

# Ülevaade riigi andmehalduse tegevuskavast 2019-2021

Veiko Berendsen, andmehalduse ekspert

12.06.2019

Ülevaade Statistikanõukogule



Andmesoo kuivendamine

# Ülevaade riigi andmehalduse tegevuskavast 2019-2021



## TEGEVUSKAVA

1. Üleriigiline ülevaade andmestikest asutuste kaupa

2. Metaandmed / andmekirjeldus

3. Andmekvaliteet

4. Andmete elukäigu haldus

5. Andmete avaldamine / avalikustamine (MKM)

6. Andmehaldur (*data steward*)

## ARENGUKAVA

### Meede 3. Riigi andmehalduse juhtimine

1. Luua andmehalduse juhtrühm ja edendada andmehalduse eest vastutust ja tegevusi asutustes.

2. Koordineerida riigi infosüsteemi andmekogude jt statistikas kasutatavate infosüsteemide andmekirjeldusi, edendada semantilist koostalitlusvõimet ning juhtida klassifikaatorite süsteemi.

3. Arendada välja riiklik metainfosüsteem.

4. Luua riiklik andmete arhitektuurilise ülevaate keskkond.

5. Arendada välja andmekvaliteedi kindlaks tegemise ja parendamise raamistik.

6. Mõõta halduskoormust ja rakendada kooskõlastamise kaudu andmekogudes ühekordse põhiandmete kogumise põhimõte (*once-only* printsiip).

7. Luua valmisolek riikliku andmearhiivi loomiseks.



|                                                                                                                                                                                               | 2019 |    | 2020 |    | 2021 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|----|------|----|------|----|
|                                                                                                                                                                                               | I    | II | I    | II | I    | II |
| 1. Andmekirjeldus ja metaandmed                                                                                                                                                               |      |    |      |    |      |    |
| 1.1 Andmekirjelduse standardi koostamine                                                                                                                                                      |      |    |      |    |      |    |
| 1.2 Semantilist koostalitusvõimet edendavate valdkondlike sõnastike koostamine                                                                                                                |      |    |      |    |      |    |
| 1.3 Ülevaate omamine rahvusvahelistest valdkondlikest andmehaldusega seotud standarditest. Standardite ülevõtmine, tõlkimine ja juurutamine sh andmehaldusega seotud mõisted ja terminoloogia |      |    |      |    |      |    |
| 1.4 Klassifikaatorite ühtlustatud kirjelduste juurutamine ja klassifikaatorite haldussüsteemi uuendamine                                                                                      |      |    |      |    |      |    |
| 2. Andmestike ülevaade ja põhiandmete määratlemine                                                                                                                                            |      |    |      |    |      |    |
| 2.1 Üleriigilise andmehalduse raamistiku rakendamine. Infoarhitektuuriliste valdkondlike mudelite rakendamine                                                                                 |      |    |      |    |      |    |
| 2.2 Andmekataloogi koostamise ja haldamise juhised                                                                                                                                            |      |    |      |    |      |    |
| 2.3 Funktsionaalsusnõuded andmestiku haldamiseks                                                                                                                                              |      |    |      |    |      |    |
| 2.4 Riigi infosüsteemis eri tüüpi andmestike haldamiseks infoarhitektuurilise lahenduse ja rakenduse loomine (Riigi infosüsteemi haldussüsteemi RIHA uuendamine)                              |      |    |      |    |      |    |
| 2.5 Riigi põhiandmete määratlemine                                                                                                                                                            |      |    |      |    |      |    |
| 2.6 Pilotprojektid – ülevaadete loomine andmestikest valitud asutustes (3-4 asutust või funktsiooni)                                                                                          |      |    |      |    |      |    |
| 3. Andmekvaliteet                                                                                                                                                                             |      |    |      |    |      |    |
| 3.1 Andmekogude / infosüsteemide kvaliteedi tagamise juhise uuendatud versiooni koostamine                                                                                                    |      |    |      |    |      |    |
| 3.2 Andmeanalüütikas sh statistikas vajaliku andmekvaliteedi tuvastamise ja kirjeldamise meetoodika ja juhise loomine                                                                         |      |    |      |    |      |    |
| 4. Andmete avaldamine / avaandmed                                                                                                                                                             |      |    |      |    |      |    |
| 4.1 Andmete usaldusväärsuse tagamise ja selle tuvastamise meetoodika koostamine                                                                                                               |      |    |      |    |      |    |
| 4.2 Anonümiseerimise ja pseudonümiseerimise ning teiste otsest tuvastamist mittevõimaldavale kujule viimise (otmimise) meetodite arendamine ja nende juurutamine                              |      |    |      |    |      |    |
| 4.3 Andmetele nende avaldamisel lisatavate litsentside kokku leppimine ja unikaalsete identifikaatorite lisamine ning juhend nende kasutamiseks                                               |      |    |      |    |      |    |
| 5. Andmete elukäiguhaldus                                                                                                                                                                     |      |    |      |    |      |    |
| 5.1 Protsessi kui elukäigu haldus: funktsionaalsusnõuded data lineage'i juurutamiseks                                                                                                         |      |    |      |    |      |    |
| 5.2 Andmete väärtuse hindamine meetoodika ja juhise väljatöötamine                                                                                                                            |      |    |      |    |      |    |
| 5.3 Asutuses andmestike repositooriumite pidamise nõuete või soovitude väljatöötamine                                                                                                         |      |    |      |    |      |    |
| 5.4 Avaldatud andmestike haldamise ja säilitamise repositooriumi (andmearhiivi) loomise Statistikaametis                                                                                      |      |    |      |    |      |    |
| 6. Andmehaldur ja täiendkoolitus                                                                                                                                                              |      |    |      |    |      |    |
| 6.1 Andmehalduse ülesannete täitmiseks vajalike oskuste omandamiseks andmehalduri töökirjelduse ja koolitusprogrammi koostamine. Vajalike koolitusmaterjalide koostamine                      |      |    |      |    |      |    |
| 6.2 Täiendkoolituse läbiviimine                                                                                                                                                               |      |    |      |    |      |    |



Siin ta on see Andmejärv!

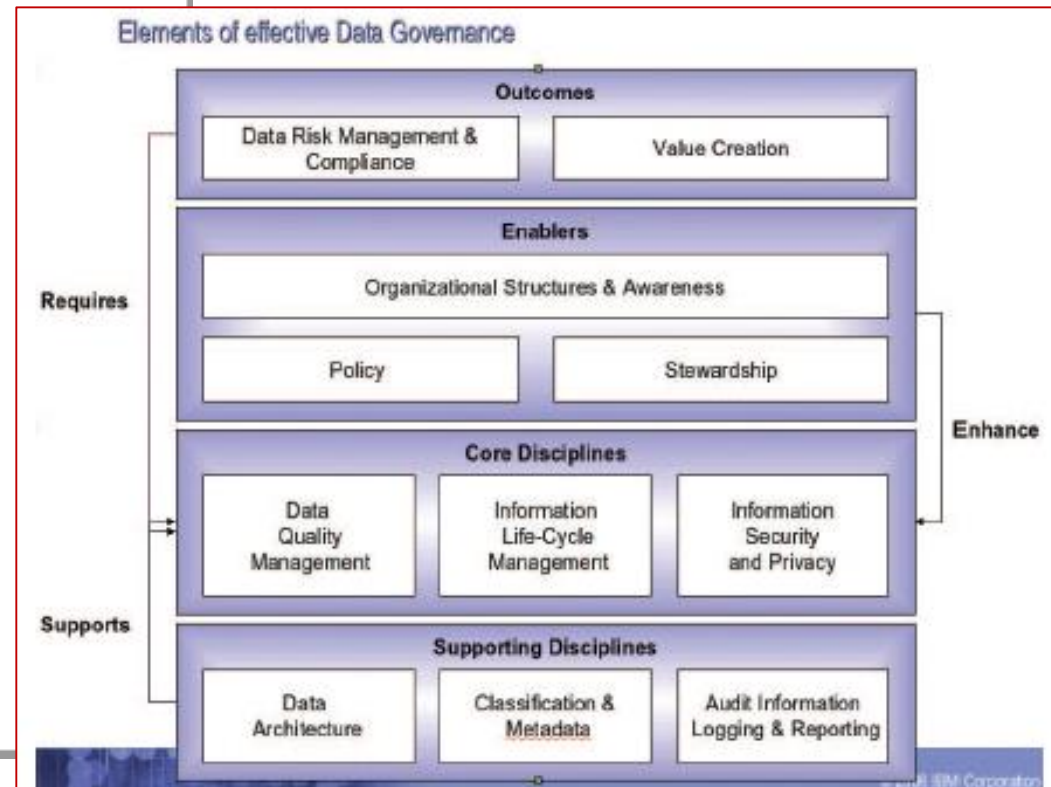
# Data Governance Council Maturity Model

IBM Data Governance Council Maturity Model  
October 2007



The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance

**O**utcomes – väljundid  
**E**enablers – võimaldajad  
**C**ore Disciplines – põhivaldkonnad  
**S**upporting Disciplines – tugivaldkonnad



| Category |                                        | Description                                                                                                                                                                              |
|----------|----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1        | Organizational Structures & Awareness  | Describes the level of mutual responsibility between business and IT, and recognition of the fiduciary responsibility to govern data at different levels of management.                  |
| 2        | Stewardship                            | Stewardship is a quality control discipline designed to ensure custodial care of data for asset enhancement, risk mitigation, and organizational control.                                |
| 3        | Policy                                 | Policy is the written articulation of desired organizational behavior.                                                                                                                   |
| 4        | Value Creation                         | The process by which data assets are qualified and quantified to enable the business to maximize the value created by data assets.                                                       |
| 5        | Data Risk Management & Compliance      | The methodology by which risks are identified, qualified, quantified, avoided, accepted, mitigated, or transferred out.                                                                  |
| 6        | Information Security & Privacy         | Describes the policies, practices and controls used by an organization to mitigate risk and protect data assets.                                                                         |
| 7        | Data Architecture                      | The architectural design of structured and unstructured data systems and applications that enable data availability and distribution to appropriate users.                               |
| 8        | Data Quality Management                | Methods to measure, improve, and certify the quality and integrity of production, test, and archival data.                                                                               |
| 9        | Classification & Metadata              | The methods and tools used to create common semantic definitions for business and IT terms, data models, types, and repositories. Metadata that bridge human and computer understanding. |
| 10       | Information Lifecycle Management       | A systematic policy-based approach to information collection, use, retention, and deletion.                                                                                              |
| 11       | Audit Information, Logging & Reporting | The organizational processes for monitoring and measuring the data value, risks, and efficacy of governance.                                                                             |

# Andmehaldus / Data Governance

## Väljundid [O]

Väärtuse loomine

## Võimaldajad [E]

Organisatsioon ja korraldus

## Põhivaldkonnad [C]

Andmekvaliteet

Elukäiguhaldus  
ja andmearhiiv

Andmeturve ja  
avalikustamine

## Tugivaldkonnad [S]

Andmearhitektuur

Metaandmed

## Organisatsioon ja korraldus [E]

- TEGEVUS: koordineerimine ja juhtimine (klassifikaatorid ja kirjeldused)

## Väärtuse loomine [O]

- TEGEVUS: (seos andmeteenusega) andmete olemasolu, vajaduse, kasutatavuse (s.o väärtuse) väljaselgitamine ja statistilise väljundiga lisandväärtuse pakkumine

## Andmearhitektuur ja metaandmed [S]

- TEGEVUS: mudelite ja standardite juurutamine, semantika edendamine

## Andmekaitse ja teabe/andmete avalikustamine [C]

- TEGEVUS: isikuandmekaitse, üksikandmete kasutamine, avaandmete pakkumine

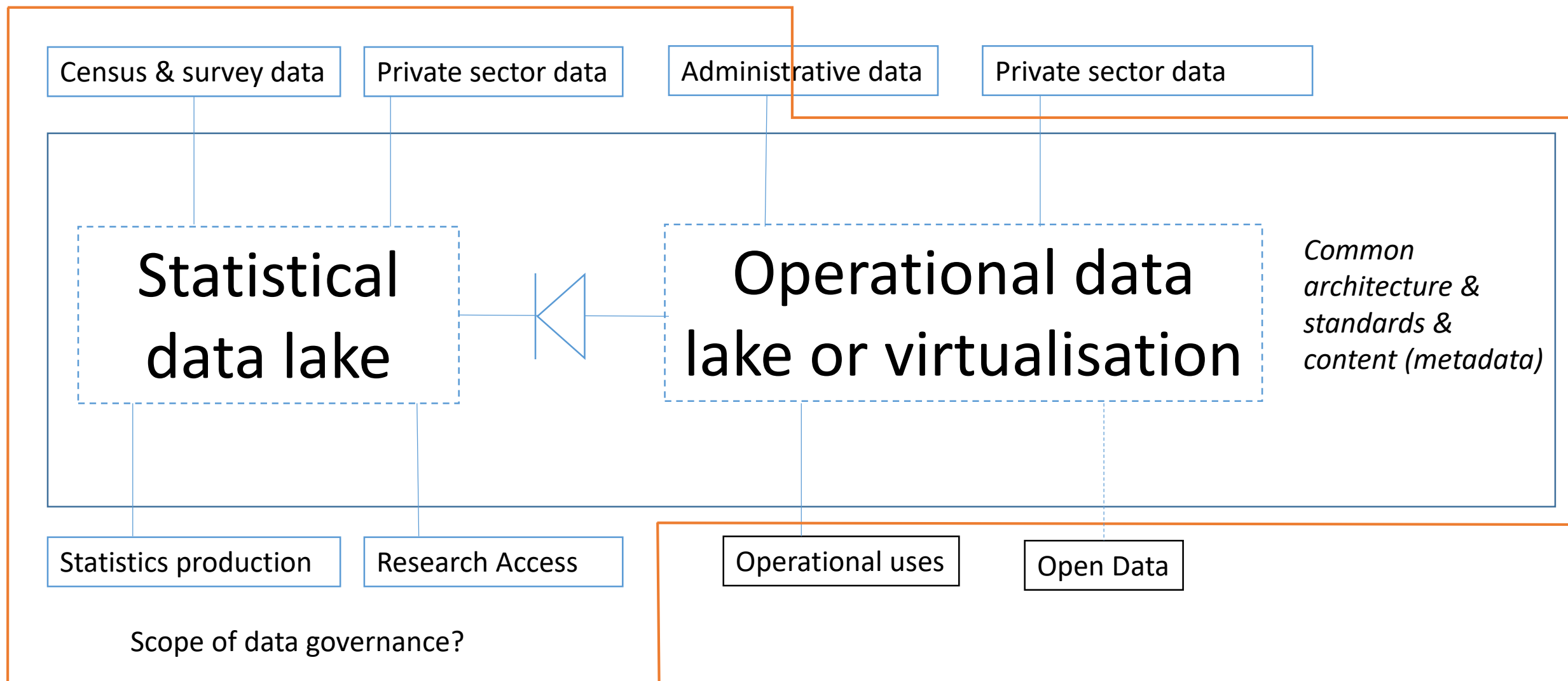
## Andmekvaliteet [C]

- TEGEVUS: andmekvaliteedi edendamine riigis ja selle jälgimine (audit)

## Elukäiguhaldus ja andmearhiiv [C]

- TEGEVUS: andmete elukäiguhalduse juurutamine; andmearhiivi loomine

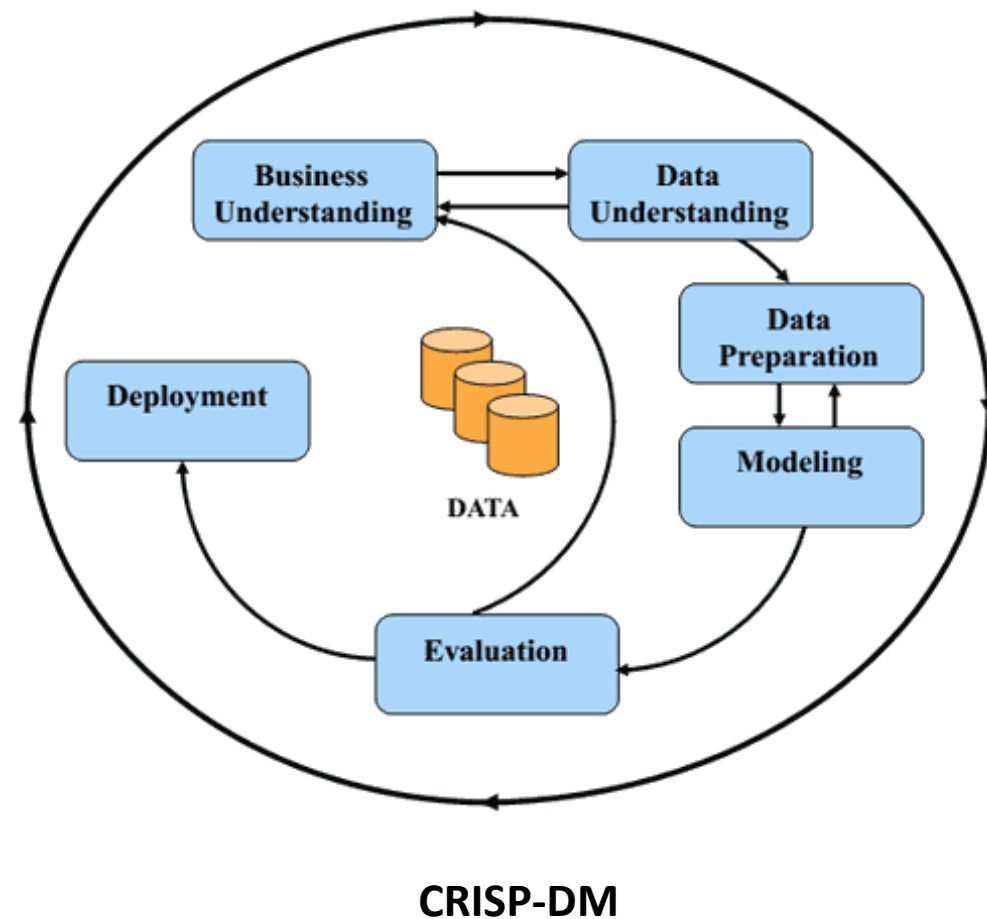
# Rahvusvaheline statistikaasutuste juhtide vaade andmehalduse ulatusele





Andmehaldus ja andmeteadus > CRISP-DM

- Andmeteaduse elutsükli ülevaade ja selle komponentide baaskursused vastavalt CRISP-DM klassifikatsioonile
  - \*ärivaldkonna vajaduste teadmine ja hindamine;
  - \*andmete ettevalmistus, puhastamine, andmelaod;
  - \*masinõpe ja hindamine, lahenduste juurutamine, visualiseerimine ja tõlgendamine; ning
  - \*privaatsususe ja legaalsed aspektid.



CRISP - cross-industry standard process for data mining



# Pearuga kahasse

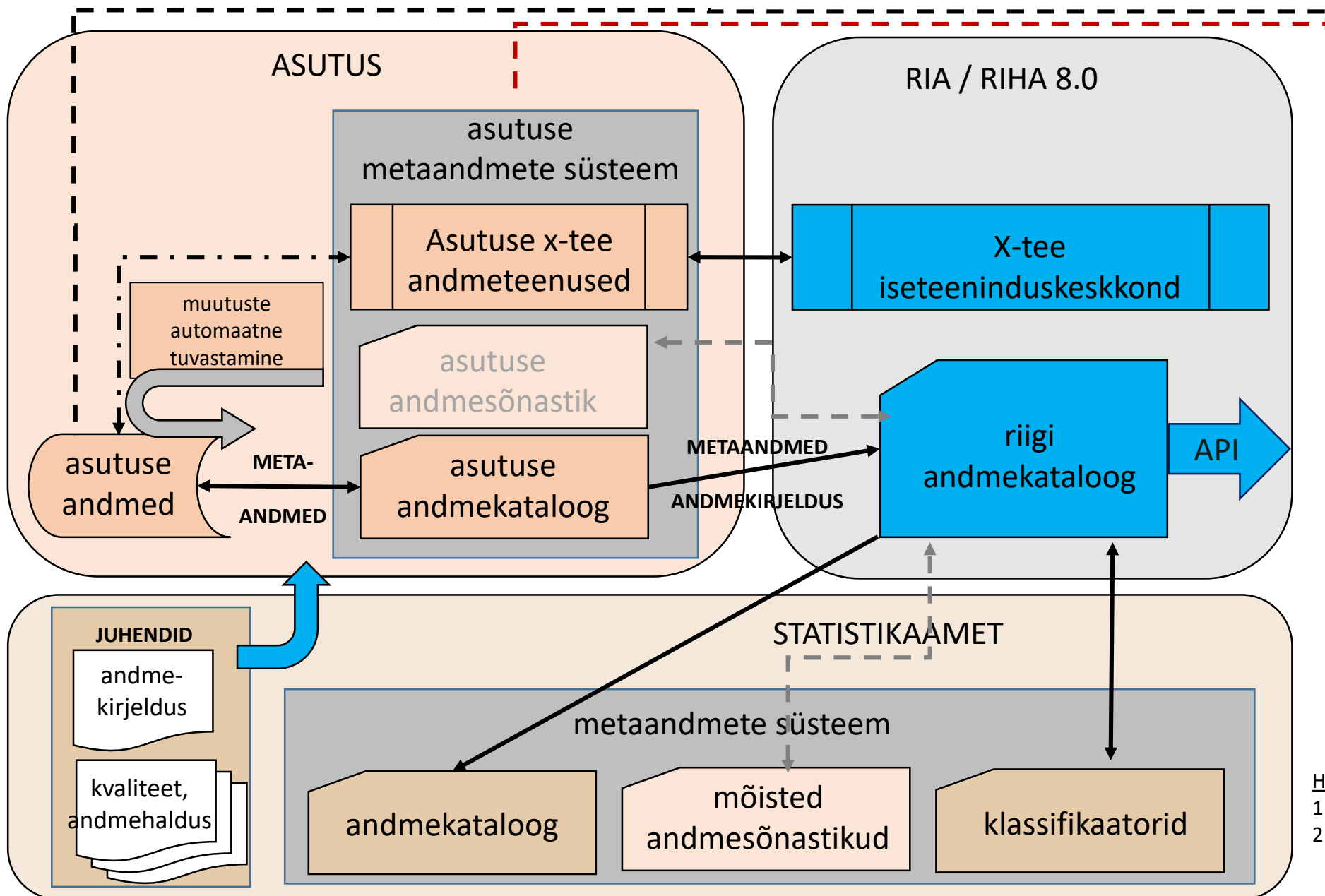
KOOSTÖÖ

# Koostöö uue RIHA loomisel

- Ülevaatamisel riigi infosüsteemi arhitektuur laiemalt!
- RIHA selle osa (MKM lähtekoht)
  - 1. muuta kasutatavaks IS arendajale
  - 2. tagada andmekirjelduste vastavus tegelikele andmemudelitele
- Statistikaameti ettepanek
  - Luua andmekataloogid
- Arutelu 16.-17.05 ning 13.06
  - Mitme rolli (IT, äripool, välised kasutajad) silmas pidamine
  - Uus mudel – kataloogipõhisus; uus tehniline dokumenteerimine – XML manifest
  - Uus protsess – kooskõlastamise asemel registreerimine ja järelevalve

# SA ettepanek keskne andmekataloog

avaandmed



## RIHA 8.0 funktsioonid

- andmevahetus asutuse metasüsteemiga
- hierarhiline andmekataloog – sirvitav
- põhiandmete määratlemine
- otsing andmekirjeldustest
- otsing mõistete sõnastikud
- (ametlike) klassifikaatorite kasutuse tuvastamine ja liidestamine
- IS sisese klassifikaatorite (loendite) väärtuste tuvastamine
- teavituste tellimine ja saatmine
- andmestruktuuride automaatne tuvastamine (RIHAKE)

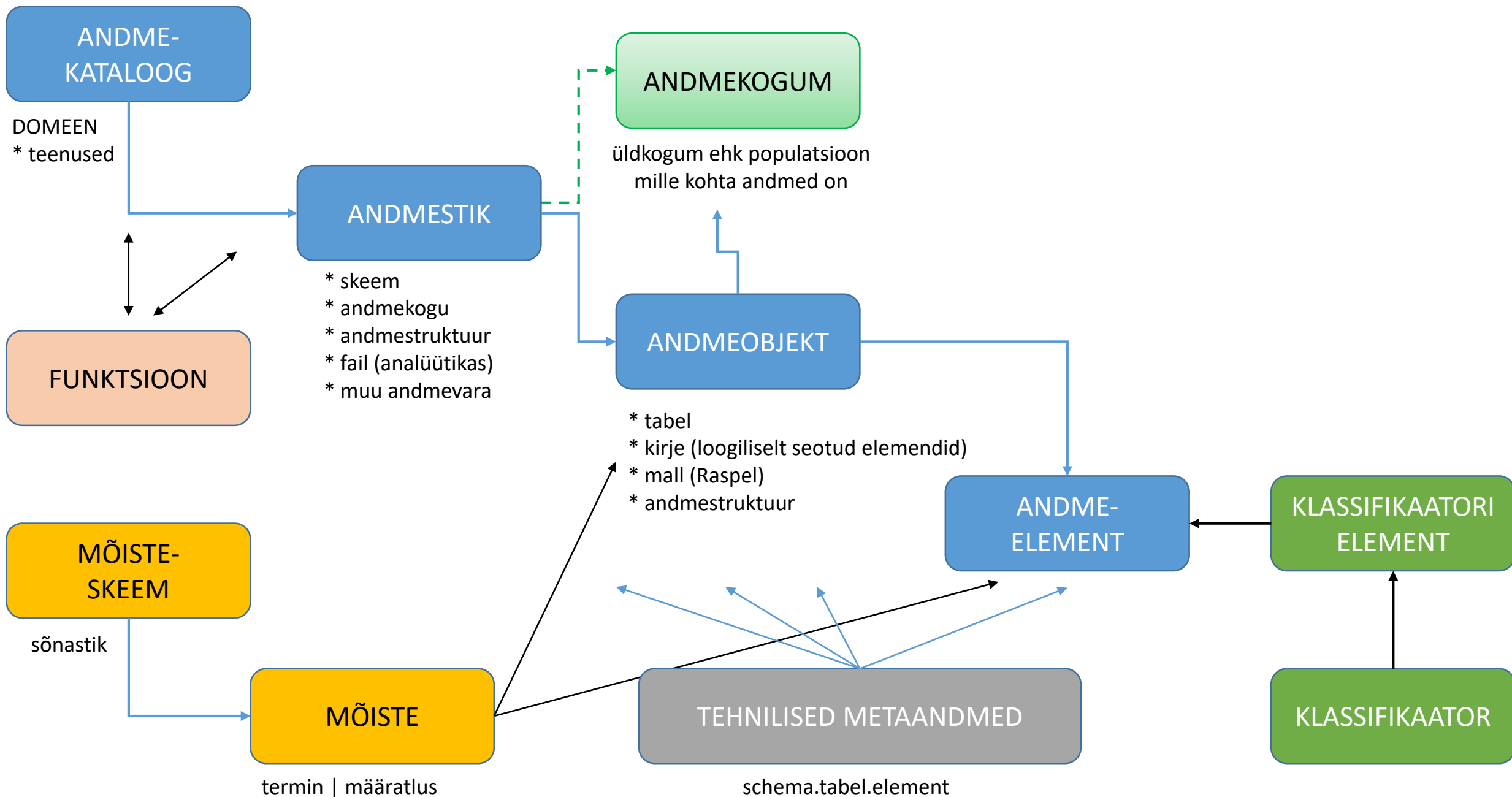
## Hierarhiline andmekataloog:

1. andmekogu – andmetabel – andmeelement
2. andmeobjekt <> kategooria <> tag <> mõiste





„Andmed ja info“  
metaandmed ja andmekirjeldus



# Andmekirjelduse standard

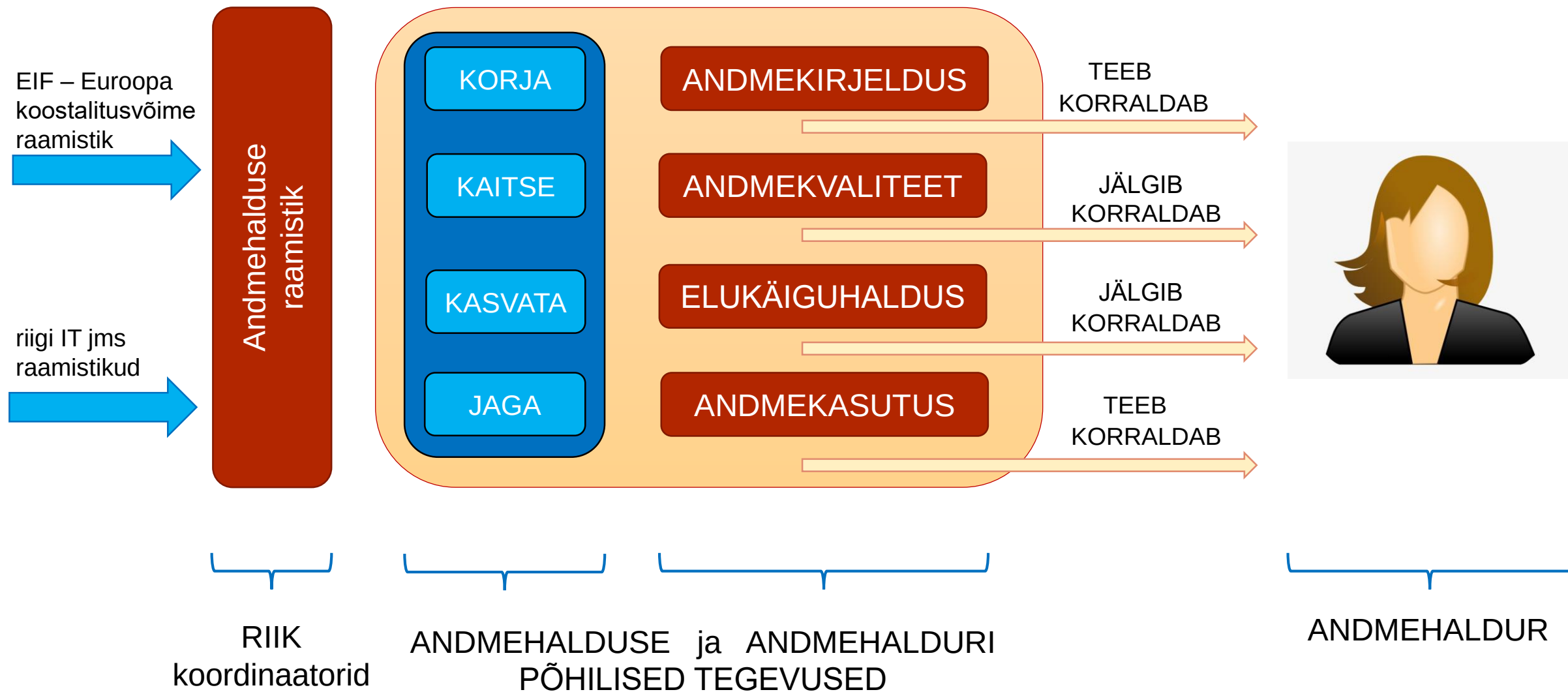
- Standardi tööühm ja kavand
  - Tööühma seisukoht – pigem juhend kui standard s.o selgitava teksti lisamine
- SISU
  - Andmeelemendi kirjeldus
  - Klassifikaatori kirjeldus
  - Tehnilised metaandmed – viide andmevälja kohale
  - Andmestiku kirjeldus
    - Data Catalog Vocabulary (DCAT) - revised edition (W3C Working Draft 16 October 2018)
  - Mõiste (*concept*) käsitletus
    - Kust ja kuidas mõisted 'võtta'?
    - Valdkondlike sõnastike vajadus
    - Rahvusvaheline standard – SKOS (Simple Knowledge Organization System) ei ole tuntud; see on tehniline semantilise veebi lahendus



# Andmehaldur

VASTUTUS JA TÖÖ

# Andmehalduse sisu ja ulatus





# roll, teadmised, oskused

täiendkoolitus

töötamine asutuses

Oskused

Teadmised

ROLL  
ASUTUSES

- Juhtimistegevus ja koostöö
- Arendustegevused
- Rutiinsed tegevused

- töövahendid
  - andmekataloog
- töö sisu
  - kirjeldamine
  - juurdepääsud
  - elukäiguhaldus
  - avaldamine
  - kvaliteet

