusagentuur (ETAg)) andmepäringu Rahvastikuregistrile.

Rahvastikuregister koostab **andmete sidumiseks vajalikud andmebaasid** ehk isikukoodide ja genereeritud unikaalse koodidega andmebaasi (edaspidi andmestik 1) ja välismaalaste lühiajalise Eestis töötamise registreerimise andmekogu eripärasid (sageli puuduv isikukood) arvestava andmete sidumiseks vajalik andmestiku, mis sisaldab genereeritud unikaalseid koode, isikukoodi, inimese sünnikuupäeva ja nime (andmestik 2). Rahvastikuregister saadab andmete sidumiseks vajaliku andmebaasi (andmestik 1) krüpteeritult teiste andmepäringusse kaasatud registrite volitatud töötlejatele. Erandiks on välismaalaste lühiajalise Eestis töötamise registreerimise andmekogu volitatud töötleja, kellele saadetakse andmestik 2.

Teised registrid lisavad Rahvastikuregistri saadetud andmete sidumiseks vajalikule andmebaasile nende registrist päritavad andmed, eemaldades isikukoodid (lühiajalise töötamise registri puhul ka nimed ja sünnikuupäevad) ja jättes alles genereeritud unikaalsed koodid. Seejärel saadavad kõik registripidajad andmefailid Tartu Ülikoolile, kes seob kõikidelt registritelt saadud andmefailid, mille tulemusena loodud andmebaas on kvantitatiivanalüüsi aluseks. Kuna analüüsi teostamiseks on vaja Rahvastikuregistrist ka muid andmeid peale andmete sidumiseks vajalike andmestike, siis nende andmetega viib Rahvastikuregister läbi sama protseduuri, mis teised registrid.

Registrite vahel muud andmevahetust peale Rahvastikuregistrist teistele registritele saadetavate andmete sidumiseks vajalike andmebaaside (andmestik 1 ja andmestik 2) liikumise ei toimu. Töö teostajale isiku otsest tuvastamist võimaldavaid andmeid (isikukood, nimi) ei edastata. Töö tellijale edastatakse vaid lõpparuanne.

Joonis 1. Andmete liikumine registrite vahel ja nende jõudmine TÜ-ni



Sellist andmete ühendamise lahendust on varasemalt korduvalt kasutatud. Näiteks võib siin tuua Riigikogu Arenguseire keskuse poolt tellitud uuringu „Tööga seotud sotsiaalkaitse mudelid ja nende sobivus alternatiivsete tööturu arengute korral Eestis“ (AKI luba antud 31.05.2018, otsuse number nr 2.2.-1/18/22) või käesoleva uuringuga sarnaselt ETAgI poolt RITast rahastatud uuringu „Muutuvast tööturust ja kasvavast rändest tulenevad võimalused ja väljakutsed haridus --, tööturu ja sotsiaalsüsteemis“ (AKI luba antud 29.11.2018, otsuse nr 2.2 1/18/45, luba taotles konsortsiumi liige Eesti Rakendusuuringu Keskus CentAR).

Andmete päringud registrite kaupa

Need andmed edastatakse TÜ-le umbisikustatult st. isikukoodid on eemaldatud ja iga valimiisik on tähistatud unikaalse koodiga (vastab Rahvastikuregistri edastatud valimiisikute andmebaasi koodile isiku kohta).

1. Rahvastikuregister (Siseministeerium)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** kõik 2016. aasta jaanuari seisuga vähemalt 15-aastased ja 2020. aasta 1. novembri seisuga kuni 75-aastased inimesed
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused**
 - **Andmestik 1** (võtme dokument, mis saadetakse teistele registritele, väljaarvatud Välismaalaste lühiajalise Eestis töötamise registreerimise andmekogu):
 - Anonümiseeritud ID
 - isikukood
 - **Andmestik 2** (võtme dokument, mis saadetakse Välismaalaste lühiajalise Eestis töötamise registreerimise andmekogule):
 - Anonümiseeritud ID
 - isikukood
 - Nimi
 - Sünnikuupäev
 - **Andmestik 3** (analüüsis kasutatavad andmed, mis saadetakse töö teostajale):
 - Anonümiseeritud ID
 - Sünniaeg
 - Surmaaeg
 - Sugu
 - Kodakondsus
 - Kõrgeim omandatud haridustase
 - Elukoht omavalitsuse tasandil
 - Kõrgeim omandatud haridus

2. Elamislubade ja töölubade register (Politsei- ja Piirivalveamet)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadatud andmestikus (andmestik 1) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused:**
 - Anonümiseeritud ID
 - Eestis viibimise alus
 - Elamisloa taotlemise, pikendamise ja kehtetuks tunnistamise andmed
 - **Kõige esimese** elamisloa väljastamise alus ja aeg kehtivuse lõppemise aeg ja kehtetuks tunnistamise aeg
 - Kõigi perioodil 01.01.2016 – (väljavõtte tegemise aeg) isikule väljastatud elamislubade väljastamise alus ja aeg kehtivuse lõppemise aeg ja kehtetuks tunnistamise aeg
 - Elamisloa taotleja kodakondsus
 - Elukoha aadress omavalitsuse täpsusega
 - Töökoha aadress omavalitsuse täpsusega

- Elamisõiguse kehtivusaeg
- Haridus taotluse esitamise ajal

3. Välismaalaste lühiajalise Eestis töötamise registreerimise andmekogu (Politsei- ja Piirivalveamet)

Lisaks eelpool loetletud andmekogudele on uuringu läbiviimiseks vaja andmeid ka Välismaalase lühiajalise Eestis töötamise registreerimise andmekogust (LTR), mille vastutav töötleja on Politsei- ja Piirivalveamet. Selle andmekogu eripära seisneb aga selles, et sinna ei kanta üldjuhul inimese isikukoodi. Seetõttu pole võimalik selle andmekogu andmeid ka isikukood alusel siduda. Seetõttu tuleb kasutada teisi tunnuseid – nimi ja sünnikuupäev. LTR-i andmete sidumine käib seetõttu järgmise mehhanismi alusel:

- Rahvastikuregister saadab LTR-i modifitseeritud RR I andmebaasi, millele sisaldub lisaks pseudoisikukoodile inimese nimi ja sünnikuupäev (nimetame selle RR I andmebaasiks).
- LTR lisab nime ja sünnikuupäeva alusel RR I andmebaasi oma andmed, kustutab isiku nime ja sünnikuupäeva (jättes alles pseudoisikukoodi) ning edastab selle uuringu läbiviijatele.
- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 2) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused:**
 - Anonümiseeritud ID
 - Eestis viibimise alus
 - Kontaktaadress omavalitsuse täpsusega
 - Töökoha aadress
 - Tööandja põhitegevusala (EMTAK koodi alusel)
 - Töökoha nimetus, millel välismaalane tööle asub
 - Töötamise liik vastavalt välismaalaste seaduses sätestatud alusele;
 - Töötamise eest makstava brutotöötasu suurus
 - Töötamise planeeritav algus- ja lõpuaeg
 - Haridus taotluse esitamise ajal (kui on olemas)

4. Maksukohustuslaste register (Eesti Maksu- ja Tolliamet)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 1) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused (indiviid):**
 - Anonümiseeritud ID
 - Tööandja anonümiseeritud ID
 - Kodakondsus
 - Sugu
 - Vanus
 - Tööandja põhitegevusala (EMTAK koodi alusel)
 - Töökoha asukoht omavalitsuse täpsusega
 - Ametnimetus
 - Isiku (bruto) sissetuleku kogusumma tululiikide lõikes
 - Tööleping
 - ATS
 - võlaõiguslik leping (käsundusleping + töövõtuleping)
 - juhtimis-, kontrollorgani liige

- muu
- Isikule tema haigestumise tõttu tööandja poolt deklareeritud töövõimetushüvitis
- **Tunnused (ettevõtte):**
 - Tööandja anonümiseeritud ID (isiku anonümiseeritud ID ja tööandja ID ei pea olema kokkuviidavad, ehk siis võib olla teise võtme alusel anonümiseeritud ID, kui indiviidi andmete juures esitatav tööandja ID)
 - Deklareeritud käive
 - Deklareeritud tööjõukulu
 - Deklaratsiooni kohane töötajate arv
 - Deklareeritud ettevõtte tulumaks (võimalused eristada dividenditulu või sellelt makstud tulumaks)

5. Tööotsijate ja tööturuteenuste riiklik register (Eesti Töötukassa)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 1) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused:**
 - Anonümiseeritud ID
 - Töötasu hüvitise saamise kuupäevad
 - Töötasu hüvitise kuine määr
 - Registreeritud töötuna arveloleku kuupäevad
 - Töötutoetuse saamise kuupäevad
 - Töötutoetuse päevamäär
 - Töövõime (osaline, puuduv) (kõik kirjed)
 - Algus
 - Lõpp
 - Protsent
 - Töövõimetoetuse summa

6. Töötuskindlustuse andmekogu (Eesti Töötukassa)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 1) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused**
 - Anonümiseeritud ID
 - Töötuskindlustushüvitise saamise algus- ja lõppkuupäevad
 - Töötuskindlustushüvitise päevamäär esimesel sajal päeval
 - Töötuskindlustushüvitise päevamäär 100+ päeval

7. Sotsiaalkaitse infosüsteem (Sotsiaalkindlustusamet)

- **Ajaperiood:** 01.2017 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 1) olevatest inimestest
- **Andmete ajaline täpsus** – kuine
- **Agregeerituse tase:** indiviidi tasemel

- **Tunnused:**
 - Anonümiseeritud ID
 - Töövõimetus (kõik kirjed)
 - Algus
 - Lõpp
 - Protsent
 - Puude raskusaste
 - Keskmine, raske või sügav
 - Puude raskusastme tuvastamise päev

8. Eesti Hariduse Infosüsteem (Haridus- ja Teadusministeerium)

- **Ajaperiood:** 01.2016 – andmete väljavõtte hetkel viimane seis
- **Valikukriteerium:** Väljavõtte Rahvastikuregistri saadetud andmestikus (andmestik 1) olevatest inimestes
- **Agregeerituse tase:** indiviidi tasemel
- **Tunnused:**
 - Anonümiseeritud ID
 - Isiku kõrgeim omandatud haridustase (ISCED2011 klassifikaatori järgi ja iga kalendriaasta lõpu seisuga (2020. aasta andmed viimane teadaolev info))
 - Isiku kesk-/kõrg-/kutsehariduse (ISCED2011 klassifikaatori järgi) õpingutele sisseastumise aeg
 - Isiku kesk-/kõrg-/kutsehariduse (ISCED2011 klassifikaatori järgi) õpingute katkestamine või õppepuhkuste kasutamine (kõik sündmused, kuu täpsusega)