

Rahandusministeerium | Statistikaamet

# Eesti andmehalduse juhtimise tegevuskava

2018 – 2022

# Kokkuvõte

Andmete keskne arusaamine organisatsioonide ja ühiskonna toimimisest 21. sajandil on muutumas valdavaks lähenemiseks. Andmetes nähakse sarnast muutuste käivitajat, nagu seda on olnud kaasaegse ühiskonna kujundamisel naftal. Andmeid ei saa kasutada info- ja kommunikatsioonitehnoloogiata (IKT). „Eesti infoühiskonna arengukava 2020“ postuleerib, et IKT on tänapäeval üks peamisi tööriistu mistahes majandus- ja eluvaldkonna konkurentsivõime tõstmisel. Materjal, mida IKT tööriistad kasutavad, on andmed.

Andmetega toimuv vajab eraldi tähelepanu. Distsipliiniks, mis sellega tegeleb, on andmehaldus. Eestis on andmehaldus seni olnud IKT tegevuskavade ja teabehalduse korraldamise varjus. Käesolev dokument võtab esmakordselt kokku küsimused, millega andmehalduse korraldamisel üleriigiliselt on vaja süstemaatiliselt, pidevalt ja juhitult tegeleda.

Andmehaldus (*data governance*) on kokku lepitud järgmises tähenduses:

**Andmehaldus on andmetega seotud otsuste tegemine ja nende rakendamine, mis võimaldab hinnata ja parandada andmete kvaliteeti ning nendele juurdepääsu.**

Andmehalduse juhtimise tegevuskava sisaldab täpsemat ülevaadet andmehalduse kuuest kokkulepitud põhimõttest ja nende elluviimiseks vajalikest tegevustest.

## 1. Andmekirjeldus ja metaandmed

- Andmed on arusaadavad selgitatult ja kontekstis, esitatuna andmete metaandmetes. Kokkulepe metaandmetes tagab andmete mõistetavuse ja semantiliselt koosvõimelise masinloetavuse. Kirjeldused parandavad info leitavust, tõstavad selle usaldatavust.

## 2. Andmestike ülevaade ja põhiandmete määratlemine

- Andmetest on vaja omada kogu aeg ülevaadet. Seda vajavad nii andmete igapäevased kasutajad kui ka IT. Ülevaade peab olema nii andmemudelitest kui ka andmestikest. Eesmärgiks on andmekogumise dubleerimise vähendamine (*once-only-printsii*). Selgus andmete tekke kontekstist võimaldab määratleda riigis, mis on põhiandmed.

## 3. Andmekvaliteet

- Andmeid saab edukalt kasutada siis, kui nende kvaliteet on teada ja see on kõrge. Valdkondade ülene andmekvaliteedi käsitlus annab andmetele usaldusväarsuse nii tekkimise kui ka hilisemates kasutuskontekstides.

## 4. Andmete avaldamine ja avaandmed

- Andmed on kasutamiseks. Andmete kasutamisel tuleb tagada privaatsus ja konfidentsiaalsus. Avaldatud andmetele tuleb lisada kasutuskorda märkiv litsents. Tuleb luua võimalused andmekirjelduste masinloetavaks vahetamiseks ning andmetele viitamiseks.

## 5. Andmete elukäigu haldus

- Andmetel on väärtus. Selle ajas suurenemine või vähenemine sõltub nii andmetest enestest kui ka nende haldamisest. Usaldusväärsed andmed on hallatud kontrollitud keskkonnas. Väärtusetud andmed hävitatakse, kõrge väärtusega arhiveeritakse.

## 6. Andmehaldur ja täiendkoolitus

- Andmete eest tuleb vastutada. Organisatsioonide andmeanalüütika aluseks on ühene mõistmine, kes andmete eest vastutab. Andmete eest vastutaja organisatsioonis on erialaste teadmiste ja oskustega andmehaldur. Andmehaldur liidab IT ja äripoole, omades andmetest ülevaadet, tegeledes andmete kirjelduse, kvaliteedi, avaldamise, juurdepääsu ja elukäiguga.

# Sisukord

Kokkuvõte .....	1
Sissejuhatus .....	3
Mõisted.....	4
1. Andmekirjeldus ja metaandmed .....	7
2. Andmestike ülevaade ja põhiandmete määratlemine.....	8
3. Andmekvaliteet .....	10
4. Andmete avaldamine / avaandmed .....	12
5. Andmete elukäiguhaldus.....	13
6. Andmehaldur ja täiendkoolitus .....	14
Ajakava 2019-2021 .....	16

## Sissejuhatus

Järgnevalt on esitatud andmehalduse tegevuskava, mis lähtub Vabariigi Valitsuse kabineti otsusest 29.03.2018 punkt nr 3: Statistikaameti uuendatud rollist riigi andmehalduses ja teadmuspõhises juhtimises. Otsus on, et „Statistikaameti arengukavas 2018–2022. Statistikaamet riikliku andmehalduse (*data governance*) juhtrollis“ toodud uuenduste rakendamiseks peab Rahandusministeerium koostöös Majandus- ja Kommunikatsiooniministeeriumiga töötama välja vajalikud riikliku statistika seaduse ja teiste seaduste muudatuste eelnõud. Seaduse eelnõude väljatöötamisega paralleelselt kutsus Rahandusministeeriumi asekancler Raigo Uukkivi kokku andmehalduse töörühma. Töörühma eesmärk oli kindlaks määrata ja läbi arutada andmehalduse põhimõtted ja alategevused, määrata elluviimise jaoks vajalikud rollid ja nende täitjad ning panna kokku tegevuskava.

Töörühm leppis kokku ja arutas läbi järgmised põhimõtted:

1. üleriigiline ülevaade andmestikest asutuste kaupa;
2. metaandmed / andmekirjeldus;
3. andmekvaliteet;
4. andmete elukäigu haldus;
5. andmete avaldamine / avalikustamine;
6. andmehaldur.

Töörühmas osalesid: Rahandusministeerium, Justiitsministeerium, Majandus- ja Kommunikatsiooniministeerium, Statistikaamet, Andmekaitseinspeksioon, Riigi Infosüsteemi Amet ning Infotehnoloogia ja Telekommunikatsiooni Liit

Töörühma koosseisu kuulusid: Raido Roop ja Keir Pool (Rahandusministeerium), Mart Mägi ja Veiko Berendsen (Statistikaamet), Evar Sõmer, Liisi Seene ja Mirjam Rannula (Justiitsministeerium), Ott Velsberg (Majandus- ja Kommunikatsiooniministeerium), Joonas Heiter ja Riho Kerge (Riigi Infosüsteemi Amet), Urmo Parm (Andmekaitseinspeksioon), Andres Lille ja Andres Raieste (Infotehnoloogia ja Telekommunikatsiooni Liit)

Töörühm pidas 4 koosolekut: 07.08, 04.09, 25.09 ja 02.10.2018.

Tegevuskava koostas Statistikaamet,  
(Veiko Berendsen, andmehalduse ekspert, veiko.berendsen@stat.ee, 625 8491).

## Mõisted

Termin	Määratlus ja selgitus
<b>andmed</b> ( <i>data</i> )	– on informatsiooni taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötluks Määratlus standardist: <i>ISO/IEC 2382:2015 Information technology – Vocabulary.</i>
<b>informatsioon, teave</b> ( <i>information</i> )	– on teadmus, mis puudutab objekte, näiteks fakte, sündmusi, asju, protsesse või ideid, sealhulgas mõisteid, ja millel on teatavas kontekstis eritähendus Määratlus standardist: <i>ISO/IEC 2382:2015 Information technology – Vocabulary.</i>
<b>metaandmed</b> ( <i>metadata</i> )	– on andmed, mis määratlevad ja kirjeldavad teisi andmeid Määratlus on koostaja tõlge standardist: <i>ISO/IEC 11179 Information technology -- Metadata registries (MDR) -- Part 3: Registry metamodel and basic attributes.</i>
<b>andmestik</b> ( <i>data set</i> )	– on identifitseeritav ja hallatav andmete kogum (koostaja määratlus) <i>Andmestiku mõiste tuleb eristada andmebaasi (database) mõistest, mis on infotehnoloogiline ja andmekogu mõistet, mis on õiguslik.</i>
<b>andmekogu</b> ( <i>database – ainult selles kontekstis</i> )	– on riigi, kohaliku omavalitsuse või muu avalik-õigusliku isiku või avalikke ülesandeid täitva eraõigusliku isiku infosüsteemis töödeldavate korrastatud andmete kogum, mis asutatakse ja mida kasutatakse seaduses, selle alusel antud õigusaktis või rahvusvahelises lepingus sätestatud ülesannete täitmiseks Avaliku teabe seadus § 43 <sup>1</sup> lg 1 <i>Andmekogu on paljudel juhtudel registri sünonüüm.</i>
<b>andmehaldus</b> ( <i>data governance</i> )	– on andmetega seotud küsimuste üle otsuste langetamine ja nende ellu viimine Määratlus on koostaja tõlge <i>Data Governance Institute</i> definitsioonist. <i>Andmehaldus on nii data governance kui ka data management.</i>
<b>andmeteenus</b>	– on X-tee liikme teenus, mille kaudu toimub internetipõhine andmevahetus Infosüsteemide andmevahetuskiht (Valituse määrus) § 2 p 4. <i>Selle tõttu, et andmeteenuuse mõiste on juba õiguses kasutusel, on kasutusele võetud andmeajagamisteenuse mõiste.</i>
<b>andmeajagamisteenus</b> ( <i>data sharing service</i> )	– on andmete edastamine neid loonud või saanud asutustest, organisatsioonist, ettevõttest või üksikisikult nende kasutamiseks andmetöötluks teisele asutusele, organisatsioonile, ettevõttele või üksikisikule, millega võib kaasneda nende eelnev töötlu (koostaja määratlus)
<b>põhiandmed</b> ( <i>basic data</i> )	– on riigi infosüsteemi kuuluvasse andmekogusse kogutavad andmekogu unikaalsed andmed, mis tekivad andmekogu haldaja avalike ülesannete täitmise käigus

<b>Termin</b>	<b>Määratlus ja selgitus</b>
	Avaliku teabe seadus § 43 <sup>6</sup> lg 1. <i>Põhiandmetega sarnasusi omav mõiste on <u>master data</u>, mis kirjeldab unikaalse identifikaatori abil olemeid, mida toimingutes kasutatakse.</i>
<b>kvaliteet</b> <i>(quality)</i>	– on määr, mille ulatuses objekti olemasolevad omadused vastavad nõudmistele Määratlus on koostaja tõlge <i>ISO 8000-2:2018 Data quality — Part 2: Vocabulary 3.3.1 Degree to which a set of inherent characteristics of an object fulfils requirements.</i>
<b>andmekvaliteet</b> <i>(data quality)</i>	– on määr, mille ulatuses andmete olemasolevad omadused vastavad nõudmistele Määratlus on tõlge <i>ISO 8000-2:2018 3.4.8 Degree to which a set of inherent characteristics of data fulfils requirements</i> (koostaja tõlge).
<b>andmete päritolu jälgimine</b> <i>(data lineage)</i>	– on andmete elukäigu halduse osa, milles fikseeritakse andmete päritolu sh esmane allikas, nende liikumine ajas ning kirjeldatakse, mis andmetega toimub protsessides, mida nad läbivad. Määratlus on koostaja kohandatud tõlge Technopedia'st.
<b>andmekataloog</b> <i>(data catalog)</i>	– on andmebaasi objekti nagu tabelid, vaated, indeksid või muu olemi kohta esitatud metaandmete kirje või struktuur, mis võimaldab eri rollides kasutajatel andmeid leida ja kasutada. Määratlus on koostaja kohandatud tõlge Technopedia'st. <i>W3C määratleb oma soovitusel Data Catalog Vocabulary (DCAT) (16.01.2014) andmekataloogi järgmiselt: – on hallatud metaandmete kogum andmestike kohta</i> (koostaja tõlge). <i>Lähedased mõisted on <u>database catalog</u>, mis on andmebaasi objektide kirjeldus ning <u>data dictionary</u>, mis on andmeelementide kirjeldused.</i>
<b>avaandmed</b> <i>(open data)</i>	– on kõigile vabalt ja avalikult kasutamiseks antud masinloetaval kujul andmed, millel puuduvad kasutamist ning levitamist takistavad piirangud. Määratlus on võetud Wikipediast <i>Õiguslik avaandmete määratlus on avaliku teabe seaduses (RT I, 04.07.2017, 11) § 3<sup>1</sup> Avaliku teabe taaskasutamine, lõige 1: Teabe taaskasutamine on füüsilise või juriidilise isiku poolt sellise avaliku teabe kasutamine, mille üldist kasutamist ei ole seadusega või seadusega kehtestatud korras piiratud (edaspidi avaandmed), ärilisel või mitteärilisel eesmärgil, mis ei lange kokku algse eesmärgiga, mille jaoks see teave avalikke ülesandeid täites saadi või loodi.</i>
<b>klassifikaator</b> <i>(classification)</i>	– on riigi infosüsteemi kuuluvas andmekogus kasutatav täpselt kirjeldatud, üksteist välistavate ning number- või tähtkoodiga tähistatud kategooriate põhjalik ja korrastatud süsteem. Määratlus on võetud vabariigi valitsuse määrusest „Klassifikaatorite süsteem“ (RT I 2008, 4, 27).

Termin	Määratlus ja selgitus
	<p><i>Statistikas on klassifikator vaatlustega saadud põhjalik eristuvaid ja üksteist vastastikku välistavate [mõistete] hulk, mida võib omistada ühele või mitmele mõõdetavale ja/või andmeesitusel kasutatavale muutujale. (CODED – United Nations, online Glossary of Classification Terms, 8 June 2017) (koostaja tõlge).</i></p> <p><i>Klassifikaatoriga on seotud koodiloend (code list), mis on statistikas kindlaks määratud terminite hulk, millest kodeeritavatele statistilistele mõistetele omistatakse väärtusi (CODED – SDMX Glossary 2016).</i></p>
<p><b>koostalitusvõime (koosvõime) (interoperability)</b></p>	<p>– on kahe või enama süsteemi või komponendi võime vahetada informatsiooni ning seda vahetatud informatsiooni kasutada</p> <p>Määratlus on koostaja tõlge standardist <i>ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies. Thesauri for information retrieval. 2.29.</i></p>
<p><b>andmehaldur (data steward)</b></p>	<p>– on isik või organisatsioon, kellele on pandud vastutus kindlaksmääratud andmete kogumite haldamiseks</p> <p>Määratlus on koostaja tõlge <i>ISO 8000-2:2018 Data quality — Part 2: Vocabulary 3.14.4: Person or organization delegated the responsibility for managing a specific set of data resources.</i></p>
<p><b>andmete elukäik (data lifecycle või data life cycle)</b></p>	<p>– on järjestikkused etapid, mida konkreetne andmestik / andmete hulk läbib oma algsest loomisest või saamisest kuni hävitamise või arhiivi(asutusse) üleandmiseni</p> <p>Määratlus on koostaja kohandatud tõlge TechTarget'ist.</p> <p><i>Andmete elukäigu haldus (data lifecycle management – DLM) on tegevused elukäigu kestel ja nende korraldamine.</i></p>

# 1. Andmekirjeldus ja metaandmed

Andmekirjeldus ja metaandmed on andmehalduse esimene oluline nurgakivi. Kasutatavad metaandmed peavad ühilduma rahvusvaheliste standarditega. See parandab nii nende kvaliteeti kui ka kasutatavust, soodustab andmeteenuid. Standarditele toetuvad andmekirjeldused on masinloetavad, samas semantiliselt koosvõimelised. Seejuures on eriti olulised semantilist koostalusvõimet võimaldavad standardid, nagu W3C soovitusel.

Metaandmete teemal fookuseerib andmehaldus neljale erinevale metaandmete tüübile / valdkonnale: 1. kirjeldavad metaandmed ehk andmekirjeldus, 2. protsessi metaandmed, 3. administratiivsed metaandmed, 4. tehnilised metaandmed. Andmemudelid, andmeteenuid, andmekvaliteet, andmete avaldamine ja andmete elukäigu haldus kasutavad kõik metaandmeid. Selged metaandmed parandavad oluliselt ülevaate saamist andmetest, andmestikest ning loovad parema aluse põhiandmete määratlemiseks riigis.

Eraldi osana andmekirjeldustes toimib klassifikaatorite süsteem. Klassifikaator hõlmab ühetüübilisi objekte – identifitseerib need ja lisab tunnustele koodid. Klassifikaatorid on nii Eestis kui rahvusvaheliselt vastavate objektide (näiteks: kaubad, ametid, riigid) ja nende omistatud koodide loetelud. Klassifikaatorite kasutamine ühtlustab käsitletavaid objekte ja/või nende omistatud koode, sealhulgas rahvusvaheliselt.

Andmekirjeldus ja metaandmed ei ole ainult infotehnoloogiline käsitlusala. Kaasaegse andmehalduse toimimisel on aga infotehnoloogilistele standarditele toetumine vajalik. Infosüsteemides metaandmete haldamisel on üldiselt aluseks rahvusvaheline standard *ISO/IEC 11179 Information technology — Metadata registries (MDR)* (Infotehnoloogia – Metaandmete register.) Selle osa 1 *ISO/IEC 11179-1:2015 Part 1: Framework* (Raamistik) esitab andmekirjelduse kui metaandmed nende elementaarseima ühiku, milleks on andmeelement, tasemel. Nende metaandmeelementide hallatud nimestik ongi metaandmete register. Standardi osa 5 *ISO/IEC 11179-5:2005 Part 5: Naming and identification principles* (Nimetamise ja identifitseerimise põhimõtted) annab soovitusel andmeelementide nimetamise reeglitele. Andmeelementide täpne kirjeldus on vajalik nii sisuliseks kui tehniliseks andmete käitlemiseks. Lisaks elemendi tasemele on andmed kirjeldatud mitmel üldistuse või rühmitamise moel. Sellisteks on andmete loogilised ja mõistelised mudelid, kirjeldused andmestike, andmekogude jt inforessursside kaupa. Veebis oleva digitaalse info kirjeldamisel on üldiseks aluseks inforessursi kirjeldusstandard *EVS-ISO 15836:2011 Informatsioon ja dokumentatsioon. Dublin Core'i metaandmeelemendid*. Dublin Core on 15 elemendist koosnev kirjeldusstandard, mida kasutatakse ühtlustatud kirjelduse koostamiseks. XMLis esitatuna on see semantiliselt koosvõimeline ja masinloetav. Andmehalduses on enamlevinud kirjelduse mudeliks: *Data Catalog Vocabulary (DCAT)* (*W3C Recommendation 16 January 2014*). See soovitusel on disainitud veebis avaldatud andmekataloogide koostalusvõime toetamiseks. Keskseks kirjeldusobjektiks on selles andmestik (*data set*) ja see toetab mõistelisust ja avaldamist.

Andmekirjeldused võivad sisaldada märksõnastikke, tesauruseid, taksonoomiaid ja ontoloogiaid, mis suurendavad masinloetavuses info tähendusest arusaamist. Andmekirjeldused võivad hõlmata ka ühtluspealkirjade koostamist.



Eesmärgid:

- Andmekirjeldus peab olema semantiliselt koosvõimeline ja masinloetav.
- Ühtlustatud andmekirjeldus peab parandama andmetest arusaadavust nii tehniliste kui äriliste kasutajate poolt, kes ei ole andmete vahetud loojad.
- Andmekogud peavad kasutama kehtestatud klassifikaatoreid.
- Metaandmete abil peab saama andmete usaldusväärsust tagades hallata andmete elukäiku.

## Tegevused ja vastutajad

### 1.1 Andmekirjelduse standardi koostamine.

Vastutaja: Statistikaamet | Kaasatud: MKM, AKI, kõik huvipooled, pöörates erilist tähelepanu andmete kirjeldamisele, mis on avaandmed ning isikuandmed

### 1.2 Semantilist koostalituvõimet edendavate valdkondlike sõnastike koostamine.

Vastutaja: Statistikaamet | Kaasatud: MKM, RIA, valdkondlikud IT- asutused, muud asutused, teadusasutused ja raamatukogud

### 1.3 Ülevaate omamine rahvusvahelistest valdkondlikest andmehaldusega seotud standarditest. Standardite ülevõtmine, tõlkimine ja juurutamine sh andmehaldusega seotud mõisted ja terminoloogia.

Vastutaja: luua pidevalt tegutsev andmehalduse metaandmetega tegelev töörihm | Kaasatud: kõik huvipooled

### 1.4 Klassifikaatorite ühtlustatud kirjelduste juurutamine ja klassifikaatorite haldussüsteemi uuendamine.

Vastutaja: Statistikaamet | Kaasatud: valdkondlike klassifikaatorite haldajad, riigi infoarhitektuuri osas MKM ja RIA

## Vajalik ressurs

- Statistikaametis vastutajana ametikohad ning vahendid juhtimiseks ja koordineerimiseks.
- Rahvusvaheliste standardite üle võtmiseks ja tõlkimiseks ning rahvusvahelises standardimise koostöös osalemiseks.
- Statistikaametis klassifikaatorite halduse funktsionaalsuse uuendamine selleks sobilikus infosüsteemis.

## 2. Andmestike ülevaade ja põhiandmete määratlemine

Ülevaade andmestikest on andmehalduse teine nurgakivi. Andmed on olemas eri moel: andmekogud, andmebaasid, failid jms. Neid ka kirjeldatakse eri moel, millest siinses kontekstis on oluline eristada kahte vaadet: esmalt andmemudeli vaade ja teisalt andmete kui kogumi (andmestiku) vaade. Mudeli kirjeldus andmetest ülevaate saamiseks on tavaline andmekogude / andmebaaside puhul, kuid näiteks juba analüütiliselt töödeldud andmestike puhul või publitseeritud avaandmete puhul peaks kirjeldus käsitlema andmekogumit, mis tegelikult eksisteerib ja mille kirjeldus võib olla vajalik esitada madalaimal s.o andmelemendi tasemel. Andmed on eri moel, neid käsitletakse eesmärgist lähtuvalt eri moel, kuid nende haldamiseks

tuleb andmete kogum piiritleda. Piiritletud olemit nimetatakse erinevalt: *entity*, *data asset*, *data set* jne.

Iga ülevaate esmaseks funktsiooniks on kiirendada ja parandada asjast arusaadavust. Selle saavutamiseks on vaja üksikud andmemelemendid rühmitada. Tavalised rühmad on andmebaasi kirjed (näiteks registrikanded), andmestikud (*data set*), failid (näiteks SPSSi või Exceli fail) või dokumendid (e-arve, XBRLis tehtud aruanne, mis on küll samuti failid). Ülevaate vorm on kahesugune: 1. andmekataloog, kus kirjeldatakse andmeid detailselt, elemendi tasandil ning 2. andmestike kirjeldus. Andmestike kirjeldus võib olla andmekogu üldine kirjeldus, kuid võib olla ka analüütilise andmestiku kirjeldus.

Ülevaadet on võimalik hallata asutuste kaupa, valdkondlikult, üleriigiliselt ning veel mitmel moel. Ülevaade peab olema kasutatav nii tehnoloogia kui äripoolele. Millise rakendusega asutused ülevaate (andmekataloogi) loovad on nende valik. Keskselt on kavas anda ette koostamise põhimõtted ja funktsionaalsuse soovitusel.

Eestis on traditsioon püüda saada ülevaadet andmetest andmekogude kaupa. 1999. a loodi selleks andmekogude riiklik register (1999). See oli omamoodi registrite register, milles kirjeldati rahvastikuregister, äriregister jt. Peagi arusaam laienes ja asja hakati nimetama riigi infosüsteemiks. Andmekogude riiklik register korraldati 2009-2010 ümber riigi infosüsteemi haldussüsteemiks (RIHA). RIHA loomisel loodeti sellele lisada semantilist koostalitusvõimet toetav osa, kuid semantikavarasid on tekkinud vähe. Küll on suure eraldi osana olemas X-tee teenuste kirjeldus, mille WSDLides on edastatavate andmete kirjeldus. RIHA edasine areng on küll andmehalduse tegevuskavaga haakuv, kuid selles dokumendis seda täpsemalt ei käsitleta.

Loomisest peale on üleriigilise registri ja RIHA eesmärk olnud andmekogumise dubleerimise vähendamine. Selle tagamiseks on kasutusse võetud põhiandmete mõiste. Avaliku teabe seaduses (§ 43<sup>6</sup>) on põhiandmed määratletud kui riigi infosüsteemi kuuluvasse andmekogusse kogutavad andmekogu unikaalsed andmed, mis tekivad andmekogu haldaja avalike ülesannete täitmise käigus. Kuigi liidestusi eri infosüsteemide vahel on tehtud palju, on samas paljudel juhtudel põhiandmeteks määramine deklaratiivne.

Ülevaate saamiseks ei ole võimalik ette valmis kirjutada toimivat metoodikat. Esiteks on asutused erinevad, selge pole ka äripoolele kasutatav vorm. Selle tõttu on õige kavandada eri tüüpi andmestikega äri poolelt kokku puutuvate asutuste / funktsioonide osas andmestike ülevaate pilootprojekte. Nende realiseerimise järel on võimalik metoodika üldistada.

Eesmärgid:

- Saada olemasolevatest andmestikest tegelik ülevaade nimekirja või kataloogi kujul, asutuste või valdkondade kaupa.
- Edendada mõistelist (kontseptuaalset) andmestike käsitlust nii andmemudeli vaates (kontseptuaalne mudel) kui ka andmestike vaates, lisades nende kirjeldusele mõisted.
- Praktiliste tööde toel (pilotprojektid) töötada välja toimiv andmekataloogi haldamise metoodika.
- Panna toimima andmestikest ülevaate automaatne edastus asutustest RIHAsse ja vajadusel teistele andmehaldusega tegelevatele asutustele.

- Kavakindlalt määrata andmekogude / infosüsteemide põhiandmed ning jälgida nende kasutamist unikaalsuse (mitte-dubleeritavuse) tagamiseks.

## Tegevused ja vastutajad

**2.1** Üleriigilise andmehalduse raamistiku rakendamine. Infoarhitektuuriliste valdkondlike mudelite rakendamine.

Vastutaja: a) üleriigiline raamistik: Statistikaamet; valdkondlikult: IT asutused | Kaasatud: IT asutused, MKM, RIA, ITL

**2.2** Andmekataloogi koostamise ja haldamise juhis.

Vastutaja: Statistikaamet | Kaasatud: IT asutused, asutused; ITL

**2.3** Funktsionaalsusnõuded andmestiku haldamiseks.

Vastutaja: Statistikaamet ja MKM | Kaasatud: RIA, ITL

**2.4** Riigi infosüsteemis eri tüüpi andmestike haldamiseks infoarhitektuurilise lahenduse ja rakenduse loomine (riigi infosüsteemi haldussüsteemi RIHA uuendamine).

Vastutaja: RIA ja Statistikaamet | Kaasatud: MKM, Andmekaitseinspeksioon, asutused

**2.5** Riigi põhiandmete määratlemine.

Vastutaja: JUM ja MKM | Kaasatud: Statistikaamet, RIA, AKI, Maa-amet, Rahvusrhiiv, IT asutused, ITL

**2.6** Pilootprojektid – ülevaadete loomine andmestikest valitud asutustes (3-4 asutust või funktsiooni).

Vastutaja: Statistikaamet | Kaasatud: valitud asutused

## Vajalik ressurss

- Statistikaametis vastutajana ametikohad ning vahendid koordineerimiseks sh riigi põhiandmete üle järelevalveks.
- Uuendatud funktsionaalsusega RIHA loomine ja rakendamine.
- Andmekataloogi haldusvahendid asutustes.

## 3. Andmekvaliteet

Andmekvaliteet on kolmas andmehalduse nurgakivi. Andmekvaliteet on tavaliselt esitatava määratluse kohaselt sobivus kasutamiseks (*fitness for use*). Mis on küllalt kvaliteetne ühel eesmärgil ei pruugi seda olla teisel eesmärgil. Andmekvaliteedi hästi rakenduvad käsitlused on olemas eri valdkondade kohta. Valdkondade ülesed mudelid (*Total Data Quality Management – TDQM, Complete Data Quality Management – CDQM, Semantic Data Quality Management Framework – SDQM* jt) on seevastu raskesti rakendatavad, sest on ulatuslikud (kuni 70 tunnust). Valdkondlikud raamistikud, näiteks meditsiin, statistika või piltide osas, on aga omavahel üsna erinevad.

Andmekvaliteedi osas eksisteerib hulk mudeleid, mille hulgast tuleb rakendamiseks valida ja kasutusele võtta see, milles kokku lepitakse. Lisaks kvaliteedihalduse erinevatele mudelitele on Eestis Majandus- ja Kommunikatsiooniministeeriumi tellimisel koostatud andmekvaliteedi uuring ja koostanud andmekvaliteedi tagamise jälgimiseks ja juhtimiseks juhend. (*Andmekvaliteedi uuring. Lõpparuanne. (August, 2016); Andmekvaliteedi tagamise juhend andmekogu omanikele. Juhend. (Sept, 2016)*). Juhend esitab andmekvaliteedi juhtimise raamistiku, mille alusel toimub andmekvaliteedi juhtimise pidev parendamine küpsustasemetete tõstmiseks, selle rakendamise meetodika ning andmete kvaliteedi haldamiseks soovituslikud andmekvaliteedi tunnused (komplekt sisaldab üheksat kvaliteeditunnust). On oluline teha vahet, kas andmekvaliteedi raamistikku rakendatakse andmestikule või infosüsteemile (andmekogule). Viimase puhul lisandub kvaliteedi käsitlusele tavaliselt kvaliteedi parendamise mehhanism, mis kannab küpsusmudeli (*maturity model*) nimetust.

Andmekvaliteedi haldus on praktikas asutustes seotud *ISO 9000 Quality management* (Kvaliteedijuhtimine) kohase kvaliteedijuhtimise käsitlusega ehk kvaliteediprotsessiga. Lisaks on olemas *ISO 8000 Data quality* (Andmekvaliteet) andmekvaliteedi standardite perekond. See on andmekvaliteedi valdkonna terviklik süstemaatiline ülevaade. Selle kasutus on Eestis teadmata, standardiks see üle võetud ei ole.

Andmed, olles kasutuses, võivad muutuda ehk transformeeruda. Seal hulgas toimuvad transformatsioonid andmeanalüüsil. Andmete analüüs toimub erinevates valdkondades ja näiteks pilditöötlusel on andmekvaliteet midagi muud kui andmete statistilisel töötlusel. Teaduses ja üldiselt andmete töötlemisel nende kvaliteedi kindlaks tegemise võimalus ja andmetöötuse jälgitavus on osa andmete usaldusväärsusest. Kogu andmete elukäigu järgimine (*data lineage*) võimaldab kvaliteeti tuvastada kogu aeg. Euroopa statistikasüsteemi (ESS) kvaliteedikäsitlus vaatleb teemat kahes osas: andmetöötusprotsessi ja väljundi/toote kvaliteedikriteeriumid. Vaadeldavaid kriteeriume on kokku üheksa.

Eesmärgid:

- Leppida kokku valdkondade ülene kvaliteedimudel, mis on seotud rahvusvaheliste valdkonna käsitlustega.
- Luua asutustes lihtsalt rakendatav andmekvaliteedi tagamise süsteem.
- Luua süsteem, mis võimaldab andmete avaldamisel/edastamisel teistele ning andmeanalüütikas veenduda andmekvaliteedis.

## Tegevused ja vastutajad

**3.1** Andmekogude / infosüsteemide kvaliteedi tagamise juhise uuendatud versiooni koostamine.

Vastutaja: MKM ja Statistikaamet | Kaasatud: AKI, asutused

**3.2** Andmeanalüütikas, sh. statistikas vajaliku andmekvaliteedi tuvastamise ja kirjeldamise meetodika ja juhise loomine.

Vastutaja: Statistikaamet | Kaasatud: teadlased; rahvusvahelist statistikat kasutavad asutused

## Vajalik ressurs

- Statistikaametis vastutajana ametikohad ning vahendid juhtimiseks ja koordineerimiseks.
- MKM ressursid andmekvaliteedi tagamise juhendi uuendamiseks.
- Ressursid publitseeritud andmete sh avaandmete kvaliteedi jälgimiseks (võib olla automatiseeritud süsteem).

## 4. Andmete avaldamine / avaandmed

Andmete ja teabe avaldamine / avalikustamine on üha ulatuslikum. Selle reguleerimiseks on nii kohustavad (teabe avalikustamine) kui ka piiravad isikuandmete, ärisaladuse, konfidentsiaalsusnõude kohta käivad õigusaktid. Andmete avaldamisel piirab nende teisest kasutamist võimatus veenduda andmete usaldusväärsuses. Kuigi avaldatud on palju andmeid, on see teave pigem informatiivne ja ei võimalda sellele toetuda kui faktilisele informatsioonile, mille alusel saaks teostada toiminguid või rajada äriteenuseid. Osalt on tegemist andmekvaliteedi küsimusega, kuid õigete andmete kättesaadavuse teema on olulisuse tõttu eraldi välja toodud. Samuti on olulised andmete käitlemisel privaatsuse ja konfidentsiaalsuse tagamine. See võib põhineda erinevatel anonümiseerimise ja pseudonümiseerimise ning otsest tuvastamist mittevõimaldavale kujule viimise (*otmimise*) meetoditel.

Eestis põimuvad avaandmed ja avalik teave kui lähedased küsimused. Avaandmete õiguslik regulatsioon piirdub termini nimetamisega sulgudes avaliku teabe seaduses, määratledes, et avaandmed on teave, mille üldist kasutamist ei ole seadusega või seadusega kehtestatud korras piiratud. Avaliku teabe seadus kohustab teabevaldaja (s.o asutuse) tal oleva teabe avalikustama oma veebilehel. Teisalt on asutused teinud eraldi avaandmete portaale.

Avaandmete rahvusvaheline kogukond kasutab nii avaandmete (*open data*) kui ka avatud teadmise (*open knowledge*) definitsiooni. Teadmus on avatud, kui igaüks pääseb sellele vabalt ligi, saab seda vabalt kasutada, täiendada ning jagada — seadmata piiranguid muuks, kui teadmuse allikapärasuse ja avatuse tagamiseks. Avatud tähendab, et igaüks saab vabalt teadmusele juurdepääsu, et seda kasutada, muuta ja jagada tingimusel, et ta vähemalt viitab allikale (näitab ära provenientsi ehk päritolu) ja tagab avatuse. Avatud teadmust käsitletakse selles vaates ühisandmevarana (*data commons*), mille kasutamises võib igaüks osaleda.

Osa andmetest / andmestikest / andmebaasidest on käsitletavad kui teosed. See tähendab, et neile on tekkinud autorlus ja rakenduvad autoriõigused. Andmetele, mis avaldatakse või edastatakse teisele kasutajale tuleks lisada litsentsid, mis kirjeldavad edasisi kasutustingimusi.

Teadusandmete avaldamisel on kokku lepitud kindlad printsiibid. Neid tuntakse kui FAIR põhimõtteid: leitavus (*Findability*), juurdepääsetavus (*Accessibility*), koosvõimelisus (*Interoperability*) ja taaskasutatavus (*Reusability*) (*The FAIR Guiding Principles for scientific data management and stewardship, 2016*).

Eesmärgid:

- Andmete ja teabe võimalikult ulatuslik avalikustamine.
- Avaldatud andmete usaldusväarsuse tõstmine.
- Andmete mitmekordse kasutamise parandamine, järgides samas privaatsus- ja konfidentsiaalsusnõudeid.
- Litsentside lisamine avaandmetele.

## Tegevused ja vastutajad

**4.1** Andmete usaldusväarsuse tagamise ja selle tuvastamise meetodika koostamine.

Vastutaja: RIA, Statistikaamet | Kaasatud: JUM, MKM, AKI

**4.2** Anonümiseerimise ja pseudonümiseerimise ning teiste otsest tuvastamist mittevõimaldavale kujule viimise (*otmimise*) meetodite arendamine ja nende juurutamine.

Vastutaja: RIA | Kaasatud: AKI, Statistikaamet

**4.3** Andmetele nende avaldamisel lisatavate litsentside kokku leppimine ja unikaalsete identifikaatorite lisamine ning juhend nende kasutamiseks.

Vastutaja: MKM | Kaasatud: JUM ja Statistikaamet, avaandmete kogukond; teadusasutused ja raamatukogud

## Vajalik ressurss

- Statistikaametis vastutajana ametikohad ning vahendid juhtimiseks ja koordineerimiseks.
- Anonümiseerimise ja pseudonümiseerimise ning otsest tuvastamist mittevõimaldavale kujule viimise meetodite arendamine ja nende juurutamine.
- MKM litsentside ja unikaalsete identifikaatorite lisamise juurutamine.

## 5. Andmete elukäiguhaldus

Andmete elukäiguhaldus on neljas andmehalduse nurgakivi. Mitmetest erinevast elukäigu halduse käsitlesest on andmehalduse juurutamiseks oluline tegeleda 1. protsessi elukäigu haldusega ja 2. elukäigu haldusega, mis tähendab andmete säilitamist. Üheks keskseks printsibiiks elukäigu halduse juures on andmete usaldusväarsuse ja muu kvaliteedi tagamine ajas. Teiseks on andmete väärtuse hindamine. Väärtuse välja selgitamisel ja sellest tuleneva säilitustähtaja määramisel lõikuvad kolm vaadet: 1. andmete tõestusväärtus, 2. andmete kasutamine teaduses, mille alla võib lugeda ka kogu andmeanalüüsi ja statistika ning 3. andmete vastu huvi tundmise ühiskondlik dimensioon, mis annab andmete alusel ülevaatlikkuse ühiskonnas toimuvast ja võimaldab osutada andmeteenuseid.

Andmete elukäiguhaldus on ühtlasi infosüsteemi funktsionaalsus. See funktsionaalsus väljendub ühelt poolt metaandmetes ja teiselt poolt tegevustes/sündmustes/protsessis, milles andmeid kasutatakse ja mida andmehalduse kontekstis nimetatakse *data lineage*.

Eraldi vaade andmete elukäigule on nende haldamine ja säilitamine arhiivisüsteemis. Infosüsteemina nimetatakse arhiive sageli repositooriumiteks. Eristuvadki kaks käsitlust: esmalt

muutumatu andmete vaade ja teisalt *master data* (on seotud põhiandmete käsitlemisega, kuid ei kattu sellega täielikult) kui tõena käsitletava kirje (*system of record – SOR*) vaade. Viimane on siin seostatav põhiandmete temaatikaga. Repositoorium toimib nii andmete säilitamise kui ka avalikustamise kohana ning andmed on selles kirjeldatud ja leitavad. Oluline on see, et nad on selles muutumatud. Repositooriumi olemasolu toetab hallatavate andmete üldist usaldatavust.

Eesmärgid:

- Andmete elukäik on halduse all ja seda toetav metoodika on rakendunud.
- Andmete väärtuse hindamine ja väärtust enam mitte omavate andmete kontrollitud hävitamine.
- Andmete säilitamiseks ja kasutamiseks repositooriumi (andmearhiivi) loomine (Statistikaametis).

## Tegevused ja vastutajad

**5.1** Protsessi kui elukäigu haldus: funktsionaalsusnõuded *data lineage*'i juurutamiseks.

Vastutaja: Statistikaamet | Kaasatud: MKM, ITL

**5.2** Andmete väärtuse hindamine metoodika ja juhise väljatöötamine.

Vastutaja: Statistikaamet ja Rahvusarhiiv | Kaasatud: asutused

**5.3** Asutuses andmestike repositooriumite pidamise nõuete või soovitude väljatöötamine.

Vastutaja: Statistikaamet ja Rahvusarhiiv | Kaasatud: MKM

**5.4** Avaldatud andmestike haldamise ja säilitamise repositooriumi (andmearhiivi) loomine Statistikaametis.

Vastutaja: Statistikaamet

## Vajalik ressurs

- Statistikaametis vastutajana ametikohad.
- Vahendid metoodikate ja juhiste väljatöötamiseks ja juurutamiseks.
- Andmearhiivi / repositooriumi loomise ja juurutamise ressurs.

## 6. Andmehaldur ja täiendkoolitus

Organisatsioonid vajavad hea andmeanalüüsi koostamiseks ning teadmuspõhiste otsuste tegemiseks selget ja arusaadavat ülevaadet andmetest andmestikest ja andmekogudest, nende andmekvaliteedist ja andmete elukäiguhaldusest. Praegu puudub enamikel organisatsioonidel selgus, milline töötaja ja kuidas oleks vastutav andmestike ja andmete eest ehk andmehalduse vaates oleks andmete omanik.

Andmehaldus on infoühiskonna professionaalne tegevus. See tähendab, et andmehalduril on oma kutseoskused. Organisatsioonidel tuleks määrata andmete eest vastutaja (andmete omanik) ehk andmehaldur (*data steward*).

Andmehaldurite saamiseks organisatsioonidesse on vajalik nende täiendkoolitus. Praegu ei ole Eestis neile sobivaid koolitusprogramme. Koolituskava koostamine ja koolitusmaterjalide loomine peab käima koos andmehalduse valdkonna praktilise edendamisega riigis.

Eesmärgid:

- Andmehalduse koolitussüsteemi loomine.
- Professionaalsete teadmiste ja oskustega andmehaldurite olemasolu neid vajavates asutustes.

## Tegevused ja vastutajad

**6.1** Andmehalduse ülesannete täitmiseks vajalike oskuste omandamiseks andmehalduri töökirjelduse ja koolitusprogrammi koostamine. Vajalike koolitusmaterjalide koostamine.

Vastutaja: RM, MKM ja Statistikaamet | Kaasatud: MKM, Kutsekoda, koolitusasutused

**6.2** Täiendkoolituse läbiviimine.

Vastutaja: koolituse korraldajad ja Rahandusministeerium | Kaasatud: Statistikaamet

## Vajalik ressurss




- Koolitusmaterjalide koostamine.
- Täiendkoolituse läbiviimine.










































# Ajakava 2019-2021

Lisatud ajakava arvestab Statistikaameti strateegia „Statistikaamet riikliku andmehalduse (*data governance*) juhtrollis“ 2018-2022 meetmetega ning lähtub riigi andmehalduse töörühmas läbi arutatud teemadest.

Tabel 1: Tabeli 2 tähistuste selgitus

	- poolaasta täpsusega märgitud tegevus/projekt, millel on konkreetne tulem/väljund
	- pidev tegevus
	- pidev tegevus, mille kestel on konkreetsemaid projekte

Tabel 2: Ajakava tegevuskava valdkondade ja alategevuste ning poolaastate kaupa 2019-2021

	2019		2020		2021	
	I	II	I	II	I	II
1. Andmekirjeldus ja metaandmed						
1.1 Andmekirjelduse standardi koostamine						
1.2 Semantilist koostalitlusvõimet 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 juhise						
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 Pilootprojektid – 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 meetodika ja juhise loomine						
4. Andmete avaldamine / avaandmed						
4.1 Andmete usaldusväarsuse tagamise ja selle tuvastamise meetodika koostamine						

	2019		2020		2021	
	I	II	I	II	I	II
4.2 Anonümiseerimise ja pseudonümiseerimise ning teiste otsest tuvastamist mittevõimaldavale kujule viimise ( <i>otmimise</i> ) 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 <i>data lineage</i> 'i juurutamiseks						
5.2 Andmete väärtuse hindamine meetodika 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						