ANDMEKORRALDUSE RAAMISTIK 2.0

koolitus

15.10.2026

Veiko Berendsen

EESTI STATISTIKA

KOOLITUSE TEEMAD

1. **Kuidas saada tervikülevaade andmetest** (andmestikest, andmevaradest). Andmekirjelduse ja andmete ülevaate kaks eesmärki: (1) sisemine: andmete haldamise võimaldamine asutuses; (2) väline: andmete (üles)leidmine ja (taas)kasutamise võimaldamine. Milline on minimaalselt vajalik andmekirjeldus, milline täielikum kirjeldus ja kes on nende kasutajad? Sõnastike praegune ja edasine roll.
2. ~~Kas ja kuidas on asutuse sisemised **andmehalduse vajadused** kokkuviidavad üleriigiliste avaandmete eesmärkidega?~~
3. **Arutelu.** Milliste töövahendite ja protsessidega on võimalik saada terviklik ülevaade andmetest? Kuidas jõuda tervikülevaateni, kui alustame näiteks (tähtsamast) osast või andmekogudest?
4. **Andmekvaliteedi kontseptsioon.** Kas põhimõte fit for purpose on rakendatav? Andmekvaliteet ja ärireeglid. Andmekvaliteet andmetöötluses ja väljundis. Loendite ja klassifikaatorite roll andmekvaliteedis.
5. **Praktiline osa.** Andmekvaliteedi haldusvõimaluste rakendamine tarkvara SelectZero abil. Võimaluste ja tulemuste tutvustus (kuidas hinnata ja parandada andmekvaliteeti).
6. **Arutelu** andmekvaliteedi haldamise rakendamisest asutuses: (1) millised on dimensioonid ja indikaatorid; (2) millistele andmestikele rakendada; (3) andmekvaliteedi jälgimine protsessides.

Tervikülevaade andmetest (andmestikest, andmevaradest)

EESTI
STATISTIKA

4. peatükk
Teabehalduse korraldamine

§ 11. Üldised nõuded

- Asutuse teabehalduse korraldus peab tagama:
  1. **teabe kvaliteedi** ja kättesaadavuse;
  2. teabe hoiu, vahetamise ja kasutamisega seotud riskide maandamise ja kulude vähenemise;
  3. teabehalduse järjepidevuse ametniku või töötaja teenistus- või töösuhte lõppemisel, asutuse töökorralduse muutmisel ning ametniku avaliku võimu teostamise õiguse peatumise või töötaja ajutise äraoleku ajal.

§ 12. Teabe korrastamine

1. Asutusel on **ülevaade oma põhiülesannete täitmisel tekkiva teabe, selle allikate ja hoiukohtade kohta**. Ülevaade **koostatakse protsesside analüüsi käigus** ning viisil, mis võimaldab seda lihtsasti ajakohastada.
2. .
3. Põhiülesande täitmisel tekkivast **teabest ülevaate saamiseks määrab asutus kindlaks**:
   1. millist teavet on põhiülesande täitmisega seotud teenuste osutamiseks vaja, lähtudes õigusaktiga sätestatud tingimustest;
   2. millist lisateavet põhiülesande täitmisel ja teenuste osutamisel luuakse või saadakse;
   3. millised on teabe allikad;
   4. millistes vormingutes ja hoiukohtades teavet hoitakse;
   5. millised on teabe säilitustähtajad ja juurdepääsutingimused;
   6. kes on teabe kasutajad.

2.2 Ülevaade andmetest ja andmekataloog

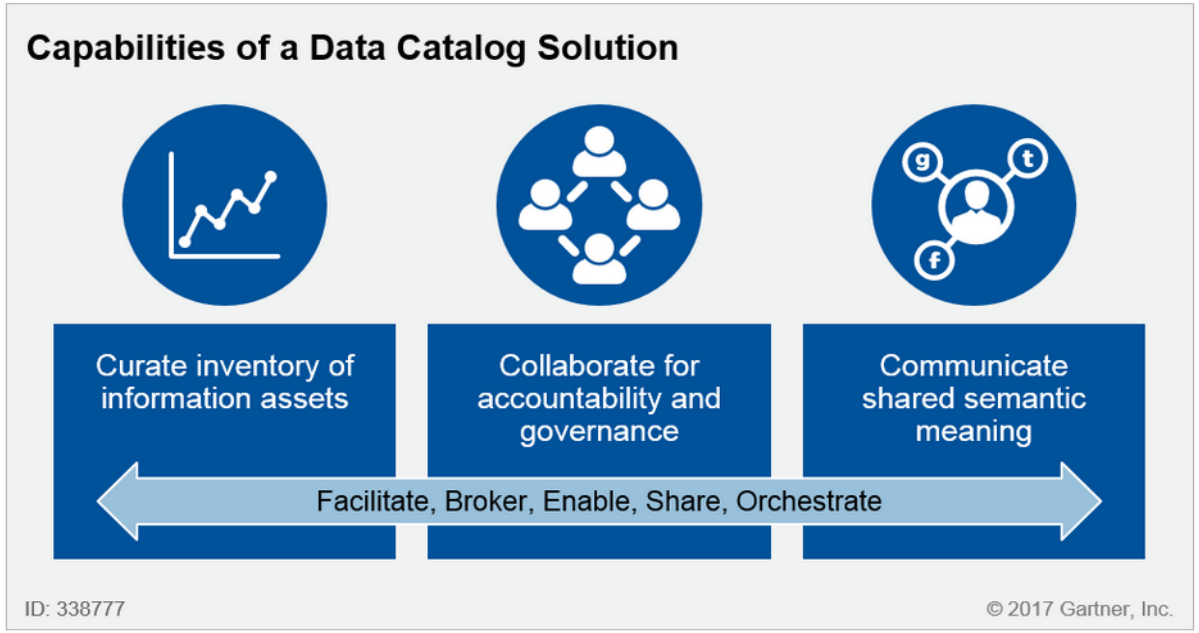
- Eesmärk
  - Organisatsioon vajab andmetest ja infost head ülevaadet, et teha põhjendatud ja kvaliteetseid otsuseid. Kuna andmemahud on suured ja mahud kasvavad pidevalt, on oluline suuta vajalik info kiiresti üles leida ning seda tulemuslikult kasutada.
  - Üheks ülevaate loomise oluliseks vahendiks on **andmekataloog**, mis toetab andmete süsteemset haldamist ning parandab nende leitavust ja kasutatavust.
- Taipamine
  - Organisatsioon peab seadma eesmärgiks saada täielik ülevaade, kuid alustada tuleb mingist kindlast otsast. Valida on vaja töövahend nagu andmekataloog. Kohe alguses on vaja teha selgeks, kes ülevaadet ajakohasena hoiab.

Eristada tuleks üsna rangelt:

- andmemudelite ja andmearhitektuuri arusaamist, mis on ülevaade andmetest
- andmevarade või andmestike nimekirja kui ülevaadet tükkide kaupa

Mida peaks võimaldama andmekataloog (või mis vahend on sobiv selleks, et saada tervikülevaade)

Gartner on andmekataloogi rakenduse funktsionaalsust kirjeldanud järgmiselt:



1. Infovarade inventuuri pidamine.
2. Koostöö andmetest ülevaate omamiseks ja andmehalduse osas.
3. Semantiliselt kirjeldatud andmete tähenduse jagamine.

(Infovara (*information asset*) ja andmevara (*data asset*) on siin sünonüümid)

Andmehalduse tööriistad ja tervikülevaade

Andmehalduse tööriistade kataloog

Tööriistatüüpide kirjeldused
Eristatud on viis peamist tööriistatüüpi koos selgitusega, millal neid vaja on ja mida neilt oodata.

- Andmekataloog
- Andmekvaliteedi platvorm
- Põhiandmete halduse platvorm
- Andmemodelleerimise tööriist
- Andmeintegratsiooni platvorm

EESTI STATISTIKA

Andmekataloog

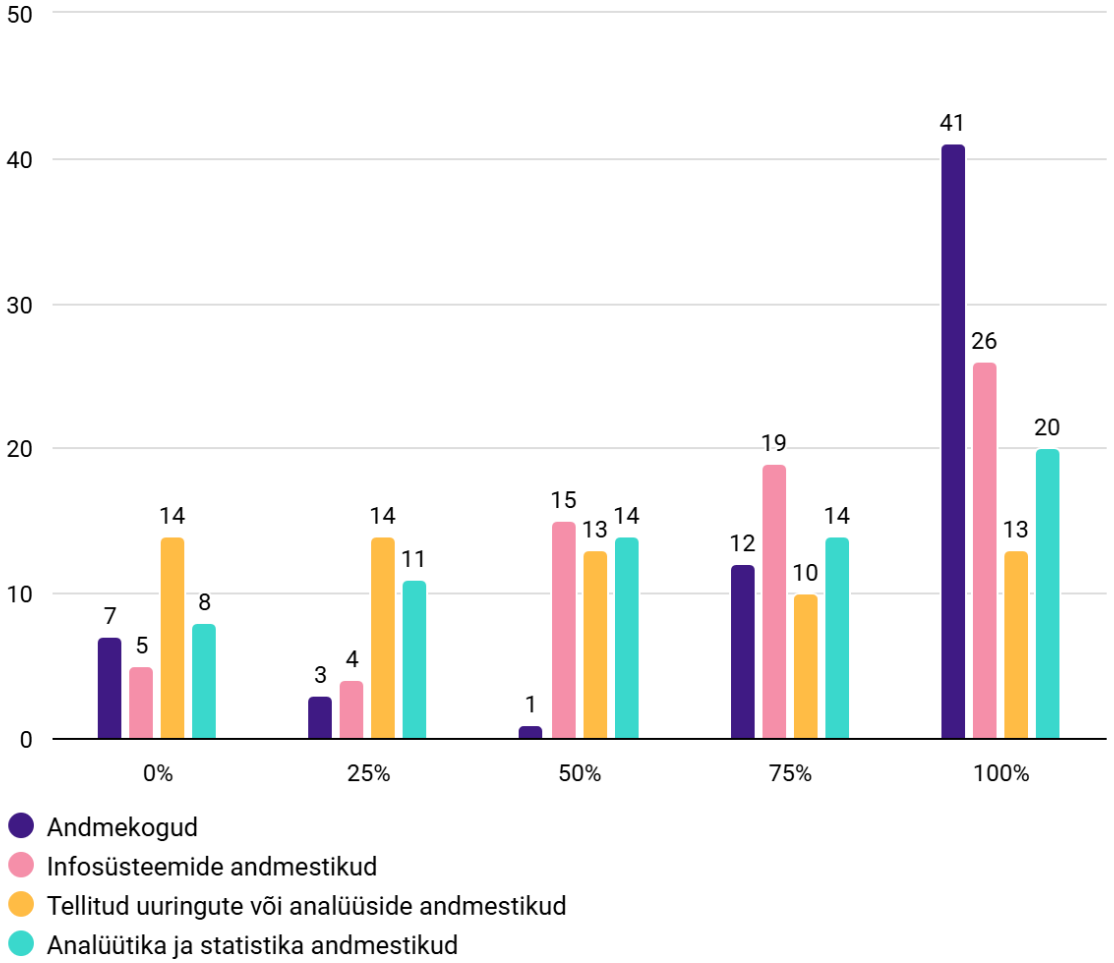
- Andmekataloog (ingl. k. *data catalog*) on organisatsiooni andmevarade otsitav ülevaade, kus hoitakse ja hallatakse andmevarade kirjeldust ehk metaandmeid (andmekirjeldust). Andmekataloog koondab eri süsteemidest pärinevad metaandmed, mis kirjeldavad andmete ärilist konteksti (nt definitsioonid, klassifikatsioonid), vastutusi, kasutusreegleid, kvaliteedisignaale ja seoseid, et kasutaja saaks ühest kohast aru, millised andmed on olemas, kus need asuvad ja kuidas neid eesmärgipäraselt kasutada.
- Baasfunktsionaalsused
  - Metaandmete haldus (ingl. k. metadata management)
  - Ühenduste haldus (ingl. k. connectivity management)
  - Andmete leitavus ja otsitavus (ingl. k. data discovery and search)
  - Ärisõnastik (ingl. k. business glossary)
  - Andmehaldurite ja -omanike haldus (ingl. k. stewardship & ownership management)
  - API tugi (ingl. k. API support)

Andmeintegratsiooni platvorm

- Andmeintegratsiooni platvorm (ingl. k. Data Integration Platform) on tarkvaralahendus, mis võimaldab koondada, sünkroniseerida ja teisendada erinevatest allikatest pärinevaid andmeid ning liigutada need kokkulepitud kujul vajalikesse sihtsüsteemidesse (nt andmelattu/andmejärve, rakendustesse)
    - Andmete teisendamistegevuste alla kuuluvad näiteks erineva päritoluga andmete struktuuri või formaadi ühtlustamine, uute väärtuste tuletamine/arvutamine, andmete rikastamine, väljade vastendamine ja muud sarnased sammud, mis toetavad andmete efektiivset ja suurimat lisandväärtust andvat kasutust äriprotsessides.
    - Sisuliselt on andmeintegratsiooni platvorm tööriist andmekonveierite (ingl. k. data pipelines) loomiseks ja haldamiseks, pakkudes vahendeid eri süsteemide või rakendustega ühendumiseks, töövoogude ehitamiseks ning andmetöötlusprotsesside sobivas järjekorras jooksutamiseks ja monitoorimiseks.
- Baasfunktsionaalsused
    - Ühenduvus (ingl. k. connectivity)
    - Andmete valmendus ja väljastus (ingl. k. data ingestion & delivery)
    - ETL/ELT-töövood (ingl. k. ETL/ELT flows)
    - Replikatsioon ja muudatuste püüdmine (ingl. k. data replication and change data capture (CDC))
    - Orkestreerimine (ingl. k. orchestration)
    - Monitooring ja logimine (ingl. k. monitoring and logging)
    - Andmeturve (ingl. k. data security)

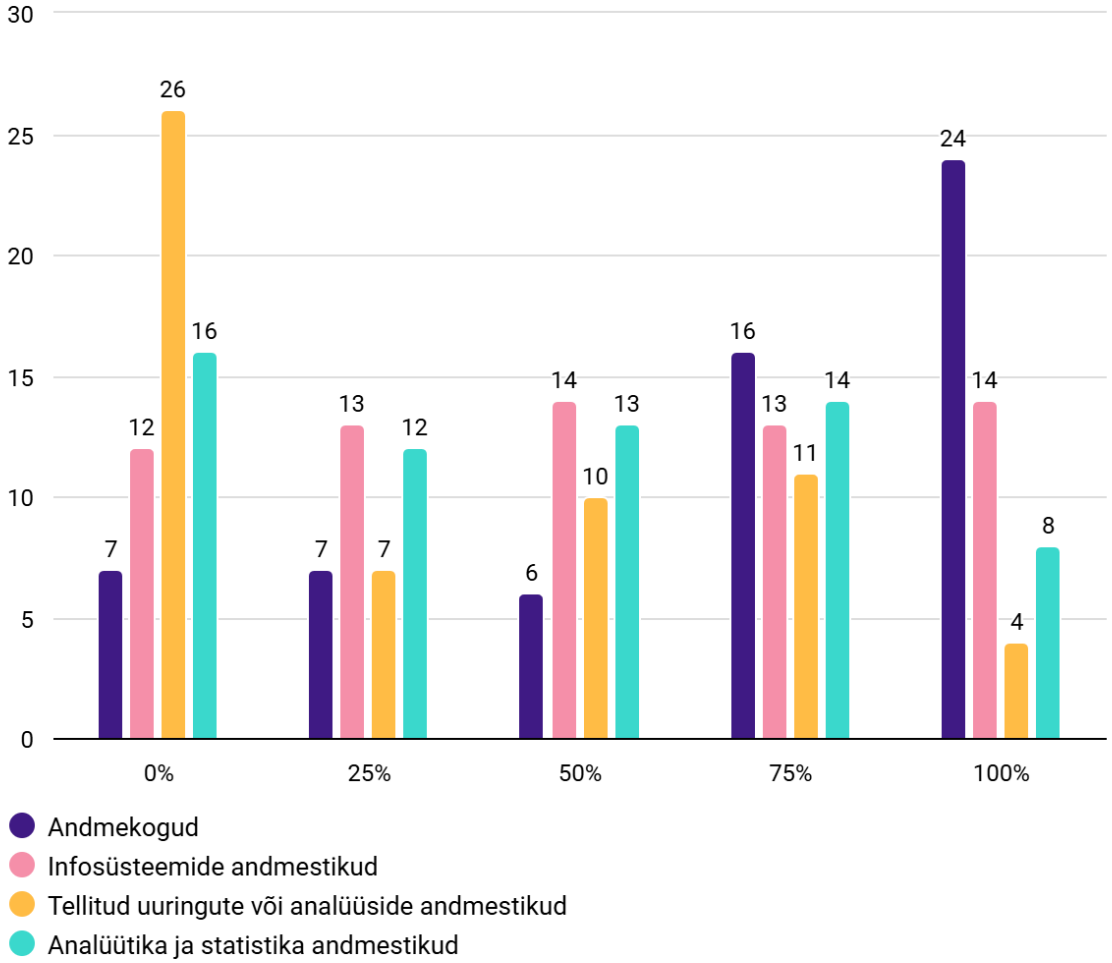
Meie on hetkeolukorra küsimustikus NELI andmestiku tüüpi!

Millises ulatuses on asutus kaardistanud oma andmestikud?



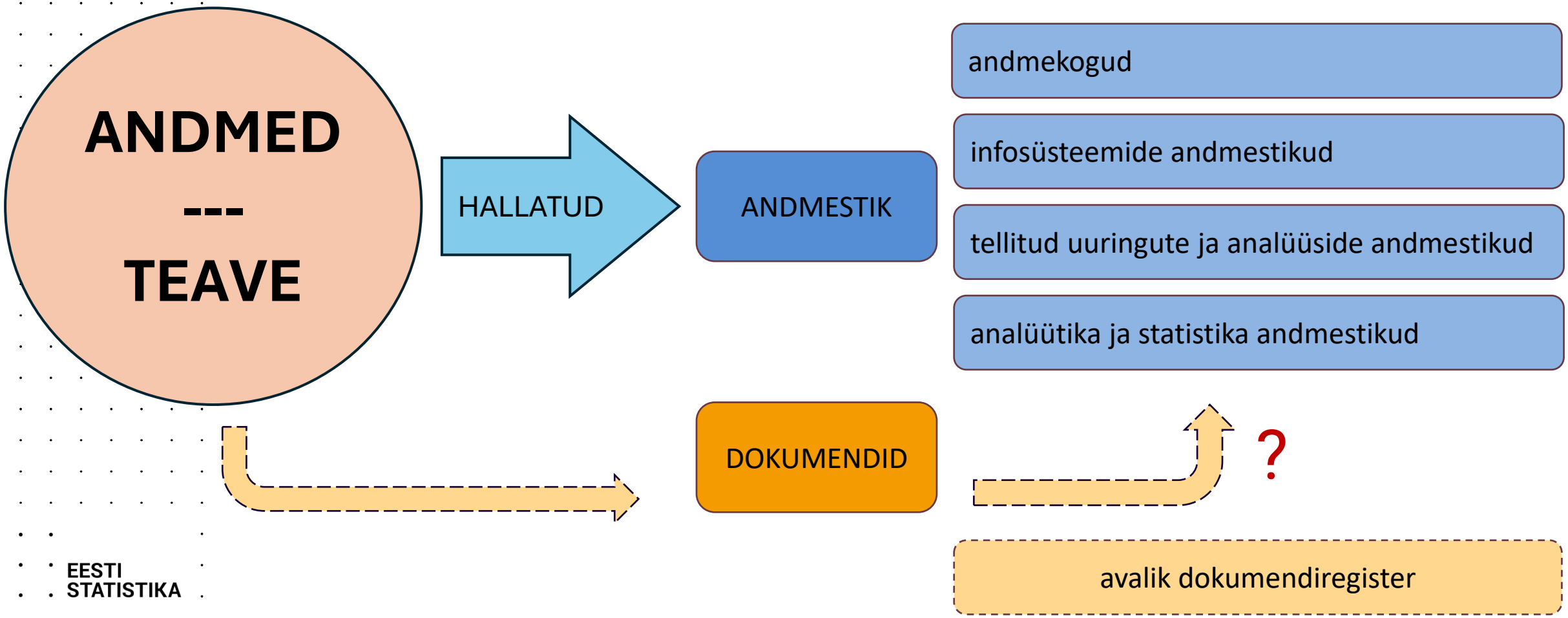
Allikas: statistikaamet

Millises ulatuses on andmestikele loodud kirjeldused?



Allikas: statistikaamet

Tervikülevaade andmetest. Seos teabe- ja dokumendihaldusega. Kas andmestike tüübid katavad terviku?



EESTI STATISTIKA

ARUTELU

Milliste töövahendite ja protsessidega on võimalik saada terviklik ülevaade andmetest?

Kuidas jõuda tervikülevaateni, kui alustame näiteks (tähtsamast) osast või andmekogudest?

EESTI STATISTIKA

Andmekvaliteedi kontseptsioon

Kas põhimõte *fit for purpose* on rakendatav?

Andmekvaliteet ja ärireeglid.

Andmekvaliteet andmetöötluses ja väljundis.

Loendite ja klassifikaatorite roll andmekvaliteedis.

EESTI
STATISTIKA

Mis on meie teada seis teie asutuses

Agregeeritud ülevaated faili või juhtimislauana (excel, tableau, sharepoint)
andmekvaliteediga tegeletakse iga infosüsteemi siseselt
Andmekvaliteeti kontrollitakse andmete sisestamisel
Andmelaos on loodud andmekvaliteedi aruanded
Andmelaos on loodud andmekvaliteedi aruanded
Andmelaos on loodud andmekvaliteedi aruanded
Andmelaos on loodud andmekvaliteedi aruanded
Ei ole kasutanud ühtegi eraldi andmekvaliteedi haldamise rakendust aga [kahel IS-il] on sisemised reeglid integreeritud. Süsteemset haldamist ei ole.
[see on STAT rida]
R, infosüsteemidesse arendatud funktsionaalsusi, eraldi scripte, Tableau. Juurutamisel Acurity andmekvaliteedi moodul.
SelectZereo, sql triggerid
SelectZero
SelectZero
SelectZero
SelectZero
SelectZero
Selectzero
SelectZero
Selectzero
SelectZero, Andmeait (SAP BO), Jira
SelectZero, Tableau
Wiki, asutuse infosüsteemi sisene kvaliteedi seiresüsteem (andmevalidaator)
#N/A
#N/A
#N/A

Andmekvaliteet õiguses – TKTA

§ 16[1]

lg 1. Riigi infosüsteemi kuuluva andmekogu vastutav töötleja dokumenteerib ja rakendab andmekvaliteedi seire ja haldamise protsessi, millega tagatakse riigi infosüsteemi kuuluvate andmekogude andmete kvaliteet vastavalt õigusaktidele.

- andmekvaliteedi seire
- andmete kvaliteet vastavalt õigusaktidele

Ptk 4 Teabehalduse korraldamine

§ 11 Asutuse teabehalduse korraldus peab tagama:

1) teabe kvaliteedi ja kättesaadavuse;

Kas teabe kvaliteet on sama kui andmekvaliteet või millest käib sisuliselt jutt?

EESTI STATISTIKA

Andmekvaliteet juhises (raamistikus)

- Eesmärk
  - Andmekvaliteedi eesmärk on tagada, et asutuse andmed on nende kasutusotstarbeks sobivad ning toetavad usaldusväärselt otsuste tegemist, teenuste osutamist ja teiste asutuse eesmärkide saavutamist.
  - Selleks peab asutus mõistma andmekvaliteedi olemust, andmete kvaliteedile esitatavaid nõudeid ning andmekvaliteedi seost asutuse üldise kvaliteedijuhtimisega.
- Taipamine
  - Organisatsioon peab pöörama tähelepanu andmekvaliteedile ning looma protsessid ja määrama vastutajad andmekvaliteedi eest.
  - Andmekvaliteet on sedavõrd oluline ja kompleksne osa andmehaldusest, et vajab hästi läbimõeldud lähenemist koos muude kvaliteedi teemadega ja tippjuhtide vastutust.

Kas põhimõte *fit for purpose* on rakendatav?

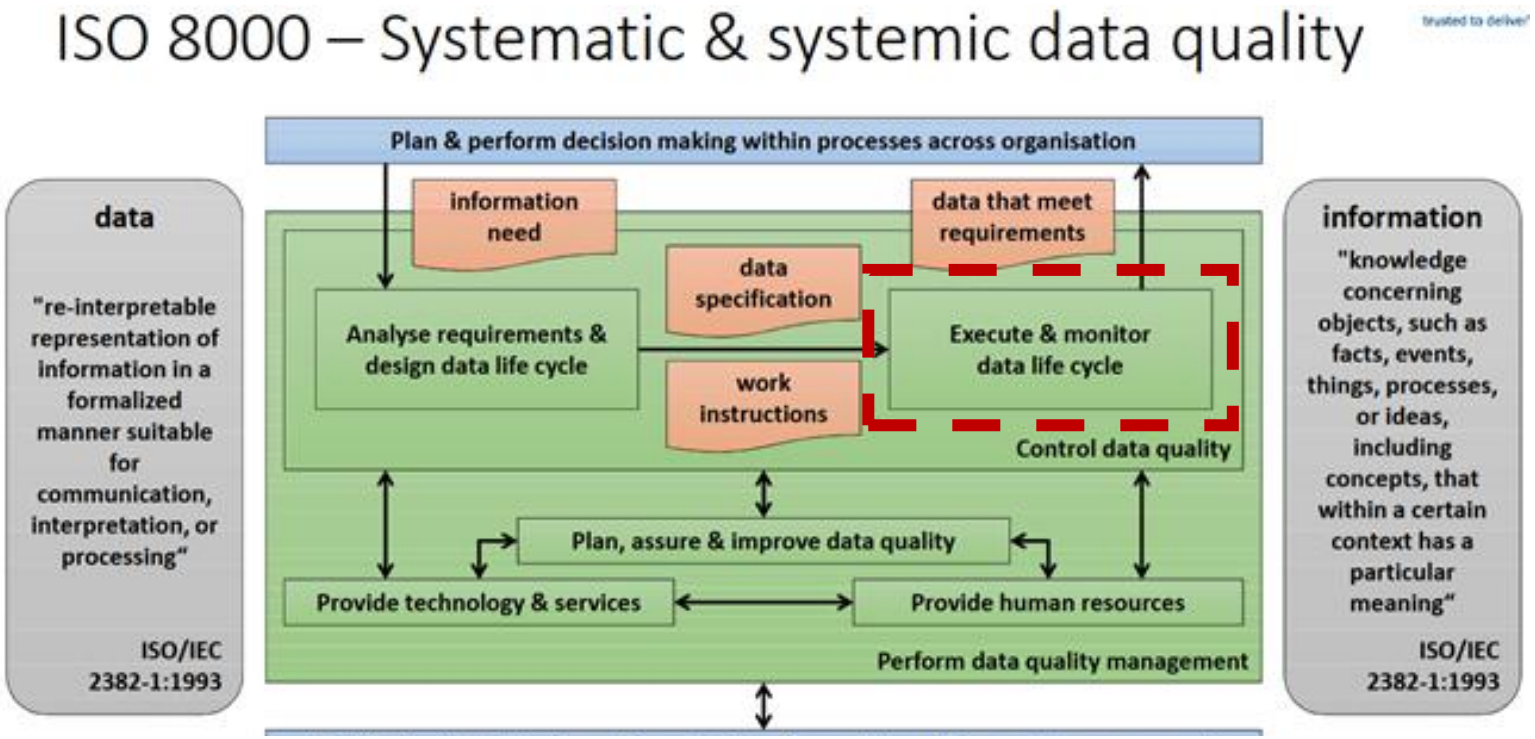
nõuded lähtuvalt kasutuseesmärgist

seotus protsessi ja elutsükliga

- DAMA käsitleb andmekvaliteeti andmehalduse põhivaldkonnana.
  - Selle järgi on andmed kvaliteetsed siis, kui need on sobivad oma kasutuseesmärgiks (***fit for purpose***).
  - Selle saavutamiseks tuleb määrata kvaliteedinõuded, mõõta andmete vastavust neile nõuetele ning juhtida pidevalt kvaliteeti mõjutavaid protsesse kogu andmete elutsükli jooksul.

- ISO 8000
  - ***Data that meets requirements***
  - ISO/TS 8000-1:2011, Data quality — Part 1: Overview
  - ISO 8000-2:2017, Data quality — Part 2: Vocabulary
  - ISO 8000-8:2015, Data quality — Part 8: Information and data quality: Concepts and measuring
  - ISO 8000-61:2016, Data quality — Part 61: Data quality management: Process reference model

ISO 8000 standardiperekond: Andmekvaliteet



3.3.1 quality – kvaliteet

Degree to which a set of inherent characteristics od an object fulfils requirements

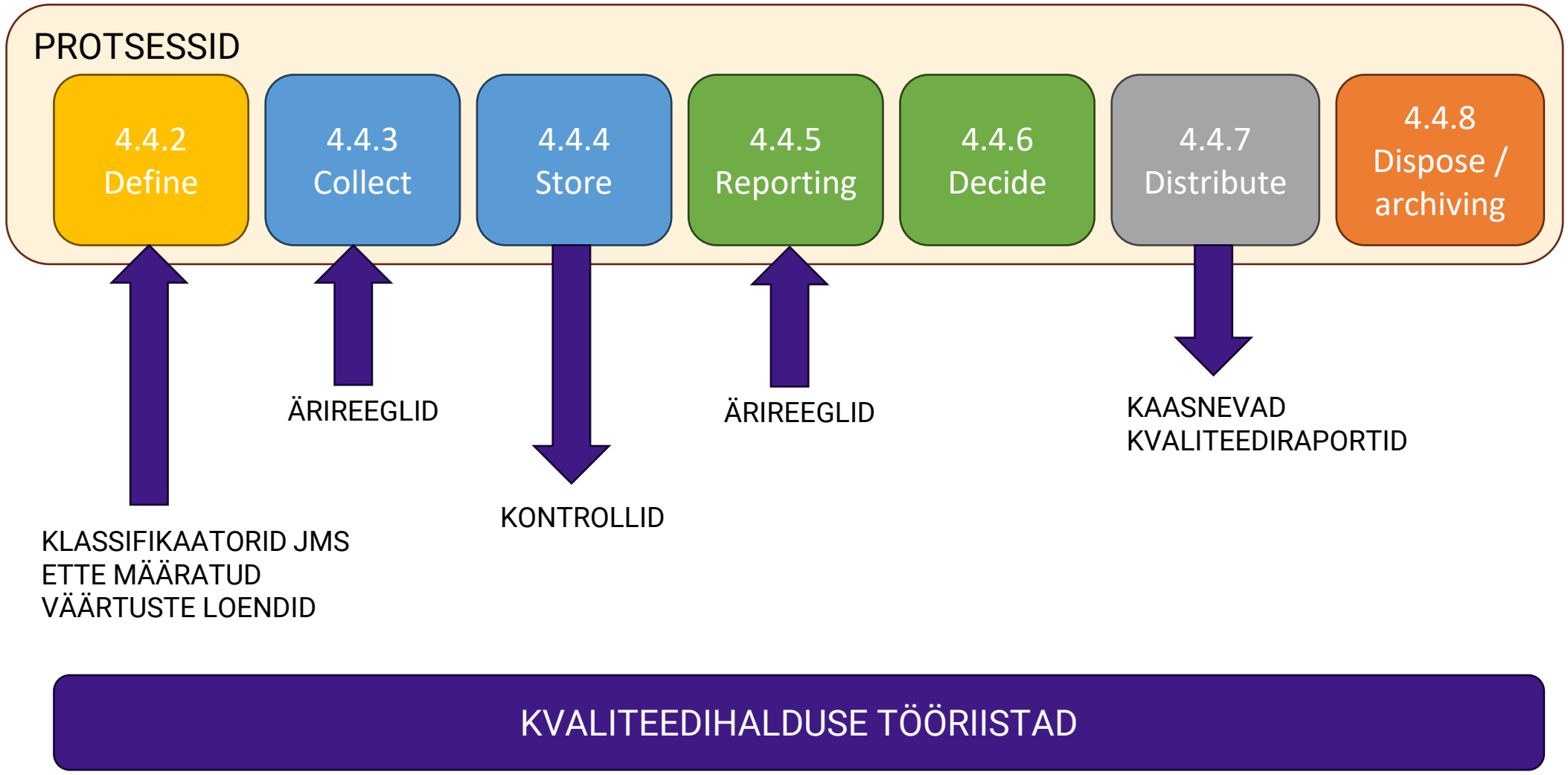
Määr, mille ulatuses objekti olemasolevad omadused vastavad nõudmistele

3.4.8 data quality – andmekvaliteet

Degree to which a set of inherent characteristics of data fulfils requirements

Määr, mille ulatuses ANDMETE olemasolevad omadused vastavad nõudmistele

Andmekvaliteet: mõõtmine ja parendamine



4.4.2
Define

Loendite ja klassifikaatorite roll andmekvaliteedis

SÕNASTIK
Thesauri | Vocabulary
---
sõnavara haldamise vahend, mis esitab mõistete hierarhilised või semantilised seosed ja vastavad terminid (sõnad, fraasid, tähised)

MÄRKSÕNASTIK
KONTROLLSÕNASTIK
Controlled Vocabulary
---
eelnevalt koostatud indeksterminite loend kirjelduste tegemiseks ja otsinguteks, mis on tavaliselt ühetasandilne

KOODILOEND
Code List
---
mõistete kui kvalitatiivsete muutujate ehk kategooriate loend, kus neile on lisatud koodid, mis võib olla hierarhiline

(STATISTILINE)
KLASSIFIKAATOR
Statistical Classification
---
mõistete (kategooriate) koodidega käsitlusala terviklikult (põhjalikult) hõlmav korrastatud ja täpselt kirjeldatud loend

ISO 25964 Thesauri and interoperability with other vocabularies

SKOS - Simple Knowledge Organization System
ConceptScheme > Concept

EESTI STATISTIKA

4.4.2 Define

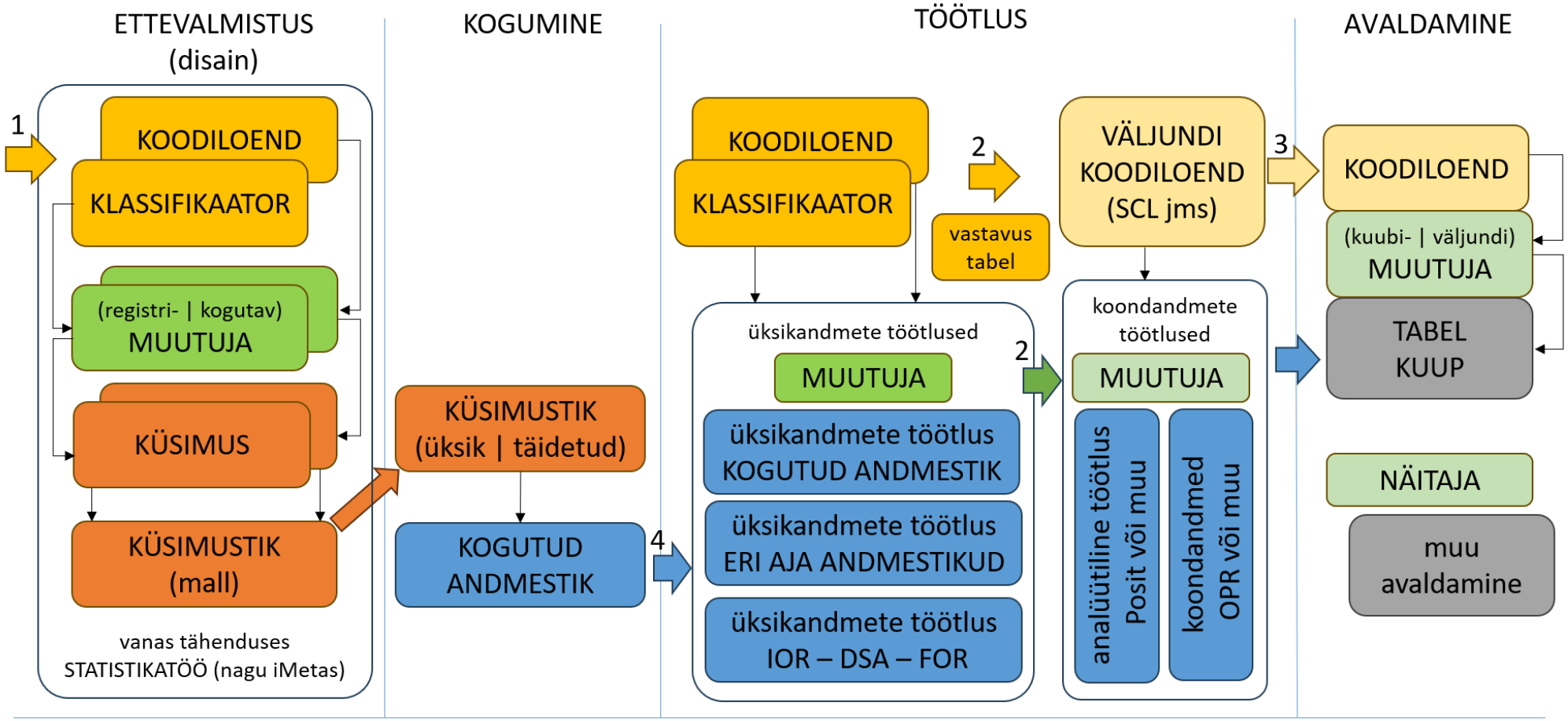
4.4.3 Collect

4.4.4 Store

4.4.5 Reporting

4.4.7 Distribute

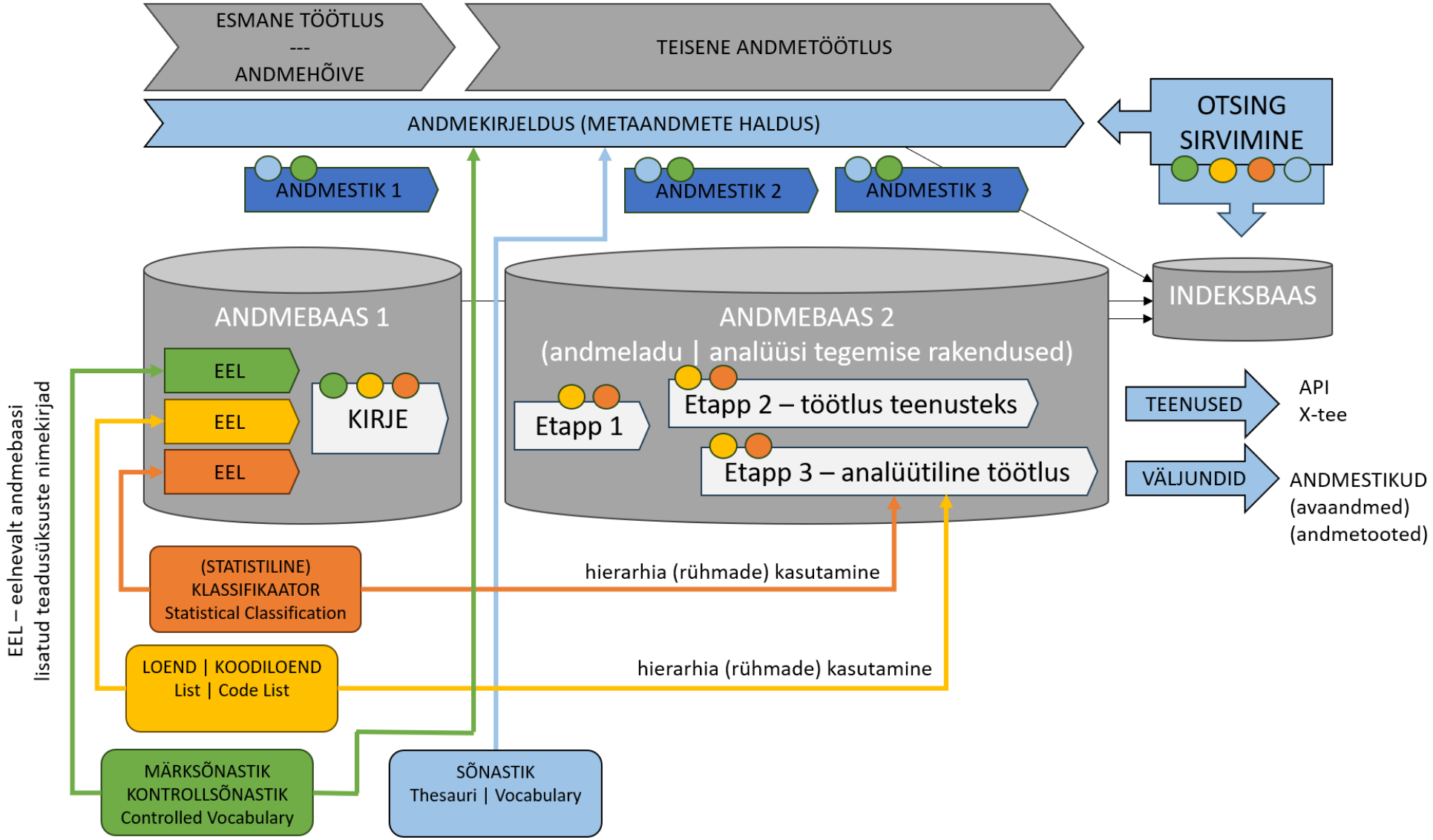
NÄIDE 1: Klassifikaatorite ja koodiloendite kasutamine statistika tootmise protsessi näitel



1. Varem koostatud või rahvusvahelise saadud klassifikaatori või koodiloendi kasutamine
2. Klassifikaatori või koodiloendi kasutamine üksikandmete töötlemisel: 2.1 hierarhiate kasutamine rühmitamisel, 2.2 vastavuste loomisel (iMeta mõttes registrimuutujast saab kuubimuutuja)
3. Konkreetse tabeli või kuubiga seotud koodiloend (sh standardkoodiloend (SCL))

EESTI STATISTIKA

NÄIDE 2: Üldisem kasutamine esmases ja teiseses andmetöötluses



EESTI STATISTIKA

NÄIDE 3

klassifikaatori tabel

CODE	PARENT	LEVEL	NAME_ET
800		1	Reisijateveo mootorsõiduk
8001	800	2	Mootorratas (üle 50 cm3)
8002	800	2	Mopeed
8003	800	2	Sõiduauto
80031	8003	3	Sõiduauto bensiiniga
80032	8003	3	Sõiduauto diislikütusega
80033	8003	3	Elektriauto
80034	8003	3	Sõiduauto muu kütusega
8004	800	2	Peamiselt reisijate vedamiseks mõeldud kaubik
8005	800	2	Takso
8006	800	2	Rendiauto
8007	800	2	Kiirabiauto
8008	800	2	Autoelamu
80081	8008	3	Haagissuvila
8009	800	2	Buss
80091	8009	3	Linnabuss
80092	8009	3	Kaugsõidubuss
80093	8009	3	Linnasõidu / kaugsõidu minibuss

klassifkaatori kasutamine kirjetes

variant 1

ID	CODE	tootmisaasta	tunnus b
1	80033	2020	abcd
2	80081	1999	efgh
3	80091	2004	ijkl

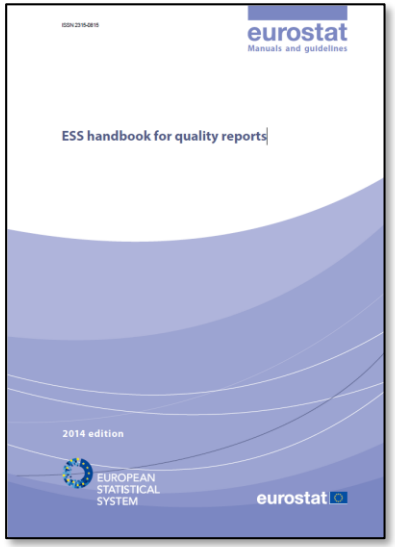
variant 2

ID	CODE	NAME_ET	tootmisaasta	tunnus b
1	80033	Elektriauto	2020	abcd
2	80081	Haagissuvila	1999	efgh
3	80091	Linnabuss	2004	ijkl

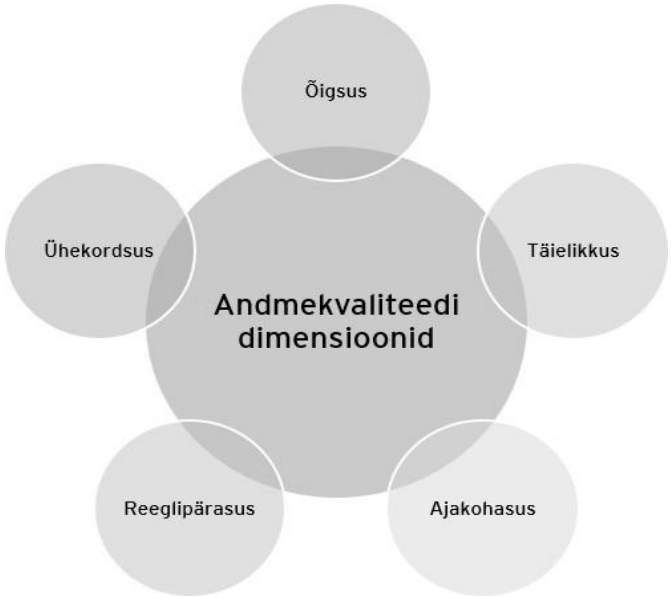
variant 3

ID	NAME_ET	tootmisaasta	tunnus b
1	Elektriauto	2020	abcd
2	Haagissuvila	1999	efgh
3	Linnabuss	2004	ijkl

Andmekvaliteedi mudelid üldisem vaade



Statistika andmekvaliteet



Andmekvaliteedi mudel

Ärireegli tüüp

Ühtlustatud mõisted
Väärtuste olemasolu ja andmete täielikkus
Õige formaat
Väärtuste domeeni koosseisu kuulumine
Õige vahemik
Väärtuste domeeni ühilduvus
Järjepidevus
Õigsuse kontroll
Ühekordsuse kontroll
Ajakohasuse kontroll
Agregeerimine – kirjete arvu loogilisus
Agregeerimine – muudatuste loogilisus
Agregeerimine – muudatuste loogilisus määratud ajavahemikul

Ärireeglid

EESTI STATISTIKA

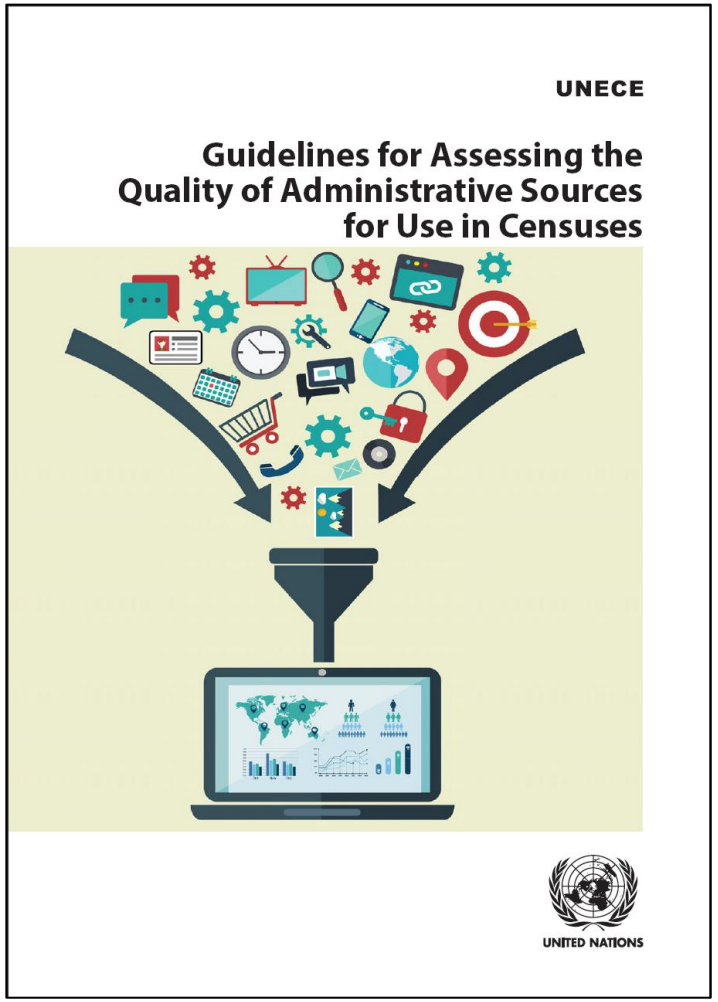
Olukord kvaliteediga (riiklikus statistikas), rõhuga andmekogude (admin.allikate) kasutamisel

Praeguse olukorra värske rahvusvaheline kaardistus, koostatud 2017 kuni 2021.

Esitab kvaliteedihindamise staatuste mudeli.

(see lähtub arusaamast, et andmete elukäigus on eristatavad ja fikseeritavad etapid (*stages*) ja igas etapis saab rakendada kvaliteedimõõtmist; kasutusel on Euroopa Statistikasüsteemi (ESS) kvaliteedimudel)

STAGE 1: SOURCE – A metadata-based QA of new or re-supplied administrative sources to be used in the census.	**STAGE 2: DATA** – QA of the raw administrative data supplied to NSOs by administrative authorities.
STAGE 3: PROCESS – The processes carried out on administrative data sources, in censuses, to transform the data for census use and/or to improve quality.	**STAGE 4: OUTPUTS** – The overall QA of the census outputs produced using administrative data.



Etapid

1. **Allika etapp (Source Stage)**

   A metadata-based QA of new or re-supplied administrative sources to be used in the census. This Stage does not require NSOs to be in possession of the actual data, but it is crucial for the Stages that follow

2. **Andmehõive etapp (Data Stage)**

   The QA of the raw administrative data supplied to NSOs by administrative authorities (data suppliers). This will require NSOs to validate the data supplied against the lessons learned from the Source Stage. As well as basic validation, this Stage includes any processing required to establish the quality of the data supplied vis-à-vis what was expected and comparisons with alternative sources,

3. **Andmetöötluse etapp (Process Stage)**

   The processes often carried out on administrative data sources, in censuses, to transform the data for census use and/or to improve quality.

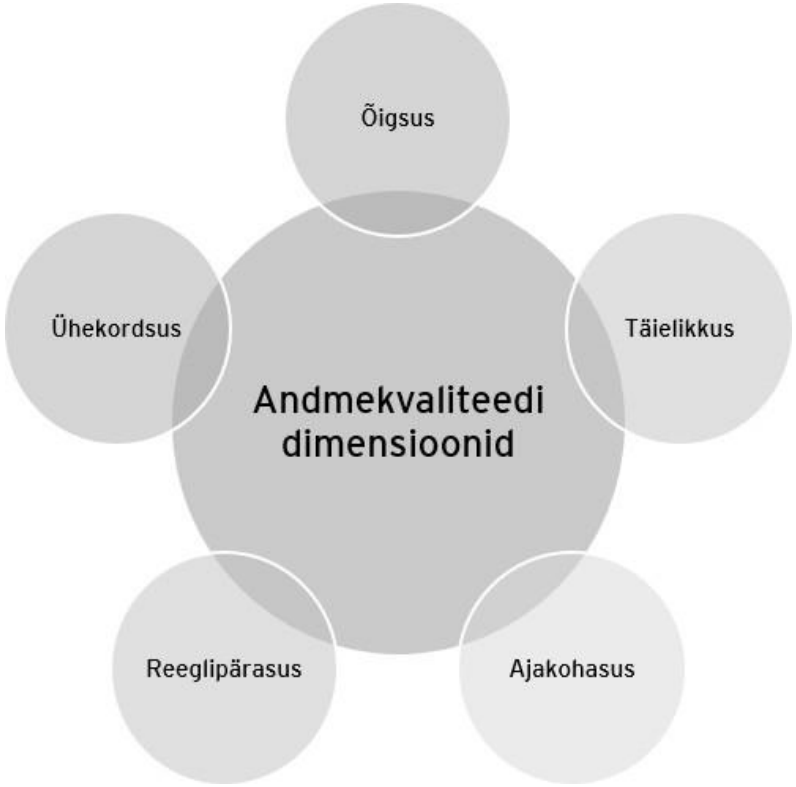
   - The processes identified include: (i) Record linkage, (ii) Statistical registers and the 'signs-of-life' methodology, (iii) Enumeration of population units: administrative data-based models, (iv) Conflict resolution/decision between sources, and (v) Editing and imputation.

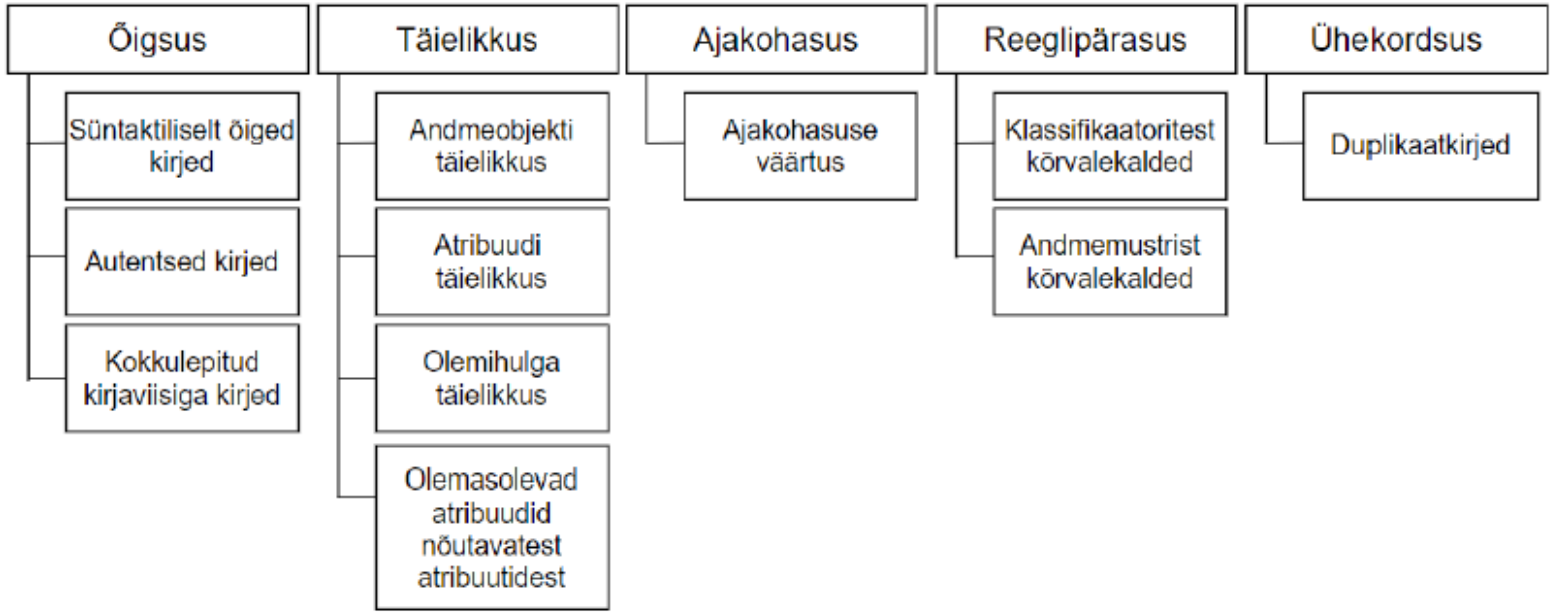
4. **Väljundi / levitamise / publitseerimise etapp (Output Stage)**

   - The overall QA of the census outputs produced using administrative data. While this is not conceptually that different from the assessment of the outputs of a traditional census, these Guidelines attempt to identify where this may differ.

EESTI STATISTIKA

Andmekvaliteedi mudel: dimensioonid ja indikaatorid





Joonis 3. Dimensioonide jaotus indikaatoriteks

EESTI STATISTIKA

Dimensioonid

- **Õigsus** (*Accuracy*) näitab, mil määral vastavad andmed tegelikkusele. Andmete õigsus jaguneb süntaktiliseks ja semantiliseks õigsuseks.
- **Täielikkus** (*Completeness*) näitab, mil määral on olemas kõik nõutud andmed. Täielikkus on vaadeldav kahes osas: kirjete täielikkus ja kogumi ehk populatsiooni täielikkus. Kirjete täielikkus näitab, mil määral on andmekirje kõigil tunnustel olemas väärtused.
- **Ajakohasus** (*Timeliness*) näitab, mil määral andmete värskus ja kättesaadavus vastab vajadustele ja nõuetele. Aja jooksul andmed muutuvad ning viide reaalsete sündmuste ning nende andmetes fikseerimise või andmete värskendamise vahel on vältimatu.
- **Reeglipärasus** (*Orderliness*) näitab, mil määral andmete formaat ja struktuur vastab nõuetele. Esiteks tähendab reeglipärasus kokkulepitud klassifikaatorite kasutamist. Teiseks tähendab reeglipärasus kokkulepitud andmemustrite järgimist.
- **Ühekordsus** (*Uniqueness*) näitab, mil määral esineb andmetes duplikaatkirjeid. Ühekordsuse probleem tekib juhul, kui ühe reaalsest elust pärineva objekti kohta on andmetes talletatud kaks või enam kirjet.

dimensioonid ja indikaatorid

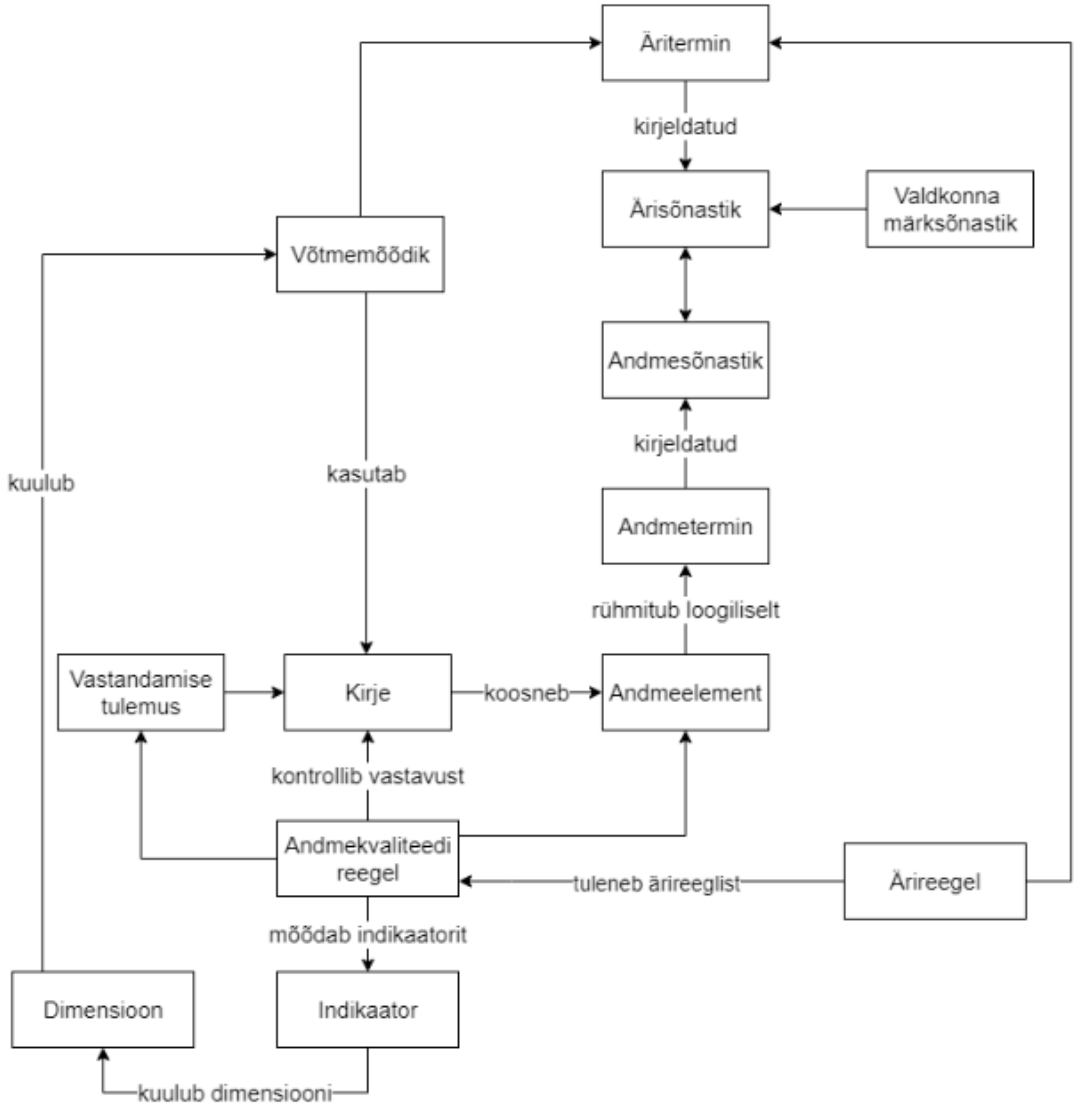
- Õigsus
  - süntaktiliselt õigete kirjete määr;
  - autentsete kirjete määr;
  - kokkulepitud kirjaviisi järgivate kirjete määr.
- Täielikkus
  - andmeobjekti täielikkuse määr;
  - atribuudi täielikkuse määr;
  - olemihulga täielikkuse määr;
  - olemasolevate atribuutide määr nõutavatest atribuutidest.
- Ajakohasus
  - ajakohasuse määr skaalal 0-1 vastavalt ajakohasuse definitsioonile $\max\left\{0, 1 - \frac{Värskus}{Volatiilsus}\right\}$
- Reeglipärasus
  - kokkulepitud klassifikaatorite kasutamisest kõrvalekallete määr;
  - kokkulepitud andmemustritest kõrvalekallete määr.
- Ühekordsus
  - duplikaatkirjete määr.

ÜLESANNE

Mõelge 1-2 andmestiku, millele andmekvaliteeti mõõta ja kirjutage iga indikaatori kohta:
1. kuidas rakendaksite, kas oskate, on rakendatav
2. mis oleks määr, mis oleks veel rahuldav

- Õigsus
  - süntaktiliselt õigete kirjete määr;
  - autentsete kirjete määr;
  - kokkulepitud kirjaviisi järgivate kirjete määr.
- Täielikkus
  - andmeobjekti täielikkuse määr;
  - atribuudi täielikkuse määr;
  - olemihulga täielikkuse määr;
  - olemasolevate atribuutide määr nõutavatest atribuutidest.
- Ajakohasus
  - ajakohasuse määr skaalal 0-1 vastavalt ajakohasuse definitsioonile
- Reeglipärasus
  - kokkulepitud klassifikaatorite kasutamisest kõrvalekallete määr;
  - kokkulepitud andmemustritest kõrvalekallete määr.
- Ühekordsus
  - duplikaatkirjete määr.

Andmekvaliteedi mõõtmise metamudel



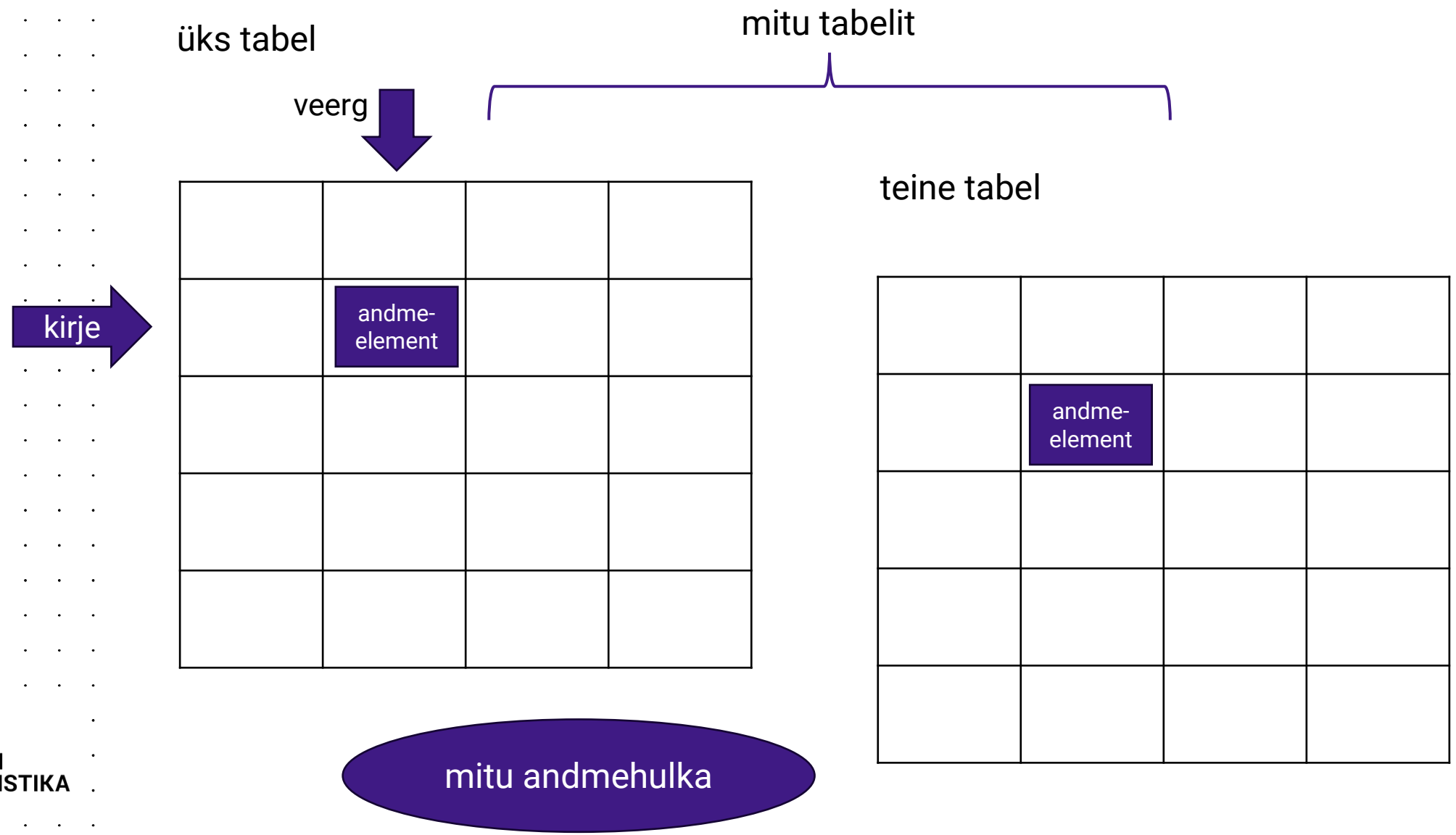
Joonis 4: Andmekvaliteedi mõõtmise metamudel

ärireeglite määramine

Mille kaupa mõõdame?

- **Andmeelement**: Andmekvaliteedi probleemid, mis ilmnevad konkreetse andmeelemendi tasemel ehk kui vaadeldakse korraga konkreetset väärtust.
- **Veerg**: Andmekvaliteedi probleemid, mis ilmnevad veeru tasemel ehk kui vaadeldakse veeru väärtusi üle kõikide kirjete.
- **Kirje**: Andmekvaliteedi probleemid, mis ilmnevad kirje tasemel ehk kui vaadeldakse korraga kirje ehk ühe tabeli rea kõiki väärtusi.
- **Üks tabel**: Andmekvaliteedi probleemid, mis ilmnevad ühe tabeli tasemel ehk kui vaadeldakse korraga ainult ühes tabelis olevaid andmeid.
- **Mitu tabelit**: Andmekvaliteedi probleemid, mis ilmnevad siis, kui omavahel on seotud mitu tabelit ehk kui vaadeldakse korraga mitmes tabelis olevaid andmeid.
- **Mitu andmehulka**: Andmekvaliteedi probleemid, mis ilmnevad siis, kui omavahel on seotud mitu andmehulka ehk kui vaadeldakse korraga mitmes andmehulgas olevaid andmeid. Näiteks Rahvastikuregistris ja Ehitusregistris hoitavad isikuandmed.

Sama joonisena



EESTI
STATISTIKA

Andmekvaliteedi probleem	Detailsuse tase						Dimensioonid				
	Andme-element	Veerg	Kirje	Üks tabel	Mitu tabelit	Mitu andmehulka	Õigsus	Täielikkus	Ajakohasus	Reegli-pärasus	Ühekordsus
Puuduv väärtus	√							√			
Süntaksi rikkumine	√						√				
Vale väärtus	√						√		√		
Väärtusvahemiku rikkumine	√						√				
Sobimatu alamstring	√						√				
Õigekirjaviga	√						√				
Ebatäpne väärtus	√						√				
Valdkonnakitsenduse rikkumine	√	√	√	√	√	√	√			√	
Unikaalse väärtuse rikkumine		√									√
Sünonüümide kasutus		√				√				√	
Pooltühi kirje			√					√			
Funktsionaalse sõltuvuse rikkumine			√				√			√	
Ligikaudselt dubleerivad kirjed				√		√				√	√
Vastuoluliselt dubleerivad kirjed				√		√	√				√
Viiteterviklus					√			√			
Vale viide					√		√		√		
Süntaksite mitmekesisus					√	√				√	
Ringlus kirjete seas					√			√			
Mõõtmisühikute mitmekesisus						√				√	
Esitluse mitmekesisus						√				√	
Homonüümide kasutus						√					√

Andmekvaliteedi reeglite kirjeldamine – andmete profileerimine

- **Struktuuripõhise profileerimise** käigus analüüsitakse andmete järjepidevust ja formaadilist korrektsust. Lisaks teostatakse matemaatilisi kontrolle (näiteks summa leidmine, miinimumväärtuste leidmine ja maksimumväärtuste leidmine). Struktuuripõhine profileerimine aitab tuvastada, kui hästi on andmed struktureeritud. Näiteks kui palju on vale pikkusega telefoninumbreid.

- **Sisupõhise profileerimise** käigus analüüsitakse konkreetseid andmekirjeid, mille tulemusena on võimalik tuvastada andmekirjetes esinevaid süstemaatilisi probleeme. Näiteks ilma suunakoodita telefoninumbrite esinemist.

- **Seostepõhise profileerimise** käigus tuvastatakse andmete omavahelised seosed, näiteks andmetabelite vahelised seosed või arvutustabelis (näiteks MS Exceli failis) hoitavate tabelite või väljade seosed.

ärireeglite koostamine

Ärireegel 1. Kõikide klientide, kelle krediidilimiit on alla 1000, laenutaotlused lükatakse automaatselt tagasi.
Selgitus: see reegel juhib laenu kinnitamise otsustusprotsessi konkreetse kriteeriumi (krediidilimiidi) alusel.
Termin: klient; laenutaotlus; krediidilimiit
Kvaliteedireegel: registris ei tohi olla kandeid isikute kohta, kelle krediidilimiit on väiksem kui 1000.

Ärireegel 2: Töötamise lõpetamine registreeritakse 10 p jooksul või töötamise lõpetamise päeval (https://www.riigiteataja.ee/akt/106072023107).
Selgitus: lõpetada töötamine, kui töötasu (TSD) väljamakset ei ole toimunud 6 kuu jooksul.
Termin: töötamise lõpetamine
Kvaliteedireegel:
- Registris ei tohi olla reaalse töötamise kandeid.
- Registris ei tohi olla töötasu väljamakseteta töötamise kandeid.

DEMOD

1. Andmekvaliteedi jälgimine / mõõtimine R-iga

2. SelectZero’ga andmekvaliteedi haldus

EESTI
STATISTIKA

ARUTELU

Kuidas saavutada andmekvaliteedi kindlakstegemine?

Seos protsessi / elukäiguga

Kuidas parendada?

EESTI
STATISTIKA

KOKKUVÕTE

Aitäh kõigile!

EESTI
STATISTIKA