

10+ aastat registreite ja suurandmete kasutamist statistikas

Kaja Sõstra, PhD

Maiki Ilves

Metoodika ja analüüsi osakond

Statistikaameti teabepäev tarbijatele

01.03.2017



REGISTRIANDMED RIIKLIKUS STATISTIKAS

Eeldused registrite kasutamiseks riikliku statistika tegemisel

- Seaduslik alus – toetab andmekogude kasutamist statistika tegemiseks
- Avalik arvamus – põhiliselt toetav, võib tekkida hirm “superandmebaasi” ees
- Unikaalsed identifikaatorid – põhiliselt olemas, probleeme on veel aadressistandardi rakendamisega
- Registrite süsteem – registrite koostoime veel nõrk, “administratiivne” kvaliteet pole alati sarnane “statistilise” kvaliteediga
- Koostöö – vajame palju infot registri sisu ja tehtavate muudatuste kohta

Andmekogude kasutamise viisid

- Täielikult andmekogude andmetel põhinevad statistikatööd
 - Andmekogust koondandmed
 - Andmekogust üksikandmed
- Osaliselt andmekogudel põhinevad statistikatööd
 - Hinnangute täpsustamiseks, imputeerimiseks
 - Väiksemate üksuste jaoks andmed andmekogudest
 - Teatud tunnused andmekogudest
 - Küsimustiku eeltäitmiseks
 - Lähteandmebaasi täiendamiseks

Klassikaline uuring

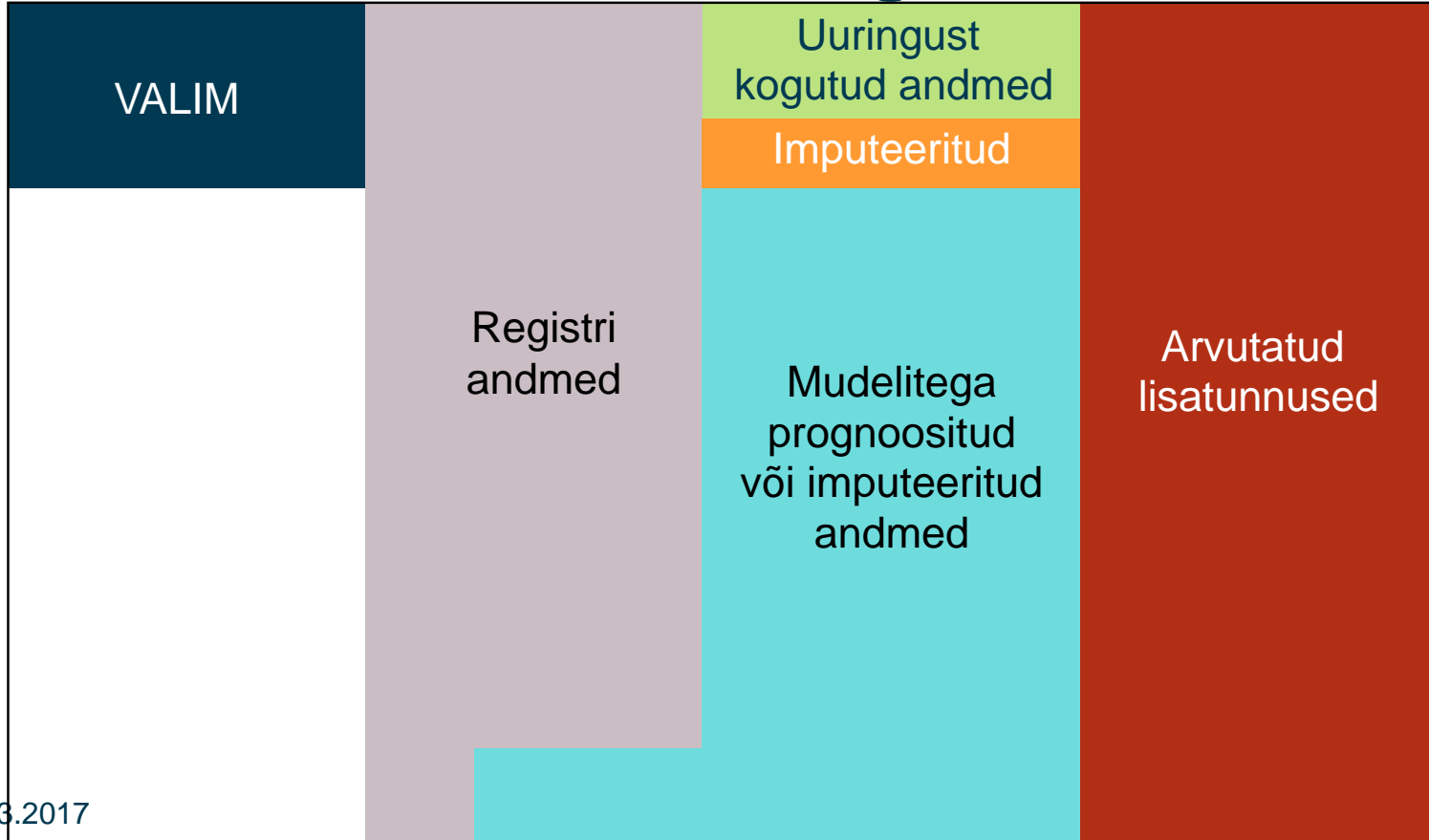
VALIM

Uuringust
kogutud
andmed

Imputeeritud

Arvutatud
lisatunnused

Kombineeritud uuring



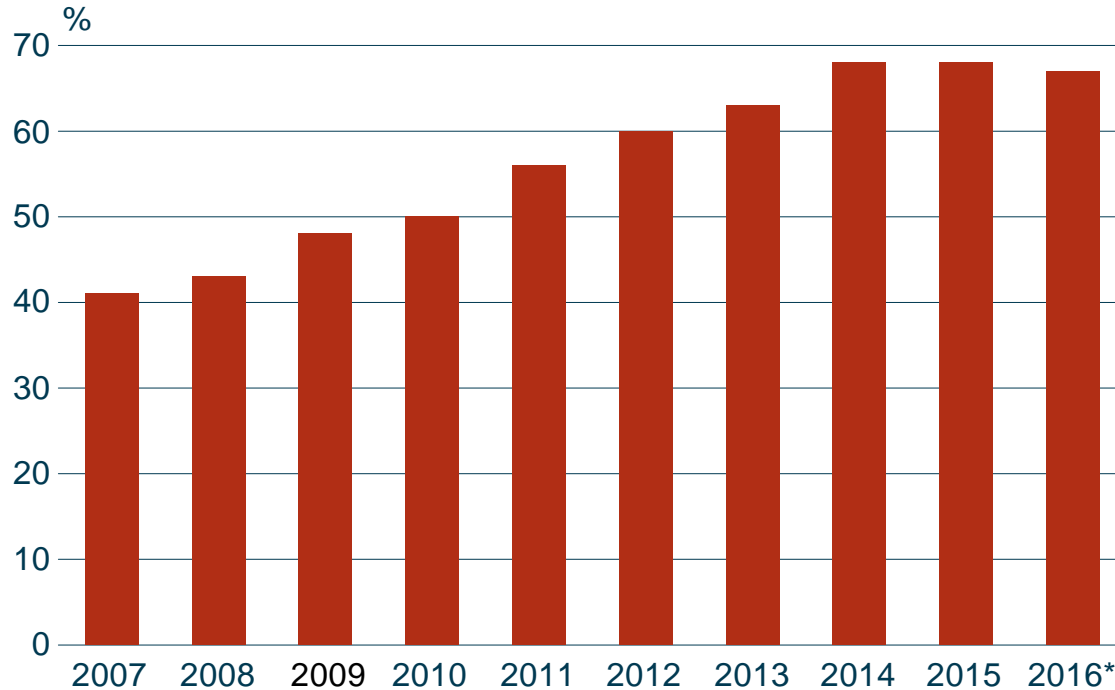
Üleminek andmekogude kasutamisele

- Ettevalmistus üleminekuks
 - Andmekogu metaanalüüs (kaetus, definitsioonid)
 - Andmekogu detailanalüüs (andmekvaliteet)
 - Andmehõive korraldamine (leping, IT-lahendus)
 - Andekogumise ja andmetöötamise juurutamine (paralleelne tootmine)
- Andmekogumise struktuuri muutus
 - Väheneb andmekogumise maht ja koormus isikutele või majandusüksustele
- Andmetöötamise struktuuri muutus
 - Andmetöötamise maht suureneb tunduvalt

Andmetöötluse mahu muutus

- Suureneb tunduvalt üksuste arv (kordades), mis suurendab andmekontrolli mahtu
- Suurematelt majandusüksustelt peab andmeid edasi koguma
- Lisandub eri andmeallikate andmete seostamine
- Suureneb imputeerimise maht, sest ka andmekogudes on puuduvaid andmeid
- Lisandub statistiliste näitajate moodustamise vajadus

Statistikatööde osatähtsus, kus on kasutatud andmekogude andmeid



2016* muudeti statistikatööde loetelu, mille tõttu andmed ei ole võrreldavad

Registrite kasutamise algus

- 1994. aastal loodi majandusüksuste statistiline register, mille põhiliseks sisendiks oli Ettevõttere register, hiljem Äriregister, alates 2003 ka Maksu- ja Tolliameti andmed
- Rahvastikustatistika aluseks olid Siseministeeriumi, Kodakondsus- ja Migratsiooniameti, AS Andmevara andmed
- Keskkonnastatistika tegemiseks andmed Keskkonnaministeeriumilt
- Haridusstatistika tegemiseks andmed Haridusministeeriumilt
- Veel saadi andmeid: Rahandusministeerium, Sotsiaalministeerium, Tööturuamet, Sotsiaalkindlustusamet, Eesti Pank, Haigekassa, Tööinspeksioon

Registrite kasutuselevõtt 2007-2011

- 2007 – aruande „Raudtee ja veeremi olem ning liiklusõnnetused” asemel saadi andmed riiklikust raudteeliiklusregistrist
- 2008 – statistikatöö “Kinnisvara” puhul võeti kasutusele Maa-ameti tehingute andmebaas
- 2008 – statistikatöö „Põllumajanduskultuuride kasvupind“ jaoks saadi andmed PRIA-lt
- 2009 – statistikatöö „Kindlustus” jaoks saadi andmed peale kindlustusseltside ka Finantsinspeksioonilt, Eesti Liikluskindlustuse Fondilt ning Maksu- ja Tolliametilt
- 2010 – RR, EHIS, MTA andmete kasutatavuse uurimine REL jaoks, REGREL ettevalmistuse käivitamine
- 2011 – isikute tulusid sisaldavate registriandmete analüüs

Registrite kasutuselevõtt 2012-2016

- 2012 – statistikatöös “Ettevõtete aasta majandusnäitajad” alustati eeltäitmist majandusaasta aruandest
- 2014 – statistikatöö “Ettevõtete kvartali majandusnäitajad” eeltäitmine MTA deklaratsioonide andmetega
- 2015 – statistikatöös “Töötasu struktuur” Haigekassa ja hariduse infosüsteemi andmete kasutamine
- 2016 – statistikatöö “Müügitulu” tegemiseks kasutatakse ainult MTA käibedeklaratsiooni andmeid
- 2014-2016 – statistikatöö “Ettevõtete aasta majandusnäitajad” väikeste ettevõtete andmed Äriregistri majandusaasta aruandest

MAA andmete kasutamine

- 2012 lõpetas SA majandusaasta aruannetega dubleerivate andmete kogumise
 - Aruanded eeltäidetakse
 - Mikroettevõtetest enamasti ei küsita aruannet
- Eelnes 4 aastat eri asutuste koostöös
- Keskmiselt 80% aruande EKOMAR lahtritest on eeltäidetud ja 20% tuleb raamatupidajal ise täita
- EKOMARi aruandluskoormus vähenes 2 korda

SUURANDMETE KASUTAMINE

Suurandmete edulood

- Liiklussensorite andmetel liiklusintensiivsuse statistika (Holland) - <https://www.cbs.nl/en-gb/background/2015/31/a13-busiest-national-motorway-in-the-netherlands>
- Kaupluste skännerite andmed hinnastatistika kokkupanemiseks (Belgia, Suurbritannia)
- Väliskülastajate statistika mobiiliandmete pealt (Eesti) - http://statistika.eestipank.ee/?lng=en#treeMenu/MAKSEBIL_JA_INVPOS/1410

Projektid Statistikaametis

- “Feasibility study on the use of mobile positioning data for tourism statistics”
- ESSnet Big Data: elektritarbimise tunniandmete pilootprojekt
- Testid veebikammimise teel andmete saamiseks

ESSnet Big Data projekt

- Projektis osaleb 22 organisatsiooni 20 riigist.
- 5 pilootprojekti perioodil 2016-2018
 1. Web scraping of job portals
 2. Web scraping of businesses' websites
 3. Smart meters
 4. Automatic Identification System data (maritime traffic)
 5. Mobile phone data

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page

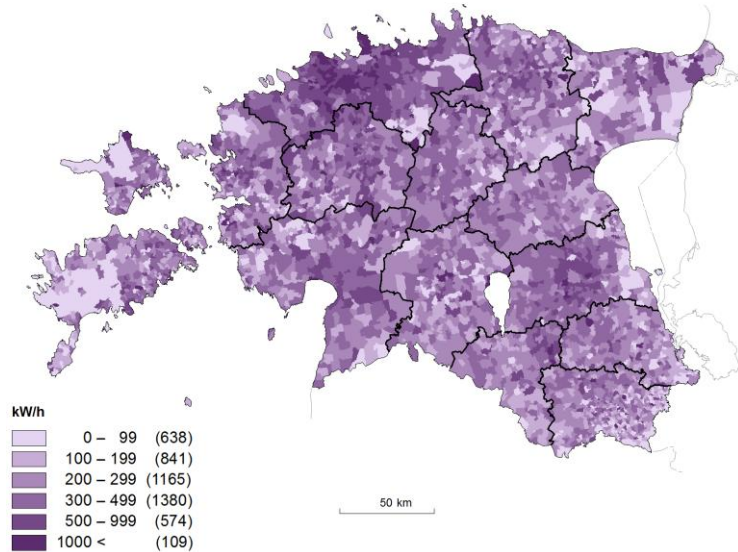
Projekti eesmärgid

- Arvutada statistika koos kvaliteedihinnangutega:
 - leibkondade elektritarbimise kohta (leibkonna tüübi ja elukoha näitajate kaupa) – uus väljund
 - Ettevõtete elektritarbimine tegevusalati – hetkel uuringuandmetel põhinev väljund
 - Eluruumide kasutatavuse hindamine (tühi, periooditi tühi, kasutusel) – uus väljund, sisend põhieluruumi määramiseks

Esialgused tulumused – leibkondade elektritarbimine

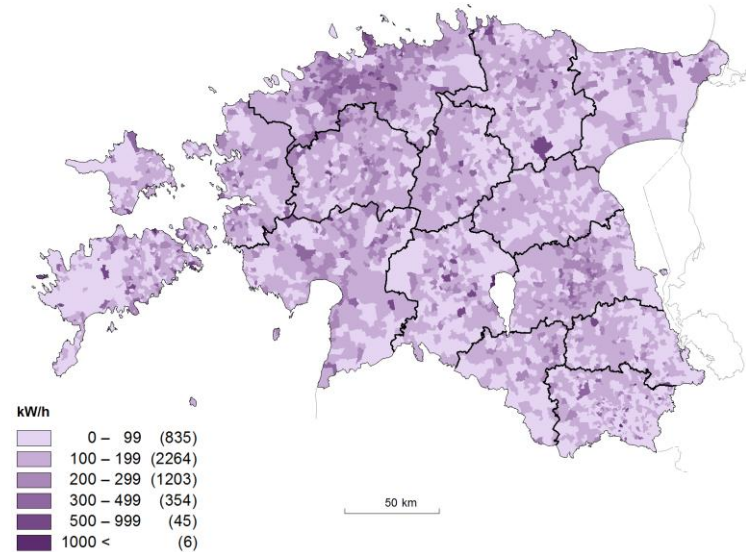
January

Average monthly electricity consumption of private persons, January 2014



July

Average monthly electricity consumption of private persons, July 2014



Registriandmete kasutamise tulevik

- SA on jõudnud registriandmete kasutamisel maksimumi lähedale
- Andmekogude kasutamise suurendamiseks on järgmised võimalused:
 - Kasutusel olevate andmekogude andmete efektiivsem kasutamine
 - Olemasolevate andmete kombineerimisel täiendava statistika tegemine
 - Andmekogude täiendamine statistika vajadustest lähtuvalt
- Suurandmete kasutuselevõtt
- Andeallikate kombineerimine – küsitlus, registriandmed, suurandmed



REPUBLIKA
SRBIJA
STATISTIKA