

**EUROPEAN COMMISSION
DG REFORM**

**Eesti andmehalduse
metoodikaprojekt**

Andmekirjelduse juhised

Mai 2020

Sisukord

1. Sissejuhatus

1.1 Andmekirjelduse juhise eesmärgid	2
1.2 Andmekirjelduse juhise sihtrühmad.....	2
1.3 Andmekirjelduse juhise ülesehitus ja põhimõtted	3
1.4 Andmekirjelduse õiguslik raamistik.....	3
1.5 Juhises kasutatud lühendid ja mõisted	3
2.1 Andmekirjelduse ja metaandmete mõiste	5
2.2 Andmekirjelduses kasutatavad sõnastikud.....	6
2.3 Andmekirjelduse koosseis	8
3.1 Andmekirjelduse objektide piiritlemine	10
3.2 Andmekirjelduse standardite ja sõnastike kaardistamine	11
3.3 Andmekirjelduse mudeli kavandamine	12
3.4 Andmekirjeldusega seotud rollide määramine	13
3.5 Andmekirjelduse loomine andmekataloogis	14
3.5.1 Ärisõnastiku loomine ja andmesõnastikuga sidumine	16
3.5.2 Andmekirjelduse loomine.....	18
3.5.3 Andme- ja ärisõnastiku täiendamine.....	20
3.5.4 Andmekirjelduse kvaliteedikontroll.....	20
3.5.5 Andmekirjelduse edastamine	21
3.6 Andmekirjelduse avaldamine	21
3.7 Andmekirjelduse seostamine organisatsiooni tööprotsesside ja teenustega	22
4.1 Andmekirjeldust reguleerivad õigusaktid.....	24
4.2 Andmekirjeldusega seotud standardid	24
4.3 Juhised ja soovitusel	25
4.4 Teiste riikide andmekirjelduse materjalid.....	25
4.5 Andmekirjelduse näited	25
4.6 Andmekirjelduse töövahendid	25
Andmeelemendi kirjeldus.....	26
Andmestiku kirjeldus	28
Mõiste ja sõnastiku kirjelduselemendid	32

1. Sissejuhatus

Käesolev juhise on üks osa Andmehalduse raamistikust, mille eesmärk on aidata organisatsioonil kehtestada ühtseid põhimõtteid, protseduure ning paika seada vajalikke rolle ja mõõdikuid andmehalduse korraldamisel. Andmehaldus on organisatsiooni võimekus hallata andmeid varana. Peamine kasu tõhusast andmehalduse korraldustest asutustele tekib

Andmehalduse raamistik annab organisatsioonile võimaluse

Andmekirjeldus esitab metaandmed andmetike ja neis sisalduvate andmete kohta. Andmekirjeldus hõlmab metaandmeid, mis aitavad tagada andmete:

- sisu kirjeldamise ja seletamise;
- ülevaatlisust ja leitavust;
- usaldusväärsust;
- arusaadavust nii tehniliste kui äriliste kasutajate poolt;
- masinmõistetavust ja semantilist koosvõimelisust;
- elukäigu haldamist.

Süsteemselt hallatud andmekirjeldus tagab organisatsioonile ülevaate oma andmetega seotud varadest andmetike kaupa. See võimaldab andmete leidmist, kasutamist ja neist arusaamist nii inimestele kui infosüsteemidele läbi mõistete ja andmesõnastike. Andmekirjeldus on andmehalduse närvivõrk, mis aitab asutusel hoida ülevaadet andmevaradest, tagab nende usaldusväärsuse ja võimaldab neid tõhusalt taaskasutada.

Andmekirjelduse peamine roll Andmehalduse raamistikus on tagada asutuse siseselt ja asutuste üleselt teadmine olemasolevate andmete ja nende tähenduse kohta.

[Juhise koostas Statistikaameti tellimusel ja EL finantseerimisel EY, 2020 a. kevadel.]

1.1 Andmekirjelduse juhise eesmärgid

Juhise on loodud selleks, et asutustel oleks andmekirjelduse loomiseks ja haldamiseks olemas tegevusjuhised, kvaliteedinõuded ja standardid. 2019. a. sügisel läbi viidud hetkeolukorra analüüs näitas, et andmete kirjeldamisele on seni asutustes vähe tähelepanu pöördunud ja ühtne praktika puudub. Seetõttu on andmekirjelduse juhise peamine eesmärk juhendada organisatsioone andmekirjelduse koostamisel ja haldamisel.

Juhise järgimine aitab kaasa andmehalduse raamistiku terviklikule rakendamisele organisatsioonis. Juhise järgi koostatud andmekirjelduste avalikustamine annab teistele osapooltele ja ühiskonnale tervikuna võimaluse võrreldavaks ülevaateks kogutavatest ja olemasolevatest andmetest. Sellise ülevaate omamine võimaldab andmeid nutikamalt kasutada.

1.2 Andmekirjelduse juhise sihtrühmad

Juhise sihtrühmaks on asutused ja organisatsioonid, kes töötlevad andmeid nii andmekogudes kui ka muudes infosüsteemides ja andmebaasides ning kes kasutavad neid andmeid nii esmase infoallikana toiminguteks ja teenusteks, aga ka teiseseks kasutamiseks aruandluses ja statistikas.

Juhise on suunatud eelkõige erialaspetsialistidele – andmehalduritele, kes korraldavad organisatsiooni andmehaldust ning IT spetsialistidele, kes loovad andmeteenuseid ja andmemudeleid.

Andmekirjelduse juhise peamised kasutajad organisatsiooni sees on **andmehaldusega tegelevad spetsialistid**, kes saavad andmekirjelduse kaudu ülevaate organisatsiooni andmevaradest ning on suutelised korraldama asutuste vahelist andmevahetust; **IT arendajad** ja nende andmespetsialistid, kes saavad ülevaate organisatsiooni andmevaradest ning on suutelised tagama tõhusat asutuse infosüsteemide arendamist ja uuendamist; organisatsiooni **juhid** kes saavad kvaliteetsema sisendi juhtimisotsuste tegemiseks ning andmepõhise juhtimise eest vastutavad töötajad ja teised andmete kasutajad, kes saavad ülevaate organisatsiooni andmevaradest ja nende haldamisest.

Väljaspool organisatsiooni on juhised mõeldud **riigi** keskse **andmehalduse rakenduse halduritele** ja kasutajatele, kes saavad ülevaate sellest, milliseid andmeid kus kogutakse ja hoitakse; IT arendajatele, ning laiemale avalikkusele, kes on huvitatud **avaandmetest** ja nende kasutamisest.

1.3 Andmekirjelduse juhise ülesehitus ja põhimõtted

Juhis annab suunised andmekirjelduse protsesside välja töötamiseks ja juurutamiseks organisatsioonis ning vajalike rollide ja vastutuste määramiseks nende protsesside realiseerimisel. Juhis koosneb kolmest põhiosast, mis seletavad mida ja miks on vaja andmekirjeldusega teha (ptk. 2), kuidas andmekirjelduse haldamist teostada (ptk. 3) ning milliseid abivahendeid seejuures kasutada on võimalik (Lisad). Andmekirjelduse protsessid katavad kirjeldatavate andmetike kindlaksmääramist, andmekirjelduse koostamist ja pidevat haldamist. Juhis on kasutatav koos erinevate andmekirjelduse loomist võimaldavate rakendustega.¹ Juhist on võimalik edaspidi kasutada elava *on-line* keskkonnana, mis on seotud teadmusbaasiga ja mida saab kestvalt täiendada.

Andmekirjelduse juhise koostamisel lähtuti DAMA-DMBOK2 andmehalduse mudelist.² Andmekirjelduse standardi koostamisel tugineti DCAT³ ja DDI⁴ standarditele.

1.4 Andmekirjelduse õiguslik raamistik

Eestis eraldi andmekirjeldust reguleerivat õigusakti ei ole, ega ole kehtestatud ka selgeid nõudeid andmekirjelduse haldamisele. Küll aga sisaldub andmekirjelduse aspekte eri tasandi regulatsioonides, sh. andmekogu-põhised õigusaktid, riigi infosüsteemi ja selle koosvõimet reguleerivad õigusaktid ning raamistikud ja soovituslikud juhised eri asutustelt nagu RIA, Statistikaamet või Rahvusarhiiv. Asutustes on konkreetsete andmekogude kirjeldamise nõudeid esitatud andmekogu (tehnilises) dokumentatsioonis. Ülevaade õigusaktidest, soovituslikest juhistest ja standarditest on juhise Lisas 1.

1.5 Juhises kasutatud lühendid ja mõisted

Kõigil kolmel projekti tulemil saab lõpuks olema ühine mõistete sõnastik.

Termin	Määratlus ja selgitus
Andmed	Informatsiooni taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötamiseks [ISO/IEC 2382]

¹ Käesoleva projektiga samaaegselt oli Statistikaameti ja RIA koostöös loomisel vastav tehniline lahendus projektinimega RIHAKE. Peatükkides toodud näited on loodud tavaliste tabelarvutusvahenditega. Selle juhise ja paralleelselt riigi loodava andmekirjelduse rakenduse mudelid ühilduvad täielikult.

² <https://dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf>

³ <https://www.w3.org/TR/vocab-dcat-2/>

⁴ <https://ddialliance.org/Specification/DDI-Lifecycle/3.3/>

Andmeelement	Elementaarüksusena käsitletav nimega seos käsitusvalla objektide ja neid esitavate sõnade vahel [ISO/IEC 2382-17]
Andmehaldur	Äriprotsesse esindav roll andmehalduse alal: - andmete sisu, konteksti ja metaandmete eest vastutaja; - kohustused sõltuvad kontekstist ja võivad osaliselt kattuda andmekäitleja omadega. [https://akit.cyber.ee/term/2172-andmehaldur]
Andmehaldus	Juhtimis- ja kontrollitegevuste (planeerimine, seire ja kehtestamine) rakendamine andmevaradega seotud tegevuste üle (DAMA DMBOK2)
Andmekirjeldus	Andmeelemendi ning kõigi ta nime ja ta sõnu sisaldavate andmestruktuuride formaliseeritud kirjeldus [ISO/IEC 2382-17]
Andmekogu	Andmekogu 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 [AvTS §41 ¹]
Andmekvaliteet	Näitab, mil määral andmekarakteristikud rahuldavad teadaolevaid või eeldatavaid vajadusi kasutamisel ettemääratud tingimustes [ISO/IEC 25012]
Andmeobjekt	Andmeelement või määratletud andmeelemendikogum, mis on seotud üheainsa tähendust ja kompositsiooni määrava sildiga [ISO/IEC 18013-2] vt ka https://akit.cyber.ee/term/5074-andmeolem-1-andmeuksus
Andmestik	Andmete hulk, mis on avaldatud ja mida hallatakse kindla isiku poolt ning millele saab anda juurdepääsu või seda alla laadida ühes või enamas vormingus [DCAT]
Andmevara	
Metaandmed	Andmed, mis määratlevad ja kirjeldavad teisi andmeid [ISO/IEC 11179-1]
Mõiste	Teadmisüksus [ISO 25964-1]
Märksõna	Termin või ette määratud terminite jada, mis on võetud märksõnastikust [ISO 25964-1]
Märksõnastik	Ettekirjutatud terminite, märksõnade või koodide nimekiri, mille iga liige tähistab mõistet [ISO 25964-1]
Taksonoomia	Kategooriate ja alamkategooriate skeem, mida saab kasutada teadmusüksuste või informatsiooni sortimiseks või muul viisil organiseerimiseks [ISO 25964-1]
Termin, oskussõna	Sõna või fraas, millega mõistet tähistatakse [ISO 25964-1]

2. Andmekirjelduse põhimõtted

2.1 Andmekirjelduse ja metaandmete mõiste

Eri valdkondades ja kontekstides tähendavad andmed erinevaid asju. Arvutisse salvestatud andmetest aru saamiseks tuleb neid alati tõlgendada läbi nende konteksti ja eesmärgi, et vältida vääritimõistmist. Andmete rikastamine kirjeldusega annab neile vajaliku konteksti, et saaksime neid mõista tähendust omava informatsioonina – andmed saavad tähenduse läbi oma metaandmete. Näiteks mõiste „pind“ tähistab erinevaid objekte ehisregistris, põllumaade või metsamassiivide registrites. Metaandmed kirjeldavad igas registris neid andmeid, mida selles objekti „pind“ kohta kogutakse. Andmekirjelduse kaudu oskame nii meie kui tarkvarasüsteemid kasutada andmeobjekti „pind“ õiges tähenduses.

Andmekirjeldus on struktureeritud metaandmete hulk, milles kirjelduselementide ja nendevaheliste seoste kaudu esitatakse kirjeldatava objekti tähendust väljendavat informatsiooni. Andmekirjelduse objektiks on andmestik ja andmeelement ning mõisted, mis seovad andmeelemente tähendust omavateks andmeobjektideks.

Metaandmete kaudu saame vastuse tüüpilistele küsimustele andmete kohta: kes?, mida?, millal?, kus? ja miks? Andmete metaandmete (andmekirjelduse) haldus peab olema korraldatud ja järjepidev, et tagada andmete usaldusväärsuse püsimine ja põlvnemise tuvastatavus (*data lineage*). Metaandmete haldus hõlmab reeglite kehtestamist ja tööprotsesside rakendamist, et kogu organisatsioonil oleks võimalik andmekirjeldust kasutada, jagada, linkida, analüüsida ja töödelda.

Täpsed metaandmed (andmeelementide täpne kirjeldus) on vajalikud nii sisuliseks kui tehniliseks andmete käitlemiseks. Lisaks elemendi (andmevälja) tasemele on andmed kirjeldatud mitmel üldistuse või rühmitamise moel: andmete loogilised ja mõistelised mudelid, andmestike, andmekogude jt inforessursside kirjeldused.

Metaandmete kaudu avaldub andmete väärtus varana,⁵ sest andmekirjelduses sisalduvate mõistete (andmesõnastik) kaudu on organisatsioonil võimalik saada vastused olulistele küsimustele, nagu:

- Mis andmed meil olemas on ja mille kohta?
- Kust need andmed pärinevad?
- Kus andmed praegu asuvad?
- Kas andmed on ajakohased?
- Kuidas on andmed muutunud nende kogumisest alates?
- Kellel on õigus andmeid kasutada ja kuidas?
- Kas tegemist on tundlike andmetega ja millised riskid on nende kasutamisega seotud?

Metaandmed on nagu lisandväärtust pakkuv „keel“, mis peale nimetamise ja liigitamise toimib infosüsteeme ühendava kihina. „Metaandmete keel“ võimaldab andmete ning tööprotsesside ja teenuste koostoimet, nt. andmete sisestamine, leidmine, arvutused ja muul viisil kasutamine, nii inimeste kui tarkvara jaoks. Andmemudelid, andmeteenused, andmekvaliteet, andmete avaldamine ja andmete elukäigu haldus kõik kasutavad metaandmeid. Kvaliteetne andmekirjeldus parandab oluliselt ülevaate saamist andmetest, andmestikest ja loob parema aluse põhiandmete taaskasutamiseks riigis, vähendades seekaudu andmete dubleerimist.

⁵ vt. ka EY, Andmehalduse raamistik (2020), ptk. 3.3

2.2 Andmekirjelduses kasutatavad sõnastikud

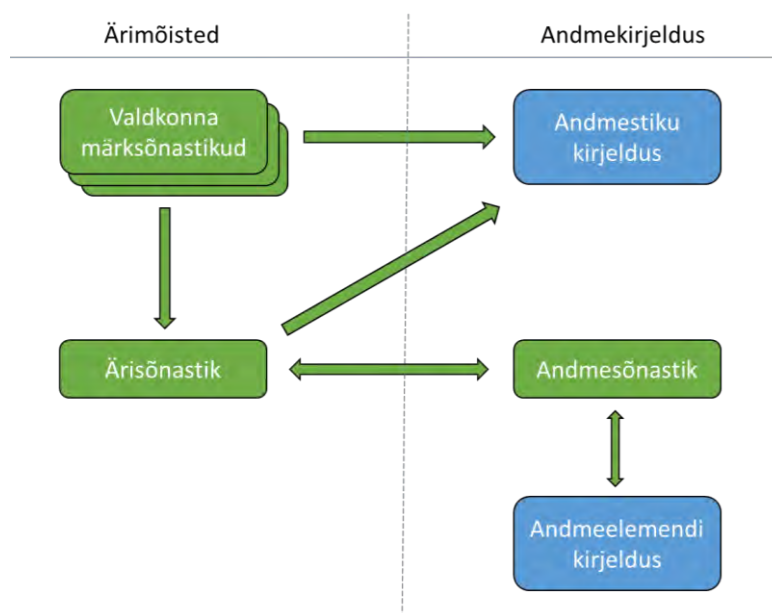
Andmekirjeldus hõlmab metaandmeid andmestike ja neis sisalduvate andmete kohta. Andmekirjelduses on vaja andmete tähistused, mis tihti on lühendid või akronüümid (näiteks relatsioonilises andmebaasis veergude tähised), seletada lahti sõnadega, mis on üldarusaadavad. Tähendust selgitavad sõnad esitavad mõisteid terminitena. **Mõiste** on inimese peas olev ettekujutus (teadmusüksus) mingist objektist või nähtusest, **termin** aga on sõnavaraüksus – sõna või sõnaühend, millega mõistet tähistatakse (nt. sõnastikus). Terminite valikust sõltub sisuline andmetest arusaadavus, ehk mõistet tähistav sõna peab omama üldteadaolevalt või kindlas valdkonnas mitmetele kasutajatele ühesugust tähendust. Et andmete kirjeldamiseks valitud sõnad ei oleks juhuslikud, vaid kokkuleppelised ja laialt arusaadavad, on tuleb sõnadena kasutada termineid ehk oskussõnu. Sellistest sõnadest moodustub märksõnastik, mida andmekirjelduses nimetatakse andmesõnastikuks (*data dictionary*).

Sõnastike kasutamine andmekirjelduses aitab kaasa kirjelduse ühetaolisusele ja arusaadavusele ning on eelduseks kirjelduse masinmõistetavusele ja andmeteaduse meetodite rakendamisele. Seetõttu ongi selle juhise fookus andmete mõistetavaks tegemine sõnastike abil.

2.2.1 Sõnastike tüübid

Andmekirjelduse koostamisel ja haldamisel kasutatakse kolme tüüpi sõnastikke (vt. ka **Error! Reference source not found.** ja Joonis 2):

- **valdkonna märksõnastik** (*controlled vocabulary*)
- **ärisõnastik** (*business glossary*)
- **andmesõnastik** (*data dictionary*)



Joonis 1. Andmekirjelduse ja ärimõistete sõnastike üldine mudel.

Valdkonna märksõnastik on märksõnade korrastatud loetelu, kus tuuakse välja valdkonna terminite ja nende seoste kirjeldused. Andmestiku taseme kirjelduses kasutatakse valdkonna sõnastikust võetud märksõnu, mis seovad andmestiku kontrollitud märksõnastikega (mis võivad olla ka klassifikaatorid), valitsemisfunktsioonidega (VFK), Eesti Majanduse Tegevusalade Klassifikaator (EMTAK/NACE), õigus

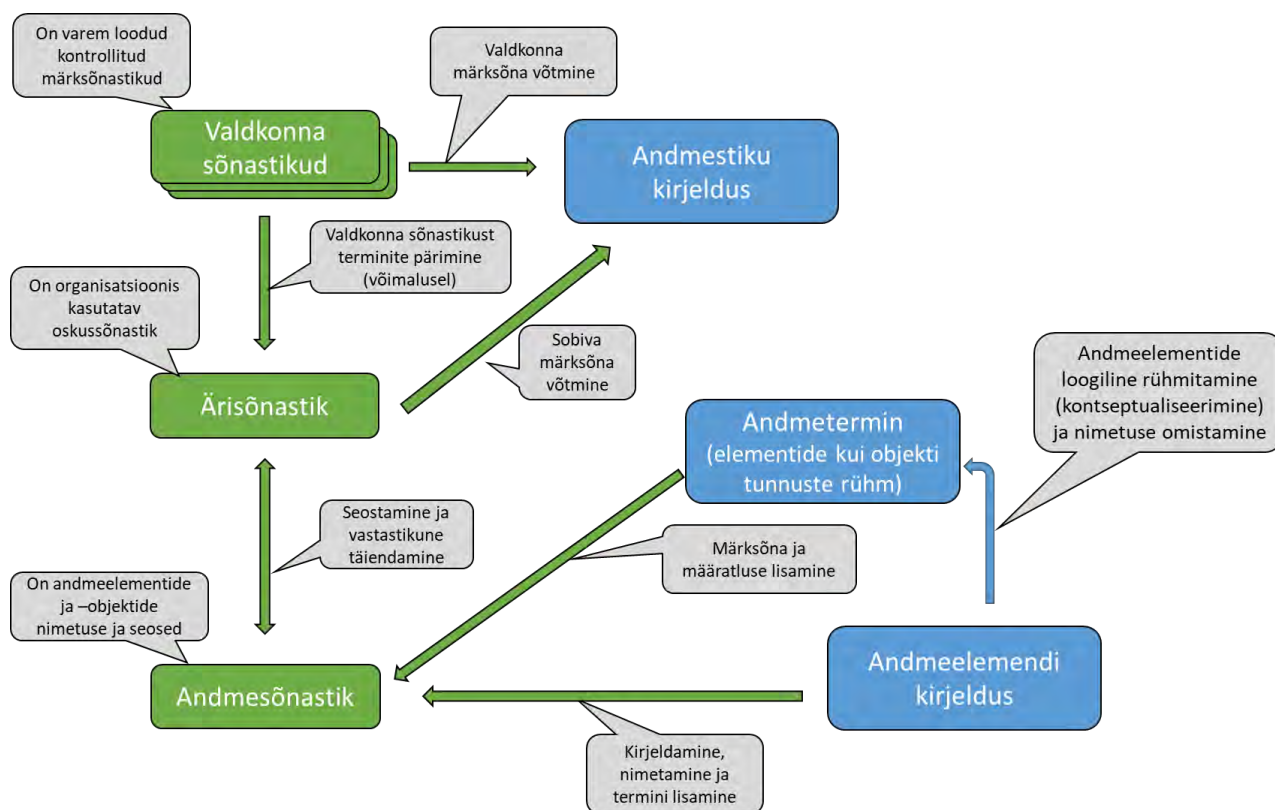
(EuroVoc) või muu valdkondlik jaotus. Valdkonna sõnastiku näiteks hariduse valdkonnas on EstCore2.⁶ Lisaks terminitele ja nendevahelistele seostele toob masinloetav versioon välja ka andmeelementide kirjeldused kasutades owl:DatatypeProperty kirjeldusi ja nende seosed mõistetega. Samuti on see märksõnastik seotud teiste märksõnastikega (<https://schema.org/> jt.). Valdkondlikke märksõnastikke kasutatakse ühe terminite allikana ärisõnastike loomisel.

Ärisõnastik on organisatsioonis kasutatav oskussõnastik ehk sõnaseletuste loend, mis fikseerib organisatsiooni terminoloogia. Ärisõnastik määratleb organisatsiooni tegevusega seotud mõisted. Ärisõnastik on reeglina tesauruse struktuuriga märksõnastik, mis võimaldab näidata mõistetevahelisi seoseid. Ärisõnastiku kaudu saavad andmestikes sisalduvad andmed omale konteksti ja andmeobjektid tähenduse. Ärisõnastiku mõistete kaudu on võimalik andmed ja nende kasutamine siduda organisatsiooni tööprotsesside ja teenustega ning kehtestada andmetele mõttekaid kvaliteedi reegleid. Ärisõnastiku koostab ja seda haldab organisatsiooni äripool ja selle kasutusala on reeglina laiem kui üksnes andmekirjelduse toetamine.

Andmesõnastik on ühest küljest (1) andmemudelil kasutatavate andmeobjektide ja andmeelementide kirjelduste loetelu ja teisest küljest (2) andmekirjelduse käigus tekkiv terminiloend. Andmesõnastik kirjeldab andmeid asutuste tegevuse terminites, aga sisaldab lisaks ka andmete kasutamiseks vajalikke metaandmeid (nt. andmetüübid, andmestruktuuride kirjeldused, turvapiirangud, jmt.). Andmesõnastiku sisu saadakse enamasti andmebaasi loogilisest mudelist ja seotakse ärisõnastiku terminitega.

Joonis 1Joonis 2 on näidatud detailsem infovahetuse loogika sõnastike vahel ning andmeelementide rühmitamine mõtteliseks andmeobjektiks, mida tähistav mõiste lisatakse terminina andmesõnastikku.

⁶ <https://schema.edu.ee/> ja vastav masinloetav sõnastik
<https://github.com/hitsa/estcore2/blob/master/ontologies/haridus.rdf>



Joonis 2. Sõnastike vahelise infovahetuse mudel andmekirjelduses.

Sõnastike näiteid Ehtisregistri baasil on toodud juhise peatükis 3 ning sõnastike haldamise vahendeid käsitletud juhise lisas 1, samuti Andmehalduse raamistiku peatükis 5.

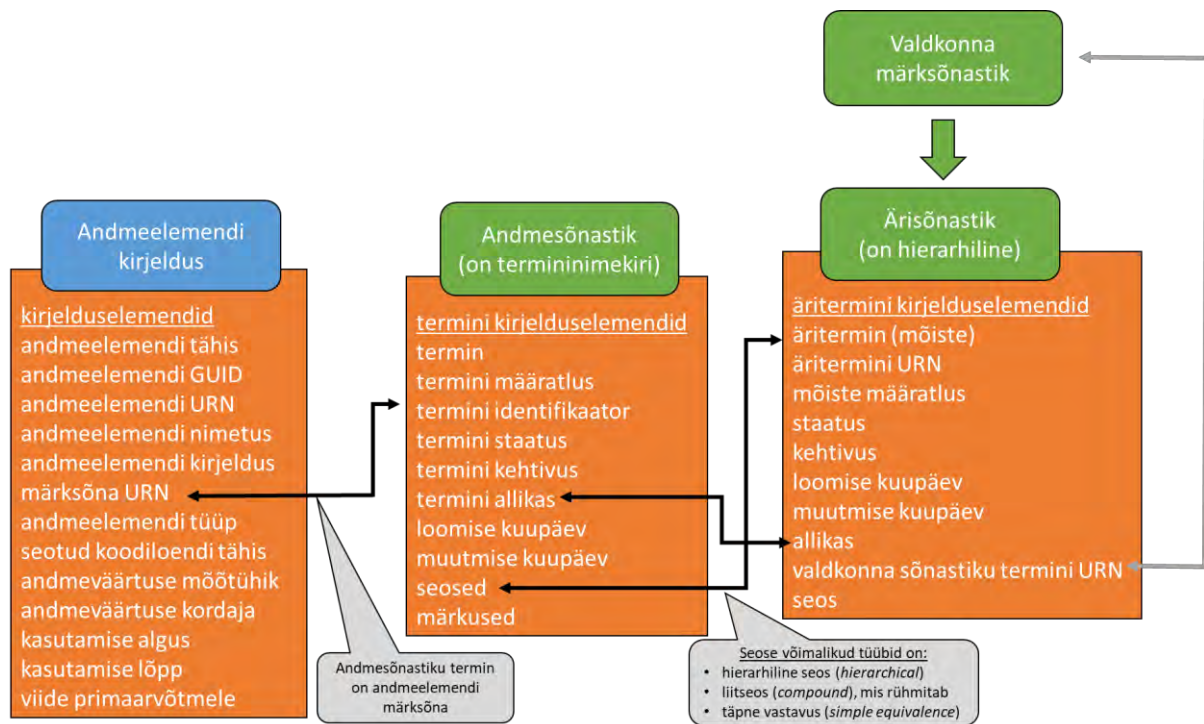
2.3 Andmekirjelduse koosseis

Andmekirjeldus koosneb kahest peamisest komponendist: andmestiku kui terviku kirjeldus ja üksikute andmeelementide kirjeldus.

Andmestiku kirjelduse aluseks on Eestis DCAT standard, nõutavad kirjelduselemendid on toodud andmekirjelduse standardis (Lisa 2). Andmestiku kirjelduse eesmärgiks on tagada andmestiku kui terviku leitavus ja taaskasutatavus. Andmestiku kirjelduses peab sisalduma üldine info selle omaniku ja kasutusvõimaluste kohta (s.h. andmestiku kasutamist võimaldav veebileht, kehtivad juurdepääsupiirangud jms). Vajadusel saab andmestiku kohta kirjeldada ka tema geograafilise kattuvuse (näiteks Tallinna linn, kogu Eesti) või andmete piirdatumid (näiteks „hõlmab andmed alates 1995. aastast“). Sellised kirjeldused aitavad kasutajatel mõista kas sama teemat katvad andmestikud on omavahel kattuvad või ajas ja ruumis erinevad.

Andmeelementide kirjeldamisel tuleb Eestis kasutada kohandatud ja lokaliseeritud DDI standardit, mille nõutavad kirjelduselemendid on toodud andmekirjelduse standardis (Lisa 2). Andmeelementide kirjelduse peamiseks eesmärgiks on andmete sisemise struktuuri (näiteks andmebaasi andmemudeli, tabelite ja väljade) kirjeldamine selliselt, et need oleks üheselt mõistetavad ja taaskasutatavad. Peamiseks kirjeldamise objektiks on üksik andmebaasi väli (veerg), mille kohta peab kirjeldaja sisestama välja tähise andmebaasis, inimloetava nimetuse, lühikirjelduse, andmetüübi ja kasutusele võtmise aja. Võimalusel peab viitama ka välja väärtuse aluseks olevale klassifikaatorile või koodiloendile.

Andmeelemendi kirjelduse seoseid sõnastikega kirjeldab Joonis 3.



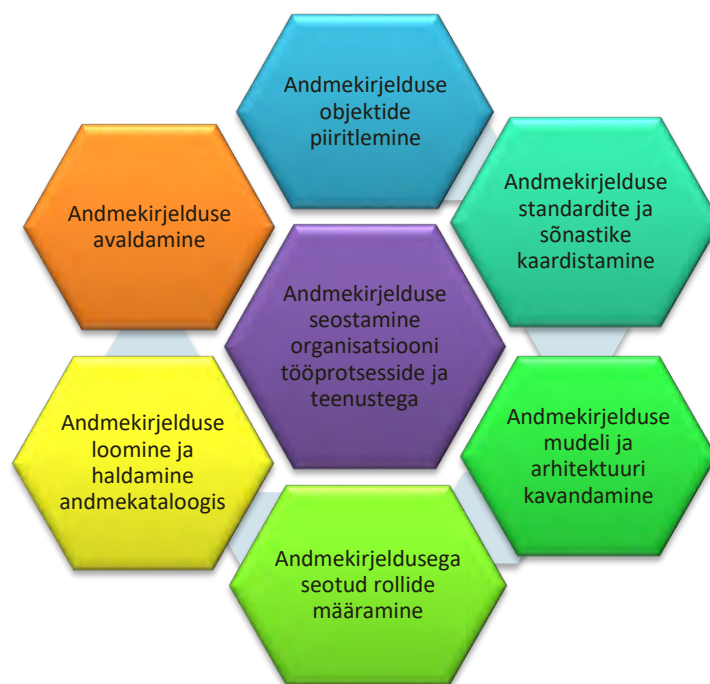
Joonis 3. Andmeelemendi kirjelduse seosed sõnastikega.

3. Andmekirjelduse juhised

Andmekirjelduse haldamine hõlmab käesolevas juhises kogu andmestike ja andmeelementide kirjeldamise ning sõnastikega seotud tegevuste komplekti. See algab andmete kirjeldamise eesmärkide püstitamisest ja lõpeb andmekirjelduse avaldamisega, sisaldades ka ärisõnastiku loomist, andmekirjelduse edastamist RIHAsse jmt.

Andmekirjelduse halduse tegevused on (vrd. Joonis 4):

- 1) Kirjeldatavate andmestike ja objektide väljaselgitamine, olemasoleva andmekirjelduse kaardistamine
- 2) Organisatsiooni jaoks sobivate kirjeldusstandardite ja märksõnastike kaardistamine
- 3) Andmekirjelduse mudeli ja arhitektuuri määratlemine
- 4) Rollide määratlemine andmekirjelduse haldamisel
- 5) Andmekirjelduse loomine, sh sõnastike koostamine ja laadimine, töötamine andmekataloogiga
- 6) Andmekirjelduse avaldamine
- 7) Andmekirjelduse sidumine tööprotsesside ja teenustega.



Joonis 4. Andmekirjelduse haldamise tegevused.

Andmekirjelduse haldamise korraldus on organisatsioonis hästi tööle saanud siis kui:

- Organisatsioon tunnetab ja tunnistab andmekirjelduse süsteemse haldamise tulemusena paranenud andmete kvaliteeti.
- Andmekirjeldus on kooskõlas valdkonna ja rahvusvaheliste vastavate standarditega ning toetavad tõhusat andmevahetust.
- Andmekirjeldus on hallatud ja kättesaadavad keskse infosüsteemi ehk andmekataloogi (nt. RIHAKE) kaudu.
- Andmekirjelduse kvaliteeti jälgitakse ja parendatakse pidevalt.
- Andmekirjelduse muutmine ja täiendamine järgib kokkulepitud protsessi, mis on tervikliku andmehalduse raamistiku osa.

3.1 Andmekirjelduse objektide piiritlemine

Andmekirjelduse koostamise esimeseks sammuks on kirjeldatava andmestiku (nt. andmekogu, register) valimine. Andmekirjelduse täpsemaks piiritlemiseks analüüsitakse selles sisalduvaid komponente (andmebaasid, tabelid, andmeelemendid).

Organisatsiooni hallatavaid andmeid võib andmestikeks jagada mitmel erineval moel. Näiteks Ehitisregistrit on võimalik defineerida kui ühte andmestikku (nn ehitisregistri andmestik), või kui mitut erinevat andmestikku (ehitise põhiandmete andmestik, järelevalve andmestik, energiamärgise andmestik jne). Andmestiku sobiva piiritlemise eesmärgiks on defineerida andmestik mis on piisavalt suur, et selles sisalduv info oleks täielik ja terviklik, samas piisavalt väike, et andmekirjelduste haldamine oleks organisatsioonile kasulik ja selle andmehaldurile määratud aja jooksul jõukohane. Otstarbekas on alustada organisatsiooni jaoks tähtsamatest andmestikest, mille andmeid kasutatakse põhitegevuses või mis on teenuste oluliseks osaks.

Andmestikus sisalduvate andmebaaside, tabelite ja andmeelementide analüüsi eesmärgiks on eristada olulisemad ja vähemolulisemad komponendid, mis omakorda lubab edaspidi keskenduda ainult oluliste (sisulist väärtust omavahete) andmete kirjeldamisele. Enamuses infosüsteemides on talletatud palju tehnilist tugiinfot mis on küll vajalik andmete turvalisuse ja terviklikkuse tagamiseks kuid pole oluline andmehalduse ja -kirjelduse vaatest.

Kirjeldatava andmestiku defineerimisel ja olemasoleva kirjelduse kaardistamisel on oluline organisatsiooni andmehalduri koostöö valdkonnaspetsialistiga (näiteks vastava infosüsteemi peakasutaja või andmestiku omanik). Andmestiku komponentide (andmebaasid, tabelid, andmeelemendid) analüüsimisel tuleb konsulteerida ka tehniliste halduritega (näiteks andmebaasi administraator).

Andmestiku piiritlemisel võib kasutada järgmiseid näpunäiteid:

- Andmekirjelduse koostamist alusta asutuse jaoks olulisematest ja rohkem kasutatavatest andmestikest. Isegi kui sa ei jõua kõiki andmestikke analüüsida ja kirjeldada, oled selliselt kõige vajalikuma ära teinud!
- Kui andmestik on asutatud ametliku andmekoguna, siis defineeri kogu andmekogu ühe andmestikuna. Kui andmete haldamiseks või taaskasutuseks on vajalik ka väiksemate andmestike defineerimine, saad need hiljem kirjeldada kui andmekogu alamandmestikud (vrdl alamregistrid).
- Andmestik peab olema defineeritud piisavalt suurena, et sisalduvad andmed oleksid terviklikud ja andmestikul kui tervikul oleks selge kasutuspotentsiaal. Üldjuhul tähendab see, et asutuse ühe tegevuse (näiteks lubade andmine, järelevalve, toetuse määramine) käigus saadud ja/või tekkivad andmed peaksid sisalduma ühes andmestikus, mitte pihustatud erinevate andmestike vahel. Näiteks pole mõistlik Ehitisregistri andmestiku piiritlemine selliselt, et selles ei sisaldu hoonele kasutusloa andmise kuupäev.
- Andmestik peab olema piiritletud piisavalt väiksena, et andmete ja andmekirjelduste haldamine oleks andmehaldurile jõukohane. Üldjuhul võiks ühes andmestikus sisalduda omavahel tihedalt seotud tegevuste käigus tekkivad andmed, ehk üks andmestik võiks katta ühe või paar organisatsiooni funktsiooni. Näiteks on mõistlik defineerida andmestikuna "riigieelarve" või "keskkonnaseire andmed", aga mitte "riigieelarve ja keskkonnaseire andmed".

Andmestiku komponentide analüüsi näpunäiteid:

- Komponentide analüüsi on hea teha tarkvara dokumentatsiooni, eelkõige andmemudeli põhjal.

- Andmekirjelduse mõttes “vähemtähtsateks andmeteks” võib üldjuhul lugeda süsteemi logi, andmebaasi kirjade loomise ja muutmise kuupäevi, failide räsiseid ja muud sellist tehnilist tugiinfot.

Nüüdseks olen ma:

- **Saanud aru mis on kirjeldatav andmestik ja millised andmed selles sisalduvad**
- **Kindel, et kirjeldamiseks valitud andmeid on mõttekas kirjeldada**
- **Otsinud üles kogu andmestiku kirjeldamist toetava dokumentatsiooni**

3.2 Andmekirjelduse standardite ja sõnastike kaardistamine

Andmekirjelduse koostamise teiseks sammuks on kirjeldamisel kasutatavas sõnavaras kokku leppimine, ehk valdkonna- ja ärisõnastike kaardistamine.

Andmekirjelduse üks peamisi eesmärke on asutuse andmetest ülevaate pakkumine – millised andmed on olemas, kes ja kus neid kogub ja hoiab. Sellist ülevaadet suudab pakkuda üksnes kirjeldus milles on läbivalt kasutatud ühtset sõnavara (mõisteid). Ühtse sõnavara kasutamine kirjeldamisel aitab organisatsioonilt tõeliselt ära kasutada andmehalduse potentsiaali üle kõigi oma andmestike – tuvastada ühelaadseid andmeid ja dubleerivat kogumist, võrrelda nende kvaliteeti jne. Ühtse sõnavara mittekasutamist võib teisalt võrrelda Google'i otsinguga, kus info otsimiseks peab tihti teostama mitu päringut erinevate märksõnadega ja erinevates keeltes, ning ühe päringu vastuseks võib olla ka omavahel mitteseotud tulemid.

Olemasoleva kirjelduse kaardistamine valmistab ette teadmusbasi, mille abil edasist kirjeldamist teha kiiremini ja efektiivsemalt. Samuti aitab teadmine muudest kirjeldustest vältida vasturääkivusi kirjelduses ning hoida kirjeldamisel ühtset sõnavara teiste allikatega.

Parimal juhul on organisatsioonil juba olemas keskne ärisõnastik mille kasutamine on sobilik ka kirjeldatava andmestiku kontekstis. Tihti on aga vaja laiemalt ringi vaadata ja otsida sobivaid sõnastikke valdkonnast.

Esimeseks võimaluseks on vaadata, kas organisatsiooni mõne muu andmestiku kirjeldamisel on sobivad sõnastikud juba kaardistatud ja kasutusele võetud. Samuti on võimalik uurida kirjeldatava andmestikuga tegelevatel meeskondadel muude tegevuste käigus dokumenteeritud ja selgitatud termineid kohta.

Lisaks organisatsiooni sees sobivate sõnastike otsimisele on võimalik abi küsida ka teistelt samas valdkonnas tegutsevatelt organisatsioonidelt, kuna mitmetes valdkondades on olemas riiklikud või rahvusvahelised valdkonnasõnastikud.

Iga leitud sõnastiku kohta on vaja hinnata selle sobivust ning koguda teavet selle haldamise ja tehnilise keskkonna kohta. Sõnastiku sobivus tähendab, et selles kasutatav sõnavara peab olema nii piisav (kaetud on enamus kirjeldatava andmestiku seisukohast olulisi termineid), kui ka kohane (sõnastiku terminid ja nende sisu vastavad valdkonna õigusaktide ja/või ekspertide poolt kasutatud terminitele). Näiteks on ehitusseaduses defineeritud olulised terminid 'hoone' ja 'rajatis' kui üksteist välistavad: „rajatis on ehitus mis ei ole hoone“. Seega pole mõistlik Ehitisregistri andmete kirjeldamisel kasutada (hüpoteetilist) sõnastikku mis defineerib rajatise ja hoone kui samatähenduslikud mõisted, või milles sellised mõisted üldse puuduvad.

Sõnastiku haldamise osas peaks organisatsioon eelkõige huvi tundma kes on sõnastiku omanik ning kas ja millised protseduurid on paigas sõnastiku muutmiseks või täiendamiseks. Soovitav on eelistada selliseid sõnastikke mille omanik ja protseduurid on selgelt ja läbipaistvalt määratud. Eelkõige on see oluline olukorras kui asutus avastab andmete kirjeldamise käigus üksikuid sõnastikust puuduvaid olulisi termineid või vigasid – sellises olukorras peab olema võimalik omanikuga ühenduda ja algatada sõnastiku täiendamine või muutmine. Vastupidiselt ei ole asutusel soovitatav panna end olukorda kus tal puudub täielikult kontroll sõnastiku edasise täienemise või muutmise üle. Põhimõtteliselt on asutusel võimalik küll väline valdkonnasõnastik üle võtta (nõ lokaliseerida) ja jätkata selle haldamist asutuse siseselt. Siin tuleb aga arvestada, et pikaajaline sõnastiku haldamine on küllaltki keerukas ja kulukas tegevus, samuti tekib selliselt oht paralleelsete sõnastike haldamiseks erinevates sama valdkonna asutustes. Seega on pigem mõistlik seda teha sama valdkonna asutuste koostöös.

Tehniliselt on soovitatav kaardistada, millises vormingus on sõnastik loodud ja kas selle poole on võimalik ka automaatselt pöörduda või liideste kaudu alla laadida. Üldiselt on tänapäeval eelistatud masinloetavat esitust (XML, RDF jms) omavad sõnastikud kuivõrd neid on lihtsam taaskasutada ja teiste süsteemidega integreerida.

Olemasolevate kirjelduste kaardistamisel võid kontrollida nende dokumentide olemasolu:

- Infosüsteemi, andmebaasi või teenuste dokumentatsioon.
- Andmekogu põhimäärus, valdkonna õigusaktid.
- Andmekogu info avaandmete portaalis.

Juhul kui asutus pole siiski endale sobilikku sõnastikku leidnud, tuleb see nõ „nullist“ ise luua. Vastava tegevuse kirjeldus on leitav peatükis 3.5.1.

Nüüdseks olen ma:

- ***Tuvastanud andmestiku kirjeldamiseks sobivad organisatsiooni sisesed ja –välised sõnastikud***
- ***Veendunud, et sõnastikud on kohased, piisavad, hästi hallatud ning tehniliselt taaskasutuseks sobivad***

3.3 Andmekirjelduse mudeli kavandamine

Andmekirjelduse järgmiseks tegevuseks on eelnevalt piiritletud andmestiku, selle haldamiseks kasutatava tehnilise keskkonna ja sõnastike ühine analüüs ning andmekirjelduse mudeli või arhitektuuri koostamine.

Andmekirjelduse koostamine võib lihtsamal juhul piirduda ühe väiksema andmestiku (milleks on näiteks paarist MS Exceli tabelist koosnev külastajate nimekiri) ning ühe lihtsama sõnastikuga. Keerulisemal juhul võib aga andmehalduse ja –kirjelduse ülesanne olla mitmete mahukate ja keeruka struktuuriga riiklike andmekogude haldamine üle mitmete valdkondade ja sõnastike. Kuigi neil kahel äärmuslikul juhul on tõenäoliselt andmekirjelduse koostamiseks ja haldamiseks kasutatavad lahendused erinevad, on mõlemal juhul tarvilik esmalt kaardistada olemasolevad lahendused ning leida nende hulgast asutuse vajadustele sobiv.

Samuti tuleb andmekirjelduse mudelit ja arhitektuuri kavandades meeles pidada asutuseväliseid kasutajaid – näiteks peab X-Teega ühineda soovivad andmestikud kirjeldama RIHAS, samuti soovib

Statistikaamet saada andmekirjeldusi riikliku statistika kogumisel ning rahvusarhiiv arhiiviväärtuslike andmete arhiveerimisel. Seega on asutuse vaatest mõistlik luua selline andmekirjelduse arhitektuur, kus kasutatavad tarkvarad ja failivormingud on võimalikult avatud ning lasevad vajadusel kirjeldusi lihtsalt eksportida ja taaskasutada.

Andmekirjelduse mudeli ja arhitektuuri kavandamisel on mõistlik esmalt vaadata kui mahukate või keerukate andmestikega tegemist on. Lihtsamatel juhtudel on andmekirjelduse haldamine võimalik ka tavapäraste tabelarvutuse vahenditega (näiteks Microsoft Office, OpenOffice, LibreOffice, Google Sheets). Sellisel juhul tähendab andmekirjelduse mudeli ja arhitektuuri koostamine peamiselt tabeli ülesehituse kokkuleppimist – luuakse omaette tabelid andmestiku, andmeelementide kirjelduse ja sõnastiku jaoks, määratakse tabelite veerud (näiteks andmestiku nimi, lühikirjeldus, piirdatumid) ja hakatakse omakorda andmeelemente ja sõnastiku termineid omavahel kokku viima⁷.

Keerukamatel juhtudel on selline (manuaalne) andmekirjelduse koostamine liialt aeganõudev ja asutustel on pigem mõttekas kasutusele võtta mõni spetsiifiline andmekataloogi tarkvara (näiteks OvalEdge, Collibra, Informatica, erwin vms).

Sobivaima lahenduse leidmiseks tuleks asutusel analüüsida:

- Milliste andmeallikatega (näiteks andmebaasimootoritega) on andmekataloogi tarkvara võimeline automaatselt suhtlema;
- Millisel kujul sõnastikke ja kui lihtsalt on võimalik andmekataloogi tarkvarasse laadida;
- Kas andmekirjelduste koostamine on asutuse andmehalduritele või teistele kirjeldustega tegelejatele piisavalt mugav;
- Kas andmekataloogi tarkvara toetab käesolevas juhises toodud andmekirjelduse standardi (vt lisa X) juurutamist;
- Kas ja millises (avatud) vormingus on andmekirjeldusi võimalik eksportida ning vajadusel välisele kasutajatele (Statistikaamet, RIHA, Rahvusarhiiv jne) edastada;
- Millises mahus on võimalik andmekataloogi tarkvara ja andmekirjelduse tegevusi integreerida teiste asutuse andmehalduse tegevustega (näiteks andmekvaliteedi haldus).

Lõpetuseks tuleb ka meeles pidada, et andmekataloogi tarkvarad võivad olla küllaltki kallid mistõttu tuleb hoolikalt jälgida ka erinevate toodete maksumust.

Nüüdseks olen ma:

- ***Saanud täpselt aru andmekirjelduse mahust ja vajadustest***
- ***Saanud ülevaate võimalikest tarkvaradest ja kirjelduse vormingutest***
- ***Valinud välja mulle sobivaima lahenduse ning loonud selles vajaliku kirjelduse mudeli ja struktuuri***

3.4 Andmekirjeldusega seotud rollide määramine

Kui kirjeldatav andmestik ja olemasolevad sõnastikud on välja valitud, tuleb paika panna täpsem rollide jaotus andmekirjelduse loomiseks, haldamiseks ja kasutamiseks. Praktikas on ärisõnastikel ja

⁷ Pikemalt on andmekirjelduse koostamise tegevusi kirjeldatud peatükis 3.5.

andmesõnastikul organisatsioonis erinevad omanikud või vastutajad, kes peavad omavahel tihedalt koostööd tegema sõnastike täiendamisel. Reeglina on neil erineva taseme juurdepääs ja muutmisõigused teise rolli hallatavas sõnastikus. Märkus: õiguste halduse teostus sõltub oluliselt sõnastike haldamiseks kasutatavast tarkvaralahendusest. Andmehalduse raamistik⁸ kirjeldab põhjalikult andmehalduses osalevaid rolle ja nende ülesandeid. Andmekirjeldusega seotud rollide jaotus on järgmine:

Andmehalduse juht (juhtkonna liige) – korraldab ärisõnastike ja mõistete mudeli loomist.

Andmeomanik (sisuteenuse juht või infosüsteemi peakasutaja) – kirjeldab teenuste/protsesside käigus loodud või muudetud ärimõisted (ärisõnastik) ning seosed organisatsiooni teiste mõistetega (ärireeglid).

Andmehaldur (andmestike eest vastutav roll) – kirjeldab andmestikud ja andmeelemendid vastavalt andmekirjelduse standardile.

Metaandmete analüütik (IT kompetentsiga andmekirjeldaja) – vastutab andmekataloogi pidamise eest, mis sisaldab andmesõnastikku ja nii andmestiku kui andmeelementide kirjeldust.

Need rollid võivad organisatsioonide praktikas olla koondatud ühte ametikohta või olla jaotatud asutuse ja selle IT teenusepakkuja vahel. Mistahes rollijaotuse puhul on oluline eristada vastutust ärisõnastiku ja andmesõnastiku eest ning näha ette järelevalvet teostav roll, mis jälgib andmekirjelduse mõõdikute täitmist ning uuendab neid perioodiliselt.

Rollide ja vastutuste jaotus on otstarbekas kirjalikult fikseerida organisatsiooni andmehaldust korraldavas dokumendis (nt. Andmehalduse reeglid, vt. Andmehalduse raamistik, ptk. 6.7). Need põhimõtted tuleb rakendada sõnastikke haldavates tarkvararakendustes juurdepääsu- ja kasutusõigustena.

Nüüdseks olen ma:

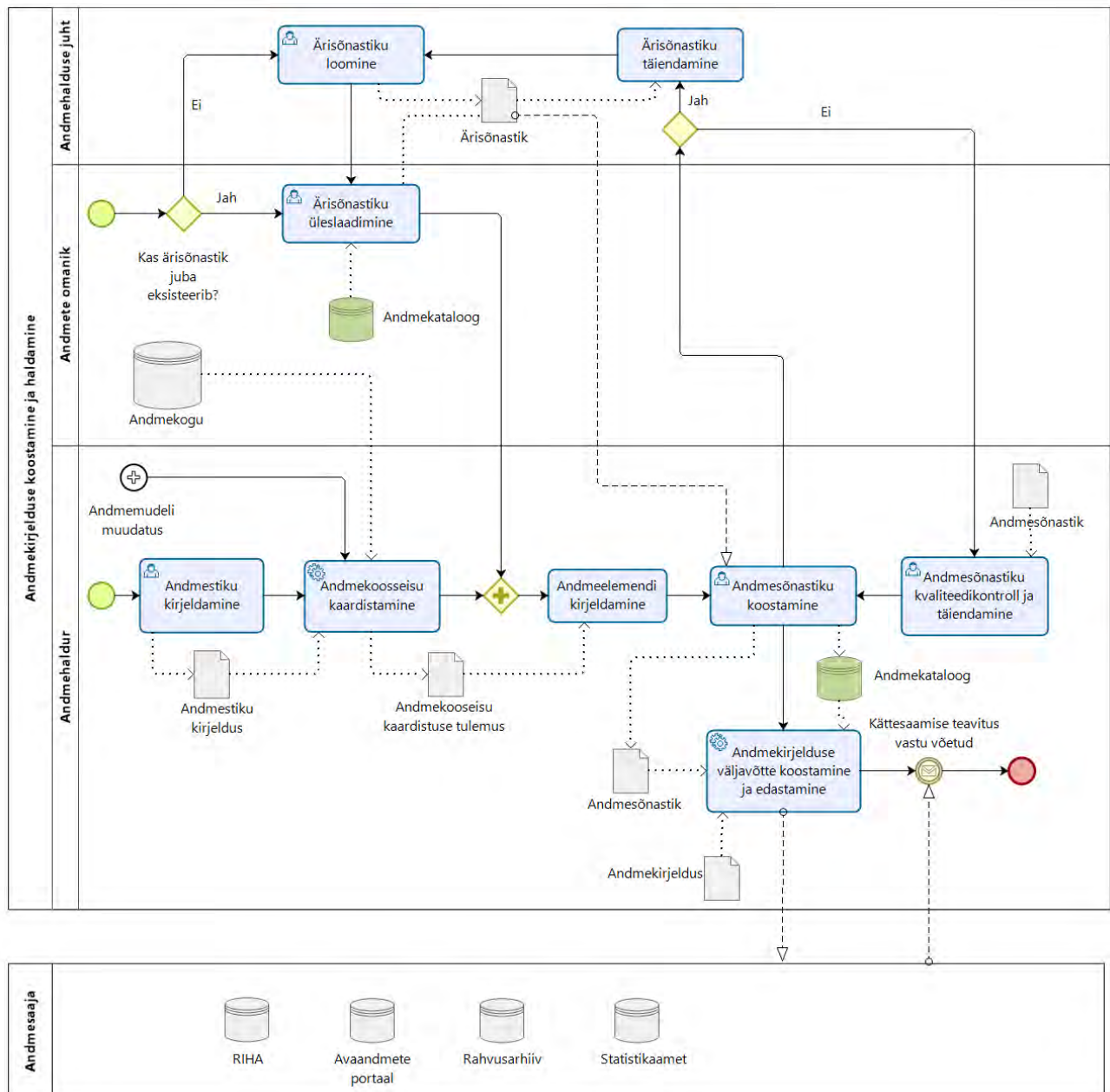
- **Leidnud asutusest üles kõik andmekirjeldusega seotud rolle täitvad isikud**
- **Dokumenteerinud nende isikute oskused ja vastutuse**

3.5 Andmekirjelduse loomine andmekataloogis

Terviklik andmekirjeldus ehk andmesõnastik moodustub kolme kirjelduskomponendi – andmestiku kirjelduse, äri- ja valdkonna sõnastike ning andmeelementide kirjelduse – sisestamisel ja omavahelisel seostamisel.

Alljärgneval joonisel (Joonis 5) on esitatud andmekirjelduse loomise detailsem protsess.

⁸ vt. EY, Eesti andmehalduse raamistik (2020), ptk. 3.3



Joonis 5: Andmekirjelduse loomine ja haldamine andmekataloogis.

Joonisel kujutatud protsessi selgitavad lahti järgnevad alapeatükid. Rollide jaotus protsessides võimaldab samaaegseid tegevusi. Näiteks saab asutuse andmehaldur töötada iseseisvalt andmeelementide kirjeldusega, lisades üksikutele elementidele pikemaid selgitusi. Valdonnaekspert (andmeomanik) saab samaaegselt tegeleda andmestikule kohaste valdkonna märksõnastike valiku ja/või ärisõnastike loomisega ning täiendada andmestiku kirjeldust. Asutuse andmehalduril on võimalik andmeelementide kirjeldustele lisada viiteid märksõnastike terminitele ja teha ettepanekuid ärisõnastiku täiendamiseks uute andmeterminitega.

Andmesõnastiku koostamine ei ole ühekordne hoogtöö vaid pidev tegevus, kus nii erinevaid kirjeldusi kui ka kirjelduste vahelisi seoseid tuleb pidevalt hallata ja vastavalt vajadusele täiendada. Andmekataloogi haldamine, andmekirjelduste parendamine ja erinevate komponentide vahel seoste loomine on seega pidev töö mida tehakse järjepidevalt erinevate rollide koostöös.

3.5.1 Ärisõnastiku loomine ja andmesõnastikuga sidumine

Eeldus: Ärisõnastiku loojal on hea ülevaade valdkonna terminoloogiast, sh olemasolevatest teistest seonduvatest ärisõnastikest, valdkonnasõnastikest, infosüsteemides ja seadusandluses kasutatavast sõnavarast (vaata ka juhise peatükki 3.2 *Andmekirjelduse standardite ja sõnastike kaardistamine* eespool).

Organisatsiooni andmehalduse tegevusi toetava andmesõnastiku aluseks on läbivalt ühtlane sõnavara kasutamine, ühtlase sõnavara aluseks omakordne kvaliteetne ärisõnastik. Reeglina peaks sobiv ärisõnastik olema asutuses loodud juba enne andmekirjelduse koostamist. Samas pole ärisõnastiku puudumine andmete kirjeldamise alustamisel takistuseks, vaid seda on võimalik luua ka kirjelduste koostamise käigus.

Ärisõnastiku koostamisel soovitame meeles pidada mõningaid lihtsaid põhimõtteid:

- Sõnastik peab olema piisav. Eelkõige peab sõnastik sisaldama piisavalt termineid, et katta kõik asutuse olulised andmed. Kõige olulisem on nende terminite täpne defineerimine, mille kohta käivaid andmeid on vaja riskasutada asutusesiseses või organisatsioonide vahelises andmevahetuses;
- Sõnastik on muutuv ja täienev. Asutus peab arvestama, et nii valdkonna sõnavara, seadusandluse definitsioonid kui ka hallatavate andmete skoop on pidevas muutumises. Asutus peab nende muutuste toetamiseks looma kindla reeglistiku mis kehtestab põhimõtted sõnastikku terminite lisamise, muutmise ja kustutamise kohta. Lihtsaimal juhul tähendab see ühe isiku määramist sõnastiku omanikuks ning talle muutmissetepanekute kogumise ja sõnastiku täiendamise ülesande andmist, üldjuhul võiks sõnastiku omanik olla ka andmeomanik;
- ...

Ärisõnastiku loomisel on asutusel kolm peamist võimalust:

- võetakse üle riiklik või rahvusvaheline valdkonna märksõnastik;
- asutus loob riikliku või rahvusvahelise märksõnastiku baasil oma ärisõnastiku;
- ärisõnastik luuakse asutuses oma tegevusvaldkonna õigusaktides ning andmestike ja infosüsteemide dokumentatsioonis kasutatud terminoloogia baasil.

Need kolm võimalust ei ole üksteist välistavad. Näiteks võib asutus kasutada samaaegselt riiklikku valdkonna märksõnastikku ja luua sellest puuduvate terminite jaoks õigusaktide definitsioonidel kasutava asutusepõhise ärisõnastiku. Mitme märksõnastiku samaaegsel kasutamisel tuleb küll jälgida, et erinevates sõnastikes poleks kattuvaid või vasturääkivaid termineid.

Ärisõnastiku loomisel peab asutus jälgima, et nii sõnastik ise kui ka selles sisalduvad terminid oleks kirjeldatud vastavalt andmekirjelduse standardi (Lisa 2) nõuetele. Sõnastiku kirjelduse alusel peab olema võimalik vähemalt aru saada millist valdkonda ja/või asutust sõnastik katab, kes on sõnastiku omanik ning millal on seda viimati muudetud. Iga termini kirjelduses peab sisalduma vähemalt termini pikem kirjeldus, kehtivus (kas termin on aktuaalne või mitte) ja termini sõnastikku lisamise ja/või viimase muutmise kuupäev.

Lisaks eelnevale on asutusel soovitatav analüüsida ja kirjeldada terminite vahelised seosed. Mõned näited sellistest seostest on:

- Hierarhiline seos: „**Hoone** ja **rajatis** on mõlemad **ehitised**“
- Antonüümid: **Hoone** vs **rajatis**

- Sünonüümid: **Hoone, maja**

Üldjuhul on asutusel mõistlik sellised seosed samuti kirjeldada. Eelkõige rahvusvahelistes sõnastikes võivad terminite vahelised seosed olla küllaltki keerulised.

Tehnilisest vaatest on mõistlik igale terminile lisada ka masinloetav identifikaator (URI) mille abil on võimalik iga terminit unikaalselt identifitseerida, muuhulgas saab seda identifikaatorit hästi kasutada ka terminite vaheliste seoste defineerimisel.

Ärisõnastiku näide Ehitisregistri baasil hallatuna MS Exceli tabelina on toodud Tabel 1.

Tabel 1. Ärisõnastiku näide.

Mõiste	Selgitus	Ülemmõiste	Infovara omanik äriüksus	Alam-mõiste	URI
Ehitis	Ehitis on hoone või rajatis.	Asi (kinnis- või vallasasi)	MKM	Hoone; Rajatis	https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/Ehitis
Hoone	Hoone on väliskeskonnast katuse ja teiste välispiiretega eraldatud siseruumiga ehitis.	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/Hoone
Rajatis	Rajatis on ehitis, mis ei ole hoone.	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/Rajatis
Ehitis-registri kood	Ehitise kohta käivate andmete esmakordsel registreerimisel antakse ehitisele unikaalne ehitisregistri kood, milleks on numbrite kombinatsioon.	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/ehitiseEHRKood
Kasutamise otstarve	Ehitise kasutamise otstarve vastavalt MKM 02.06.2015 määrusele nr 51 "Ehitise kasutamise otstarvete loetelu."	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/Kasutamisetstarve
Ehitise nimetus	Ehitise nimetus annab edasi ehitise iseloomuliku tunnust, mille järgi on võimalik seda teistest eristada.	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/ehitiseNimetus
Ajutise ehitise kasutamise aeg	Ajutise ehitise kasutamise aeg (algus- ja lõppkuupäev)	Ehitis	MKM		https://riha.eesti.ee/riha/onto/ettevotlus.toostussektorid.ehitised/2020/r1/AjutiseEhitiseKasutamiseAeg

Ärisõnastiku loomiseks võib organisatsioon lihtsamal juhul kasutada tabelarvutustarkvara, eelkõige kui ka ülejäänud kirjelduse komponendid (andmestiku ja andmeelementide kirjeldus, andmekataloog) kavatsetakse luua samas vormingus.

Kui asutusel on kasutusel spetsiifiline andmekataloogi haldamise tarkvara, on mõistlik ärisõnastik luua selliselt, et seda oleks võimalikult lihtne andmekataloogi laadida. Maailmas on enim levinud ärisõnastiku loomine XML põhiselt (näiteks OWL, RDFS, JSON-LD või SKOS skeemide alusel). Eestis on soovitusel OWL skeemi kasutamine, sõnastiku loomiseks sobib hästi vabavaraline Protege tarkvara **// viide //**.

Kui asutus on ärisõnastiku valmis saanud (või otsustanud taaskasutada teiste asutuste või valdkondade sõnastikku) tuleb see andmesõnastikku üles laadida. Vastavalt kasutatavale andmekataloogi tarkvarale võib see olla võimalik kas manuaalselt (näiteks OWL või tabelarvutusfaili laadimisena) või automaatselt. Olenemata meetodist on oluline, et asutuse ärisõnastiku eest vastutaja igal sõnastiku uuendamisel selle ka andmekataloogi lisab, võimaldamaks uuenenud või muutunud sõnavara kasutuse andmekirjelduses niipea kui võimalik.

Nüüdseks olen ma:

- **Valinud välja või koostanud ärisõnastiku**
- **Laadinud ärisõnastiku andmekataloogi**

3.5.2 Andmekirjelduse loomine

Eeldus: Andmekirjelduse looja on piiritlenud kirjeldatava andmestiku, tal on hea ülevaade andmete vormingust ja struktuurist, kirjeldatavad andmed on grupeeritud nende olulisuse alusel (vaata ka juhise peatükki 3.1 *Andmekirjelduse objektide piiritlemine*).

Andmekirjeldus koosneb kahest peamisest komponendist: andmestiku kui terviku kirjeldus ja üksikute andmeelementide kirjeldus.

Andmestiku kirjelduse aluseks on Eestis DCAT standard⁹, nõutavad kirjelduselemendid on toodud andmekirjelduse standardis (Lisa 2). Kirjeldatakse andmestiku omanikku, eesmärke, kasutamist ja selles sisalduvate andmete põhitunnuseid.

// Näide //

Enamasti on ühe andmestiku kirjeldus küllaltki üldine ja mahult väike, mistõttu on kirjeldajal soovitatav üksikute kirjelduselementide täitmine korralikult läbi mõelda ja pigem luua pikemaid ja põhjalikumaid kirjeldusi.

Andmeelementide kirjeldamisel tuleb Eestis kasutada kohandatud ja lokaliseeritud DDI standardit, mille nõutavad kirjelduselemendid on toodud andmekirjelduse standardis (Lisa 2). Peamiseks kirjeldamise objektiks on just üksik andmebaasi väli (veerg), mille kohta peab kirjeldaja sisestama välja nime andmebaasis, inimloetava nime, lühikirjelduse, andmetüübi ja kasutusele võtmise aja. Vajadusel peab viitama ka välja väärtuse aluseks olevale klassifikaatorile või koodiloendile. Lisaks andmeelementidele peab üldiselt kirjeldama ka nende füüsilise struktuuri, ehk tabelid, andmebaasi skeemid ja väljade vahelised seosed (võtmed).

// Näide ilma viiteta andmesõnastiku terminile //

Kui lihtsamaid andmestikke on võimalik kirjeldada ka käsitsi, siis suuremates andmebaasides võib sisalduda sadu tabeleid ja tuhandeid andmeelemente. Sellisel juhul on mõistlikum kirjeldamisel taaskasutada andmebaasidest eksporditud kirjeldusi, milles üldjuhul sisaldub juba suur osa ülal mainitud kirjelduselementidest (välja nimi, kirjeldus, andmetüüp, viide klassifikaatorile), ning lihtsalt täiendada seda puuduvaga (näiteks välja inimloetav nimi). Samas on siin oluline tähele panna, et enamuses

⁹ <https://www.w3.org/TR/vocab-dcat-2/>

andmebaasides on lisaks sisulist tähendust (ja taaskasutusväärtust) omavatele andmetele ka küllalt palju tehnilisi, infosüsteemi turvalisuse või terviklikkuse tagamiseks vajalikke, andmeid mille põhjalik kirjeldamine pole andmekirjelduse vaatest tarvilik.

Andmekirjelduse koostamise viimaseks sammuks on andmesõnastiku loomine. Oma olemuselt on andmesõnastik andmestiku füüsilise mudeli (üksikute andmeelementide) ja ärisõnastiku ühenduslüli, niinimetatud loogiline andmemudel. Andmesõnastik sisaldab sarnaselt ärisõnastikule üksikuid termineid, peamiseks erinevuseks on terminite kindla ärikonteksti puudumine.

Andmesõnastiku koostamine on võimalik alustada nii andmeelementide kui ärisõnastiku vaatest:

- Alustades andmeelementidest on andmesõnastiku koostaja tegevuseks üksikute andmeelementide grupeerimine (näiteks: isiku eesnimi, perenimi, isikukood), sellele andmeelementide grupile andmesõnastiku termini loomine (näiteks: isik) ja sobiva ärisõnastiku terminiga seostamine (näiteks: ehitise omanik);
- Alustades ärisõnastikust toimub sama tegevus ülalt alla – esmalt valitakse ärisõnastikust termin, luuakse sellele vastav andmesõnastiku termin ning valitakse sellele vastavad andmeelemendid või nende grupid.

// Näide: andmeelement viitega andmesõnastiku terminile -> andmesõnastiku termin viitega ärisõnastiku terminile -> ärisõnastiku termin //

Üldjuhul on korraliku andmesõnastiku koostamisel rakendada vaheldumisi mõlemat, nii alt üles kui ülalt alla lähenemist. Sellega tagatakse, et ühelt poolt on kõik ärisõnastiku terminid seotud kohaste andmeelementidega, ning teisalt kõigile olulistele andmeelementidele määratud seos ärisõnastiku terminiga.

Andmesõnastik

Mõiste	Lähtesüsteem	Tabel	Väli	Tüüp	Seletus
Ehitise nimetus	EHR	EH_EHITIS	NIMETUS	VARCHAR2	Ehitise vabatekstiline nimi. Näiteks kuur, Solarise keskus, Reaalkool.
Ehitise kasutamise otstarve	EHR	EH_EHITIS	KAOS_ID_PEAMINE	NUMBER	Ehitise peamine kasutusotstarve, milleks on kõige suurema kasuliku pinnaga kasutusotstarve. Kordab ühte EH_KASUTUSOTSTARVE.kaos_id väärtust.
Viit kasutusotstarbe klassifikaatorile	EHR	EH_KASUTUSOTSTARVE	KAOS_ID	NUMBER	Viit kasutusotstarbe ID-le
Kasutusotstarbed	EHR	KL_KASUTUSOTSTARBED	KASUTUSVIIS	VARCHAR2	Klassifitseeritud väärtus kasutusviiside kohta. Klassifikaator KASUTUSVIIS.
Ajutise ehitise kasutusele võtmise kuupäev	EHR	EH_EHITIS	AJEH_KASUTALG_KP	DATE	Ajutise ehitise kasutusele võtmise kuupäev.
Ajutise ehitise kasutuse lõppkuupäev	EHR	EH_EHITIS	AJEH_KASUTLOPP_KP	DATE	Ajutise ehitise kasutuse lõppkuupäev.

Nüüdseks olen ma:

Loonud põhjaliku ja läbimõeldud andmestiku kirjelduse

- **Kirjeldanud piisavalt andmeelemente**
- **Tekitanud seosed andmeelementide (füüsilise andmemudeli), andmesõnastiku (loogilise andmemudeli) ja ärisõnastiku vahel**

3.5.3 Andme- ja ärisõnastiku täiendamine

Eeldus: asutuses on loodud esialgne andmekirjeldus, andmesõnastik ja ärisõnastik

Juba andmesõnastiku esialgse koostamise käigus võib ilmned, et kõigi oluliste andmeelementide jaoks ei eksisteeri ärisõnastikus sobivat terminit. Samuti võib aja jooksul muutuda andmestiku ja selles sisalduvate andmete ulatus, näiteks andmebaasi füüsilise muutmise uuendamise või asutuses uute teenuste pakkumise alustamise tõttu mis omakorda toob kaasa vajaduse andme- ja ärisõnastiku täiendamiseks.

Üldjuhul on andmestiku füüsilise mudeli haldaja ja ärisõnastiku omanik erinevad isikud, seega on selles tegevuses oluline roll omavahelisel suhtlusel ning selgelt paika pandud protseduuridel. Soovitav protseduur selleks suhtluseks on järgmine:

- Andmeelementide kirjeldamisel luuakse andmesõnastikku termin kuid sellele ei leita sobivat vastet ärisõnastikust;
- Andmesõnastiku omanik teeb ettepaneku märksõna lisamiseks ärisõnastikku;
- Ärisõnastiku omanik kaalub märksõna lisamist ja kinnitab või lükkab ettepaneku tagasi;
 - a. Kinnitamisel lisatakse märksõna kinnitatud staatuses ärisõnastikku ja lisatakse andmesõnastikku viide märksõnale
 - b. Tagasilükkamisel lisatakse teade sellest märksõna juurde ja kuvatakse kohas, kus seda märksõna kasutati
- Uuendatud ärisõnastik laetakse andmekataloogi

Nüüdseks olen ma:

- **Defineerinud protseduurid ärisõnastiku täiendamiseks**

3.5.4 Andmekirjelduse kvaliteedikontroll

Andmekirjelduse koostamine on küllaltki mahukas tegevus ja piisava kvaliteedi saavutamine ei ole üldjuhul võimalik ühekordse pingutusena. Asutuse vaatest on oluline käsitleda andmekirjeldust sarnaselt andmetega – kehtestada andmekirjelduse kvaliteedile selged nõuded ja regulaarselt kontrollida kas kirjeldused neile nõuetele ka vastavad.

Andmekirjelduse kvaliteedi peamiseks mõõdikuks on enamasti taaskasutus, ehk esmalt peab pidevalt küsima „kas kirjeldus on piisav, et tagada andmete ühene mõistetavus nii majasiseste kui –väliste kasutajate jaoks“.

Võimalikud andmehalduri tegevused andmekirjelduse kvaliteedi tagamisel on:

- Regulaarse andmekirjelduse ja –sõnastiku ülevaatuse teostamine. See võib toimuda kas pisteliselt või/ja hiljuti muudetud ja lisatud kirjelduste osas;
- Valdkonnaspetsialistide intervjuerimine andmekirjelduse ja –sõnastiku kasutatavuse ja ühese mõistetavuse osas;
- Andmete kasutajatele (nii asutusesiseste kui –väliste) tagasiside andmise võimaluse pakkumine.

Nüüdseks olen ma:

- **Defineerinud protseduurid andmekirjelduse regulaarseks ülevaatuseks**
- **Taganud andmekirjelduse kvaliteedi hindamise ja tagasiside kogumise**

3.5.5 Andmekirjelduse edastamine

Eeldus: Asutuses on koostatud piisavalt põhjalik andmekirjeldus

Kuigi andmekirjelduse olemasolu on oluline ka asutusele endale, on olemas ka mitmeid asutusi kellel on huvi andmekirjelduste saamiseks. Nende hulgas on näiteks RIHA (riigi andmetest tervikülevaate haldamine), Statistikaamet (riikliku statistika kogumine) ja Rahvusarhiiv (arhiiviväärtuslike andmete kogumine riigi digitaalarhiivi). Samuti võib asutusel olla huvi andmekirjelduse (osaliseks) avalikustamiseks avaandmetena, ehk edastamiseks avaandmete portaali.

Andmekirjelduse edastamisel on eelkõige oluline, et ei edastataks ainult „suvalisi andmeid“ vaid läbi selgelt defineeritud mõistete konkreetssesse konteksti asetatud infot. Teisiti sõnastatuna peaks asutuse eesmärgiks olema, et edastatud andmekirjeldus on üheselt mõistetav ka väljaspool valdkonnaspetsialistide kitsast ringi.

Andmekirjelduse edastamise peamised tegevused on:

- Andmekirjelduse saaja vajaduste ning toetatud vormingute analüüs;
- Vajadusel kogu andmekirjeldusest saajale vajaliku alamosa defineerimine ja eraldamine;
- Vajadusel andmekirjelduse väljavõtte viimine saajale sobivale kujule (näiteks XML teisenduse teostamine).

Lihtsamal juhul on nende tegevuste läbiviimine võimalik ka käsitsi, näiteks tabelarvutustarkvara abil. Kui asutusel on aga pidev või regulaarne vajadus väliste osapooltele andmekirjelduse edastamiseks, on mõistlik need sammud defineerida omaette tegevusena ning võimalusel rakendada ka andmekataloogis (näiteks protseduurina „Statistikaametile hoonete andmekirjelduse väljavõtte tegemine“).

3.6 Andmekirjelduse avaldamine

KA: Kas eristub kuidagi sisuliselt 3.5.5? Sinnagi ei osanud midagi väga mõistlikku lisada; kui kaheks jagada jääb kummagi peatüki tekst veelgi õhemaks..

Andres Kukke: Siia lisaks kindlasti „automatsieerimise“ s.t oluline on mitte üks kord kirjeldada vaid automaatne hõive tekitada

Tegevuse eesmärgid:

-

Teostajad:

-

Toimingud:

-

Oodatavad tulemid:

- Semantilise kirjelduse kasutamine andmevahetuse toetamiseks
- URId e loomine (kas oskame selle tarvis ka mingeid hinte ja tippe jagada või vähemalt viidata mingitele õpetlikele materjalidele?)
-

3.7 Andmekirjelduse seostamine organisatsiooni tööprotsesside ja teenustega

Andmehalduse raamistik (vt. ptk. 4.1.3, 4.1.5) paneb paika reeglid, kuidas andmeelementidega seotud andmekvaliteedi nõuded seotakse organisatsiooni tööprotsesside ja teenustega. Selliste reeglite kaudu on võimalik tagada, et andmete kvaliteet vastab tööprotsesside ja teenuste kvaliteedi nõuetele. Igale mõistega kirjeldatud andmeelemendile määrab andmeomanik andmekvaliteedi reegli, mis kirjeldab millistes piirides on tulevased andmed kvaliteetsed - näiteks lihtne nõue, et mingi kindel andmeväli ei või kunagi jääda tühjaks (nt. kliendi aadress, hüvitise summa, vmt.), võib oluliselt tõsta andmete kasutatavust ja kasulikkust organisatsioonis (vt. ka Andmekvaliteedi juhise¹⁰ ptk. 3.2). Sõnastikel ja andmekirjeldusel on teenuste kvaliteedi tagamisel oluline roll.

Organisatsiooni tegevust kajastavatest peamistest mõistetest moodustuv ärisõnastik on seotud andmevaldkonda kirjeldava andmesõnastikuga. Nii ärisõnastikku kui andmesõnastikku ja selles sisalduvaid andmeelementide kirjeldusi tuleb pidevalt võrrelda organisatsiooni kehtivate tööprotsesside ja teenuste nõuetega. Nende muutumisel täiendatakse nii sõnastikke kui ka andmekvaliteedi reegleid. See on pidev töö andmekirjelduse kvaliteedi tõstmiseks (vt. Ptk. 3.5.4 eespool) ja andmekvaliteedi reeglite rakendamiseks. Selle saavutamiseks peavad ühelt poolt eksisteerima toimivad andmehalduse protsessid ning teiselt poolt protsessijuhtide ja teenuseomanike strateegiline huvi. Andmehaldurite töö mõistete ühtlustamisel nii organisatsiooni sõnastike piires, andmevaldkondade kui ka organisatsioonide vahel toetab andmekvaliteedi tõstmist ja sujuvat andmete taaskasutust (vt. ptk. 3.6 eespool).

Sõnastike halduse ja andmekataloogi tarkvarade kasutamine on üks võimalus tagada tõhus kontroll asutuse mõistete ja nendega seotud kvaliteedireeglite üle. Ärisõnastike haldamise vahendeid on kirjeldatud Andmehalduse raamistiku peatükis 5.2 ja neid on võimalik juurutada nii laiemate platvormide osana (nt. Atlassian tooteperekond) kui ka iseseisvate rakendustena (nt. Informatica, Colibra, Erwin). Andmekataloogi tarkvaradest on lühidalt juttu käesoleva juhise Lisa 1 peatükis 4.6. Juhise koostamise ajal on RIA-on kavas välja arendada ja asutustele kasutamiseks anda RIHAKEse

¹⁰ EY, Andmekvaliteedi juhise (2020)

rakendus, mis täidaks andmekataloogi rolli, hallates nii andmesõnastikku kui andmestiku ja andmeelementide kirjeldusi ja võimaldades neid edastada RIHA-le ja teistele seotud osapooltele.

4. Lisa 1: Kasulikku lisalugemist

4.1 Andmekirjeldust reguleerivad õigusaktid

Andmekogude peamised andmeobjektid, mida andmekirjeldus katma peab, on reeglina loetletud andmekogu tegevust reguleerivad seaduses või põhimääruses andmekoosseisuna, mille kõrval on määratletud andmekogu pidamise eesmärgid, põhimõisted ja mõnikord ka andmevahetuseks kasutatavad andmed. Näiteks Töökeskonna andmekogu põhimääruses on toodud loetelu faktidest, mille kohta infot kogutakse ja salvestatakse. Haigekassa andmekogu põhimääruses on toodud andmeväljade loend, mida andmekogus hoitakse. Töövõime hindamise ja töövõimetoetuse andmekogus on toodud objektide loend ja iga objekti kohta väljade loend (lühikese kirjeldusega), mis andmekogus salvestatakse.

Kohustuslikke andmekirjelduse elemente loetleb ka VV määruse „Riigi infosüsteemi haldussüsteem“¹¹ § 18, mis käsitleb ka andmekirjelduse haldamise ja edastamise põhimõtteid. Nendega on käesolevas juhises arvestatud.

Riigi infosüsteemi koosvõime puhul eristatakse õiguslikku, organisatsioonilist, tehnoloogilist ja semantilist koosvõimet.¹² Andmekirjelduse kontekstis, mille üheks eesmärgiks on saavutada parem andmetest arusaamine ja kvaliteet, on oluline eelkõige semantiline koosvõime, mis tähendab erinevate organisatsioonide võimet mõista vahetatavat andmeid ja informatsiooni ühtmoodi. Riigi semantilise koosvõime raamistik pärineb aastast 2007¹³ ja ei kajasta vahepeal toimunud arenguid semantikavahendites.

Andmekirjeldusega on seotud ka VV määrus „Klassifikaatorite süsteem“,¹⁴ mis määratleb klassifikaatorite haldamise ja kasutamise ühtsed põhimõtted.

4.2 Andmekirjeldusega seotud standardid

ISO/IEC 11179 Information technology — Metadata registries (MDR) —

ISO/IEC 11179-1:2015 Part 1: Framework (Raamistik)

ISO/IEC 11179-3:2013 Part 3: Registry metamodel and basic attributes (Registri metamodel ja põhiomadused)

ISO/IEC 11179-4:2004 Part 4: Formulation of data definitions (Andmemääratluste koostamine)

ISO/IEC 11179-5:2015 Part 5: Naming and identification principles (Nimetamise ja identifitseerimise põhimõtted)

DDI Lifecycle 3.3 (2020) – Data Documentation Initiative (Andmete dokumenteerimise algatus)

<https://ddialliance.org/Specification/DDI-Lifecycle/3.3/>

GSIM (2019) – Generic Statistical Information Model (Üldine statistiline info mudel) Version 1.2

<https://statswiki.unece.org/display/gsim/GSIM+v1.2+documents>

DCAT (2020) – Data Catalog Vocabulary (Andmekataloogi sõnastik) Version 2

<https://www.w3.org/TR/vocab-dcat-2/>

¹¹ <https://www.riigiteataja.ee/akt/106082019018>

¹² <https://www.mkm.ee/et/tegevused-eesmargid/infouhiskond/riigi-infosusteem>

¹³ https://www.mkm.ee/sites/default/files/riigi_infosusteemide_semantilise_koosvoime_raamistik.pdf

¹⁴ <https://www.riigiteataja.ee/akt/12910889>

SDMX (2013) – Statistical Data and Metadata eXchange (Statistikaandmete ja metaandmete edastamine)
Version 2.1 https://sdmx.org/?page_id=5008

- ISO/IEC 11179 – provides a standardised metadata format to describe and represent data to make it easier to understand the meaning and content of data.
- AS/NZS ISO 19115:2005 – provides a standardised metadata format for describing geographic information and services.
- AS/NZS ISO 15836:2016 – establishes a standard for cross-domain description and defines the elements typically used in the context on an application profile.

KÜSIMUS: kas tuua sisse ka rahvusvaheline mõõde?

- EL oma DCATi ja avaandmete direktiiviga
- Rahvusvahelise kirjeldusstandardid

4.3 Juhised ja soovitus

Selle projekti teised tulemid

Andmekvaliteedi juhis

SEHKE

Arhiiviväärtuslikuks hinnatud andmekogude üleandmist Rahvusarhiivile kirjeldab rahvusarhiivi juhis “Arhivaalide üleandmine”.¹⁵ Rahvusarhiiv arhiveerib ainult arhiiviväärtuslikuks hinnatud andmekogusid,¹⁶ millel eeldatakse andmekirjelduse olemasolu ja selle üleandmist arhiveerimisel.¹⁷

[Siia vaja tekst luua veel nende alateemade ümber:]

- Sõnastike ja ontoloogiate loomise reeglid ja juhised

4.4 Teiste riikide andmekirjelduse materjalid

National Archives of Australia Metadata for Interoperability Guide – this guide provides information on how to develop an organisational Metadata strategy, information on metadata harvesting tools and protocols, tips for building a metadata repository and links to relevant resources and standards.

4.5 Andmekirjelduse näited

4.6 Andmekirjelduse töövahendid

Andmekataloogi tööriistad

Selliseid „andmekataloogi“ tarkvararakendusi on mitmeid ja erineva funktsionaalsusega, kuna need põhinevad erinevatel andmekirjelduse käsitlusel või mudelitel.

¹⁵ <http://www.ra.ee/arhiivihaldus/juhised/#6>

¹⁶ <http://www.ra.ee/wp-content/uploads/2016/11/LISA-1-HO-nr-51-31.10.2017-ANDMEKOGUD.pdf>

¹⁷ <http://www.ra.ee/arhiivihaldus/digitaalarhiivindus/andmekogude-arhiveerimine/>

5. Lisa 2: Andmekirjelduse standard

Andmeelemendi kirjeldus

Tabeli veergude selgitus:

- elemendi nimetus – veergude nimetused andmekirjeldustabelis
- määratlus – elemendi definitsioon (vt **Error! Reference source not found.**)
- täiendav info – viide elemendi täitmise kohustuslikkusele ja elemendi täitmise näited
- seosed – viited standarditele / skeemidele millega on loodud semantiline seos

Tabel 2: Andmeelementide kirjeldused

	Elemendi nimetus	Määratlus	Täiendav info	DDI viide
1	andmeelemendi tähis	tehniline tähis, mis võib olla täheline, numbriline või muu lühend või akronüüm	KOHUSTUSLIK Kood võib olla semantiliselt arusaadav, kuid ei pruugi seda olla. Näide: algus_kuupv, haridus, eluk_EHAK, jt28,	DDI 3.2: VariableName > String
2	andmeelemendi nimetus	on kas määratlus (ehk definitsioon), pealkiri (mis järgib pealkirjastamise reegleid) või terminoloogiline tähistus (märksõna), mis avab elemendi mõistelise sisu	KOHUSTUSLIK Sotsiaaluuringutes on mõnikord nimetus lähedane küsimusele. Näide: Küsitluse alguse kuupäev, Haridustase, Püsielukoht EHAK-tasemel	DDI 3.2: Label > Content
3	andmeelemendi kirjeldus	ühelauseline kirjeldus andmeelemendi allikast, kontekstuaalsest tähendusest, kasutusest jms	Näide: andmeelemendi „Püsielukoht EHAK-tasemel“ kirjeldus: Püsielukoht on elukoht, kus isik veedab enamiku oma igapäevasest puhke- ja uneajast. Püsielukoht loetakse olevat Eestis, kui isik on pidevalt elanud vähemalt 12 kuud enne küsitlust Eestis. EHAK on Eesti haldus- ja asustusjaotuse klassifikaator.	DDI 3.2: Description

	Elemendi nimetus	Määratlus	Täiendav info	DDI viide
4	andmeväljatüüp	Ühe tüübi valik järgmiste andmetüüpide hulgast: ¹⁸ text, numeric, decimal, date, datetime, time boolean, blob, geometry	KOHUSTUSLIK Kasutusel koos andmeväljatüübi vorminguga. Kui andmeelement on seotud koodiloendi (klassifikaatoriga), siis peab olema väljas 'seotud koodiloendi tähis' märgitud, mis koodiloendiga on tegu	DDI 3.2: VariableRepresentation TextRepresentation NumericRepresentation CodeRepresentation DateTimeRepresentation
5	andmeväljatüübi vorming	Andmeväljatüüpide kuju kirjeldus/tehniline esitus - text: tähemärkide arv - numeric: siis numbri pikkus, täisosa ja komakohtade pikkus - decimal: number, millel on märgitud komakohtade arv - date: aasta, kuu, päev eri kujul - datetime: aasta, kuu, päev, tund, minut, sekund eri kujul - time: tund, minut, sekund eri kujul - geomety: eukleidilise koordinaatsüsteemi tähis	KOHUSTUSLIK (v.a boolean ja blob) Kasutusel koos andmeväljatüübiga. - boolean (elme, et on 1 ja 0, mitte true ja false, ei ja jah/jaa)	DDI 3.2 On VariableRepresentation struktuuris Kui NumericRepresentation > NumericTypeCode Kui CodeRepresentation > CodeRepresentation > TypeOfObject Kui DateTimeRepresentation > DateTypeCode
6	seotud koodiloendi tähis	Koodiloendi või klassifikaatori andmekogus asutuse siseselt või üldiselt kokku lepitud lühend/akronüüm	KOHUSTUSLIK kui andmeväli on koodiloend või klassifikaator Näide: ER_Ravikuuritüüp, EHAK, AK2008ap, taskOutcomeSearch	DDI 3.2 CodeListName
7	andmevälja väärtuse mõõtühik	on täidetud, kui andmevälja tüüp eeldab väärtuste mõõtühiku ära näitamist	Mõõtühikud on näiteks sekund (s), meeter (m), kilogramm (kg)	DDI 3.2 MeasurementUnit

¹⁸ [SQL General Data Types](#) on toodud hulga enam andmetüüpe ning see on kirjeldus ainult SQLi vaates

	Elemendi nimetus	Määratlus	Täiendav info	DDI viide
8	andmevälja kordaja	on täidetud, kui andmevälja tüüp eeldab väärtuste mõõtühiku ära näitamist ning kordaja on kasutusel	Näitab näiteks, et element on tuhandetes tonnides, miljonites eurodes, sadades hektarites	
9	andmeelemendi väline võti	Nimest, aadressist või tähisest koosnev identifikaator, mis viitab andmeelemendile ainuliselt	Näide: URN: GUID: 123e4567-e89b-12d3-a456-426655440000	DDI 3.2: ID (on GUID) DDI 3.2: URN
10	kasutamise algus	Kuupäev, mis ajast elementi hakati andmetega väärtustama ehk andmeid selle elemendi kohta täitma	Kasutamise alguseks loetakse päeva algushetk	
11	kasutamise lõpp	Kuupäev, mis ajast elementi enam ei täideta, mille all ei mõelda mitte juhuslikku täitmist või mittetäitmist, vaid otsust elementi mitte täita		

Andmestiku kirjeldus

Siin on välja toodud olulised DCAT v2 elemendid ressursi (resource), andmekogu (Dataset) ja levituse (Distribution) kohta.

DCAT v2: <https://www.w3.org/TR/vocab-dcat-2/>

Elemendi nimetus	RDF element	Täiendav info	Andmetüüp (<i>Range</i>)
Kõigi klasside (resource, dataset, distribution) ühised elemendid			
PEALKIRI	dct:title	KOHUSTUSLIK. A name given to the item.	rdfs:Literal
KIRJELDUS	dct:description	KOHUSTUSLIK. A free-text account of the item	rdfs:Literal

Elemendi nimetus	RDF element	Täiendav info	Andmetüüp (<i>Range</i>)
IDENTIFIKAATOR	dct:identifier	KOHUSTUSLIK. A unique identifier of the item (<i>antakse automaatselt RIHA kontekstis?</i>)	rdfs:Literal
TEEMA	dcat:theme	KOHUSTUSLIK. A main category of the resource. A resource can have multiple themes.	skos:Concept
TÜÜP	dct:type	The nature or genre of the resource.	rdfs:Class
MÄRKSÕNA	dcat:keyword	A keyword or tag describing the resource.	rdfs:Literal
KEEL	dct:language	A language of the item. This refers to the natural language used for textual metadata (i.e. titles, descriptions, etc) of a cataloged resource (i.e. dataset or service) or the textual values of a dataset distribution	dct:LinguisticSystem
AUTOR	dct:creator	KOHUSTUSLIK. The entity responsible for producing the resource.	foaf:Agent
VÄLJAANDJA	dct:publisher	KOHUSTUSLIK. The entity responsible for making the item available (täidetakse automaatselt RIHAs?).	foaf:Agent
MUU SEOTUD ISIK	prov:qualifiedAttribution	Link to an Agent having some form of responsibility for the resource	prov:Attribution
AVALDAMISKUUPÄEV	dct:issued	KOHUSTUSLIK. Date of formal issuance (e.g., publication) of the item (automaatselt RIHAs).	rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2]
MUUTMISKUUPÄEV	dct:modified	KOHUSTUSLIK KUI RAKENDATAV. Most recent date on which the item was changed, updated or modified (kohustuslik kui muudetud, automaatselt RIHAs).	rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2]

Elemendi nimetus	RDF element	Täiendav info	Andmetüüp (<i>Range</i>)
KONTAKT	dcat:contactPoint	Relevant contact information for the cataloged resource. Use of vCard is recommended	vcard:Kind
VEEBILEHT	dcat:landingPage	A Web page that can be navigated to in a Web browser to gain access to the catalog, a dataset, its distributions and/or additional information.	foaf:Document
JUURDEPÄÄSUÕIGUSED	dct:accessRights	A rights statement that concerns how the resource is accessed.	dct:RightsStatement
LITSENTS	dct:license	A legal document under which the resource is made available.	dct:LicenseDocument
MUUD ÕIGUSED	dct:rights	A statement that concerns all rights not addressed with dct:license or dct:accessRights, such as copyright statements.	dct:RightsStatement
STANDARD	dct:conformsTo	An established standard to which the described cataloged resource conforms.	dct:Standard
SEOS	dct:relation	A resource with an unspecified relationship to the cataloged item	dcat:Relationship
SEOS	dcat:qualifiedRelation	Link to a description of a relationship with another resource. Used to link to another resource where the nature of the relationship is known but does not match one of the standard Dublin Core or PROV properties.	dcat:Relationship
VIIDE	dct:isReferencedBy	A related resource, such as a publication, that references, cites, or otherwise points to the cataloged resource.	dcat:Relationship
Andmekogu (Dataset) spetsiifiline kirjeldus			

Elemendi nimetus	RDF element	Täiendav info	Andmetüüp (<i>Range</i>)
LEVITUS	dcat:distribution	An available distribution of the dataset.	dcat:Distribution
RUUMILINE KAETUS	dct:spatial	The geographical area covered by the dataset.	dct:Location (A spatial region or named place)
	dcat:spatialResolutionInMeters	Minimum spatial separation resolvable in a dataset, measured in meters.	xsd:decimal
AJALINE KAETUS	dct:temporal	The temporal period that the dataset covers.	dct:PeriodOfTime (An interval of time that is named or defined by its start and end dates)
	dcat:temporalResolution	Minimum time period resolvable in the dataset.	xsd:duration
AVALDAMISE SAGEDUS	dct:accrualPeriodicity	The frequency at which dataset is published	dct:Frequency
LOOMISTEGEVUS	prov:wasGeneratedBy	An activity that generated, or provides the business context for, the creation of the dataset.	prov:Activity An activity is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.

Mõiste ja sõnastiku kirjelduselemendid

Terminite ja sõnastike kirjeldamisel lähtutakse OWL standardist.

Elemendi nimetus	OWL / RDF element	Täiendav info
Sõnastiku kirjelduselemendid (OWL: Ontology Header)		
URI	rdf:about	KOHUSTUSLIK. Sõnastiku URI (nimeruum)
NIMI	rdfs:label	KOHUSTUSLIK. Sõnastiku nimi / nimetus
KIRJELDUS / KOMMENTAAR	rdfs:comment	Pikem sõnastiku kirjeldus
EELMINE VERSIOON	owl:priorVersion	Viide sõnastiku eelmisele versioonile (URIle)
KASUTAB SÕNASTIKKU	owl:imports	Viited sõnastikus kasutatavatele teistele sõnastikele (URIdele)
Dublin Core	<owl:AnnotationProperty rdf:about="&dc;creator" />	OWL lubab lisaks eelmistele kasutada ka kõiki Dublin Core kirjelduselemente
Mõiste kirjelduselemendid (OWL: Class, Individual, Property)		
URI		KOHUSTUSLIK:
NIMI / MÕISTE		KOHUSTUSLIK:
KLASS		KOHUSTUSLIK: Viide klassile milles klass või mõiste sisaldub
KEEL		
SELGITUS		
KEHTIVUS		KOHUSTUSLIK:
LOOMISE KUUPÄEV		KOHUSTUSLIK:
MUUTMISE KUUPÄEV		
SEOS		Parent, child, sünonüüm, antonüüm, ...
MÕISTE MUUTJA		