

ANDMEHALDUSE JUHISED

Andmekirjelduse juhis

Lisa 2: Sõnastike koostamine

andmekirjeldustes. Praktiline juhis

Märts 2023

Version 0.3

Dokumendi ajalugu

ver	muutuse sisu	autor	kuupäev
0.1	Senise praktika ja alusel	Aune Suitsmart, Reana Parve, Annika Uibopuu	
0.2	Kommentaariid sisule ja struktuurile	Veiko Berendsen, Kuldar Aas	veebr 2023
0.3	Kommentaariide sisseviimine, kujundus	Veiko Berendsen	märts 2023
0.8	Saatmine juhtrühmale	Veiko Berendsen	13.03.2023
0.9	<i>Märkuste sisseviimine</i>		
1.0	<i>Lõppversioon</i>		

Kommentaariid juhise kohta on oodatud:
Statistikaamet (andmehaldus@stat.ee),
Majandus- ja Kommunikatsiooniministeerium (andmed@mkm.ee).

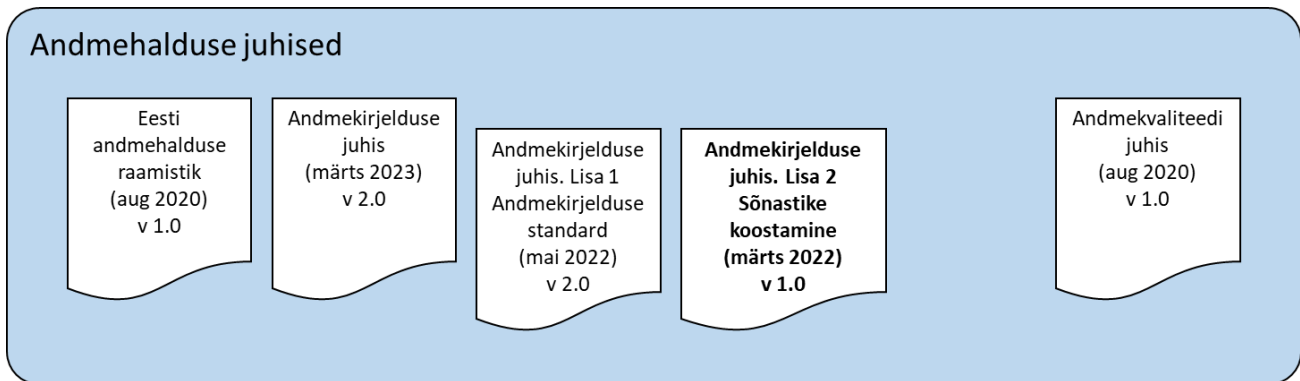
Sisukord

1	Sissejuhatus	4
1.1	Üldist	4
1.2	Sõnastike kasutamine andmekirjelduste koostamisel ja andmekataloogis	5
2	Ärisõnastiku mõiste kirjeldamine	6
2.1	Mis on ärisõnastik?.....	6
2.2	Ärisõnastiku mõisted ja nendevahelised seosed	6
2.3	Ärimõistete sisuline kirjeldamine	7
2.4	Ärimõistete vormistamine ärisõnastikus.....	10
3	Andmesõnastiku terminite kirjeldamine	13
3.1	Mis on andmesõnastik?.....	13
3.2	Andmesõnastiku termini sisuline kirjeldamine.....	13
3.3	Termini sõnastuse ühtlustamine, täiendamine ja täpsustamine.....	14
3.4	Andmesõnastiku termini vormistamine.....	15

1 Sissejuhatus

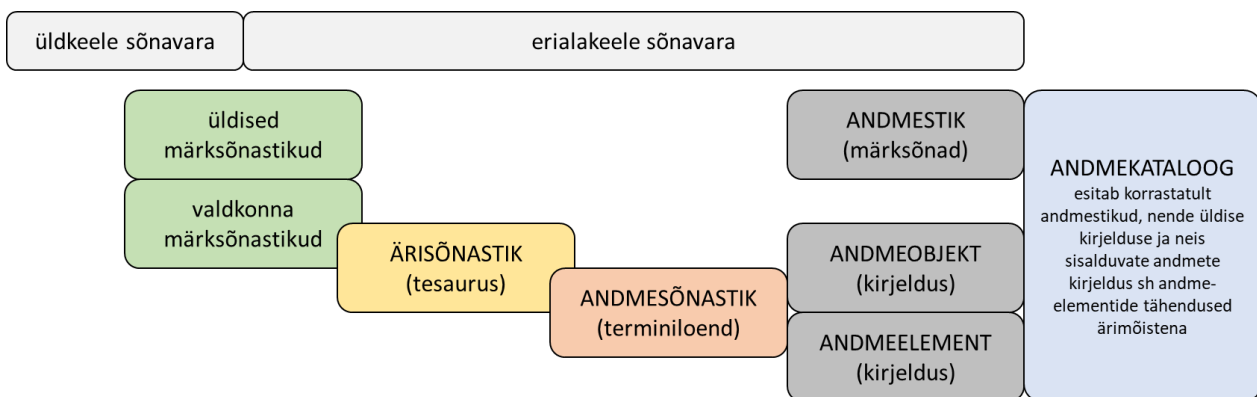
1.1 Üldist

Sõnastike koostamise praktiline juhise on osaks andmehalduse juhistest. Andmehalduse juhised (Joonis 1) on mõeldud asutustele andmehalduse sisse viimiseks ja korraldamiseks, samuti abistamiseks andmehalduse sisulisi tegevusi.



Joonis 1: Andmehalduse juhised

Juhise on koostatud lähtudes asutuste vajadustest saada sisuline ülevaade nende hallatavatest andmestikest ning nendes olevad andmed terminite ja mõistetena kirjeldada. Sõnastikud on osaks andmekirjeldusest. Juhise aluseks on terminite põhimõtted¹ terminoloogia ja sõnastike koostamise standardid² ning arusaam, et sõnastikud on praktilised sõnavara haldamise töövahendid, mis on üha enam kasutusel IT poolse andmehalduse ja valdkonna ehk äripoolse vaadete sidujana.³ Järgneval joonisel (Joonis 2) on näidatud, kuidas sõnavara korrastamine sõnastikesse on seotud andmehalduse korrastusliku vahendiga, milleks on andmekataloog.



Joonis 2: Sõnastikud andmekirjelduses

Samas on tegu praktilise juhise, mis esitab konkreetsete elementide kirjeldamise ja vormistamise nõuded. Juhise on asutuste andmekogudes kasutatavate IT ja valdkonna

¹ Erelt, T. Terminõpetus (Tartu, 2007)

² EVS-ISO 1087-1:2002 Terminoloogiatöö. Sõnastik. Osa 1: Teooria ja rakendus; ISO 25964-1:2011 Informatsioon ja dokumentatsioon. Tesaurused ja nende koostalitusvõime teiste sõnastikega. Osa 1: Infootsingu tesaurused (eesti standardiks ülevõtmisel)

³ Chaki, S. Enterprise Information Management in Practice: Managing Data and Leveraging Profits in Today's Complex Business Environment (APress, 2016). p. 115-127.

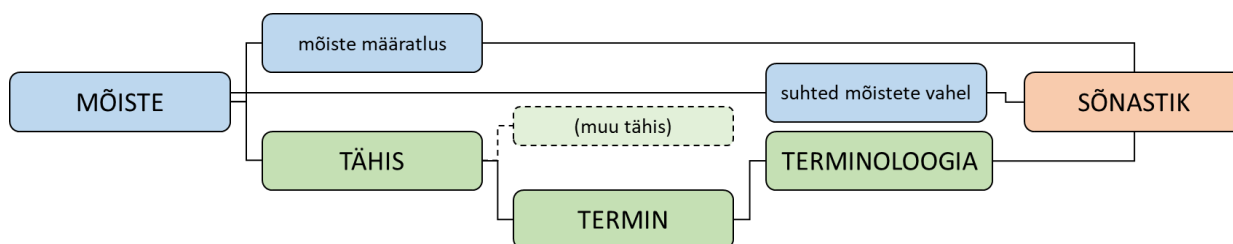
terminoloogia korrastamise ja sõnastike koostamise praktika põhjal koostanud Statistikaameti töötajad.

1.2 Sõnastike kasutamine andmekirjelduste koostamisel ja andmekataloogis

Nagu on esitatud joonisel 2 on andmekirjelduste koostamisel tegemist kahe erineva sõnastikuga. Need sõnastikud on seotud omavahel ning seotud üldiste või valdkonna märksõnastikega.

Eesti andmehalduses on põhimõte, mis lähtub nii Eesti kui rahvusvahelistest standarditest, et andmed ja andmestikud on korrastatud ja kirjeldatud andmekataloogis. Andmekataloog on kindlat tüüpi liigitusskeem. Andmekataloogis on erinevad kirjeldatavad objektid, millest olulisemad on andmestik ja andmeelement. Neile mõlemale koostatakse kirjeldus, mille üheks osaks on nende tähistamine kindlate terminitega. Andmestiku puhul on need terminid märksõnad, andmeelemendi puhul on need elemendi nimetused ning seosed teiste mõistete ja terminitega.

Sõnastikud on mõistesüsteemid, aga nad on ka terminisüsteemid ehk terminoloogiad ehk osa oskussõnavarast (Joonis 3).



Joonis 3: Sõnastikud kui terminite ja mõistete süsteemid

Andmekirjelduste koostamisel on sõnastikes sisuks olevad mõisted ja terminid nimetatud nii, et andmesõnastikus on terminid ja ärisõnastikus mõisted. Rangelt võttes tuleks lähtuda terminoloogia korrastamisel nagu see on kujutatud joonisel 3. Lihtsam on mõista igapäevatoos nii, et andmeelemendi tähised sõnastatakse andmesõnastikus terminina ning ärisõnastikus, milles on nii mõistete-terminite vahelised seosed on neid nimetatud mõisteteks.

2 Ärisõnastiku mõiste kirjeldamine

2.1 Mis on ärisõnastik?

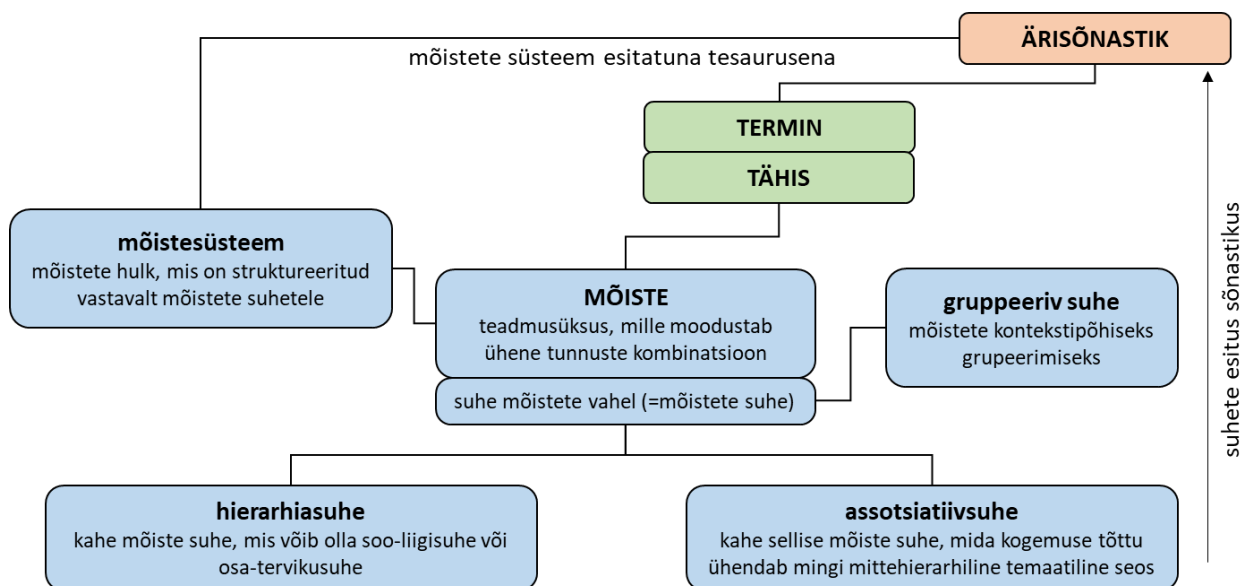
Ärisõnastik on organisatsioonis või organisatsioonideüleselt kokku lepitud sõnavara ehk mõistete ja terminite kogum, milles asutus igapäevaselt räägib või mis on kasutusel õigusaktides ja asutuse dokumentatsioonis. Sõnastik esitab organisatsiooni tegevuste ja andmete kohta piisavalt täieliku sõnavara, sisaldades nii üldistavaid kui ka detailsemaid mõisteid. Ärisõnastikke võib asutusel olla üks või mitu; nendes sisalduvad mõisted-terminid võivad pärineda nii valdkonna sõnastikest kui ka üldisest märksõnastikust.

2.2 Ärisõnastiku mõisted ja nendevahelised seosed

Ärisõnastiku mõisted on:

- põhimõiste;
- laiem mõiste;
- seotud mõiste;
- grupeeriv mõiste.

Ärisõnastiku mõisted võivad olla omavahel seostatud. Mõistetevahelisi suhteid iseloomustavad järgmised seosetüübid:



Joonis 4: Mõistetevahelised suhted ja ärisõnastik kui teaurus

- 1. hierarhiline suhe** – üks mõiste on teisest kas laiem või kitsam, nii et kitsam mõiste mahub täielikult laiemasse mõistesse. Näiteks laiem mõiste kütus on hierarhilises seoses (kitsama) mõistega biokütus või raske kütteõli; laiem mõiste isik on rolli põhjal seotud hierarhiliselt kitsama mõistega ametnik, dokumendi väljastaja, õpilane jne.
- 2. assotsiatiivsuhe** – väljendab seost semantiliselt (nt sünonüümid, alternatiivsed mõisted) või kontekstuaalselt (kasutuskonteksti kaudu) seotud mõistete vahel, mis ei ole omavahel hierarhilises suhtes. Näiteks mõiste vigastatu on seostatud mõistega liiklusõnnetus.

- 3. grupeeriv suhe** on mõistete kontekstipõhiseks grupeerimiseks (1) temaatiliselt (näiteks õigusaktist tuletatud mõiste majandustegevuse üldandmed grupeerib mõisted nagu akrediteerimine, vastutuskindlustus) või (2) vormi põhjal (näiteks dokumendi vormist tuletatud mõiste metsateatis, mis grupeerib metsateatise vormiga seotud mõisted kinnistu, puit, metsaomanik jne)

Igale andmeelemendile vastava andmesõnastiku termini kohta kirjeldatakse üks põhimõiste.

Joonisel 5 on toodud näide, kuidas on andmekirjeldus vormistatud Excelis. Andmesõnastiku termini (veerg G) kohta on kirjeldatud üks põhimõiste (eelistetermin) (veerg E). Vajaduse korral kirjeldatakse ka põhimõistest hierarhiliselt laiem mõiste (veerg D), põhimõistega seotud mõiste(d) (veerg F) ja põhimõistet grupeeriv(ad) mõiste(d) (veerud B ja C). Põhimõiste võib kuuluda mitmesse laiemasse teema ja vormiga seotud gruppi, mis tuleneb selle seostest erinevate andmeelementide / andmesõnastiku terminitega. Seotud mõistetena kirjeldatakse nii põhimõistega hierarhiliselt seotud kitsamad mõisted kui ka põhimõistega assotsiatiivsuhetes olevad mõisted.

	A	B	C	D	E	F
1	algne nr	TERMINITE GRUPEERIMINE		ÄRISÕNASTIK		
2		LAIEM GRUPP / TEEMA	VORMIGA SEOTUD GRUPP	LAIEM MÕISTE	MÕISTE	SEOTUD MÕISTE
68	66	loomeliidu liikmete muudatuse kanne	muutmiskanne	isik	loomeliidu liige	loominguline aktiivsus
69	67	loomeliidu liikmete muudatuse kanne	muutmiskanne		muutmiskanne	
70	68	loomeliidu liikmete muudatuse kanne	muutmiskanne		muutmiskanne	
71	69	loomeliidu liikmete muudatuse kanne	muutmiskanne		isik	loomeliidu liige
72	70	loomeliidu liikmete muudatuse kanne	muutmiskanne	isik	loomeliidu liige	isikukood
73	71	loomeliidu liikmete muudatuse kanne	muutmiskanne		isik	loomeliidu liige
74	72	loomeliidu liikmete muudatuse kanne	muutmiskanne		isik	loomeliidu liige

D	E	F	G	H
ÄRISÕNASTIK			ANDMESÕNASTIK	LOEND / MÕÕTÜHIK
LAIEM MÕISTE	MÕISTE	SEOTUD MÕISTE	ANDMESÕNASTIKU TERMIN	
isik	loomeliidu liige	loominguline aktiivsus	loomingulise aktiivsuse staatus muutmiskandes	JAH/EI
	muutmiskanne		loomeliidu liikme avalduse fail muutmiskandes	
	muutmiskanne		loomeliidu liikme muutmiskande identifikaator	ID
isik	isik	loomeliidu liige	isiku eesnimi muutmiskandes	
isik	loomeliidu liige	isikukood	isiku isikukood muutmiskandes	KOOD
	isik	loomeliidu liige	isiku perekonnanimi muutmiskandes	
	isik	loomeliidu liige	isiku surmakuupäev muutmiskandes	KUUPÄEV

Joonis 5: Andmekirjelduse koostamise näidis Excelis

2.3 Ärimõistete sisuline kirjeldamine

Järgnevalt on esitatud sõnastikes olevate ärimõistete praktilise kirjeldamise soovitused:

- Kirjeldama asudes võta aluseks **andmekogu põhimäärus** või muu andmekogu kehtestav õigusakt. Toetu ametlikes dokumentides kasutatud sõnavarale sh legaaldefiniitsioonidele. Kui andmekogu toimimiseks on koostatud juhiseid või metoodikadokument, kasuta nendes esitatud sõnavara.
- Ärisõnastiku põhimõisted tuletatakse andmesõnastiku terminitest, mis väljendavad andmeelementide või andmeobjektide sisu. **Põhimõiste** moodustatakse andmesõnastiku termini sellest osast, mis on kõige lähemal andmete kontekstile ja näitab, kelle või mille kohta on andmed. Näiteks andmesõnastiku termini KODULINDUDE ARV põhimõiste ärisõnastikus on KODULIND ja põhimõiste laiem mõiste LIND; andmesõnastiku termini MUNEJATE KANADE VÄLJAHAUTUD TIBUDE ARV põhimõiste on TIBU. Juhul kui andmesõnastiku termin / andmeelement sisaldab mitut konteksti, kirjeldatakse termini

ülejäanud kontekstid põhimõistega seotud mõisteteks. Antud näite puhul kirjeldatakse mõiste MUNEJA KANA mõistega TIBU seotud mõisteks. Andmesõnastiku termini DIISELRONGIVAGUNI KASUTATAVA KÜTUSE LIIK põhimõiste ärisõnastikus on DIISELRONGIVAGUN ja laiem mõiste RAUDTEEVEEREM.

- Kui **andmeelement** sisaldab mitut üksteist täiendavat konteksti, tuleks mõisted ja nendevahelised seosed andmekogu töötlejaga kokku leppida. Näiteks andmesõnastiku termini METSATEATISE, EKSPERTIISI VÕI VÄLITÖÖ OBJEKTI OMANIKU ISIKUKOOD puhul võib kirjeldada põhimõisteks METSATEATISE, EKSPERTIISI JA VÄLITÖÖ OBJEKTI OMANIK ning põhimõistega seotud mõisteteks METSATEATISE OBJEKTI OMANIK, EKSPERTIISI OBJEKTI OMANIK, VÄLITÖÖ OBJEKTI OMANIK.
- Kui andmekogu kasutab mõne teise registri/andmekogu andmeid, soovitame kasutada lähteandmekogu (põhi)mõisteid ja määratlusi (vt alltoodud näiteid).
- Kui andmed pärinevad **rahvastikuregistrist**, soovitame kasutada põhimõistet ISIK. Sel juhul kirjeldatakse andmesõnastiku termini ISIKU SUGU puhul põhimõisteks ISIK ning selle määratlus ja allikaviide võetakse rahvastikuregistri sõnastikust. Näiteks:
 - andmesõnastiku termini VEDURIJUHI ISIKUKOOD (rahvastikuregistrist pärit andmed) põhimõiste on ISIK ja põhimõistega seotud mõisted on VEDURIJUHT ja ISIKUKOOD;
 - andmesõnastiku termini SADAMATEENUSE OSUTAJA REGISTRIKOOD VÕI ISIKUKOOD (rahvastiku- ja äriregistri andmed) põhimõiste on ISIK või ETTEVÕTJA ning põhimõistega seotud mõisted SADAMATEENUSE OSUTAJA, SADAMAOPERAATOR, ISIKUKOOD, REGISTRIKOOD;
 - andmesõnastiku termini LIIKLUSÕNNETUSEGA SEOTUD ISIK ja VIGASTATU andmed on pärit rahvastikuregistrist.; sellistel juhtudel tuleb seosed andmekogude sõnastike vahel iga kirjeldatava andmekoguga eraldi kokku leppida ning määrata põhimõisted, kuna andmekogud kasutavad teiste registrite andmeid oma kontekstis.
- Kui isikuga seotud andmed ei ole pärit ühestki teisest andmekogust ja põhimõiste tähistab isiku rolli (näiteks SPORTLANE või TREENER), siis saab laiemaks mõisteks kirjeldada ISIK. Näiteks:
 - andmesõnastiku termini VEDURIJUHI ALLKIRJA NÄIDIS põhimõiste on sel juhul VEDURIJUHT ning põhimõiste laiem mõiste on ISIK.
 - Märkus: andmete teistest andmekogudest pärinemise puudumisele osutab RIHA andmekoosseisus see, kui väli „Infosüsteem” on tühi.
- Kuna ISIK võib olla ka nii füüsiliste kui juriidiliste isikute ühiseks laiemaks mõisteks, siis tuleks nende mõistete kasutamine asutusega kooskõlastada.
- **Äriregistrist** pärit andmete puhul soovitame kasutada põhimõisteid ETTEVÕTJA, AVALIK-ÕIGUSLIK JURIIDILINE ISIK, ASUTUS, ÄRIÜHING, firma ehk ÄRINIMI jt äriseadustikus ja tsiviilseadustiku üldosa seaduses toodud mõisteid. Kasutatavad mõisted tuleks kirjeldatava andmekoguga kokku leppida.

- Mõistete kirjeldamine **aadressiandmetele** juhul, kui aadressiandmed on pärit aadressiandmete süsteemi infosüsteemist (ADSi infosüsteem). Maa-amet soovib kirjeldamisel kasutada aadressiandmete süsteemi mõisteid ja tunnuseid „[ADSiga liidestumise juhendist](#)”, mille värskem dokument on kättesaadav [Geoportaalist](#). Juhendi peatükk „Põhisüsteemi mõisted” sisaldab nii mõisteid kui ka nendega seotud tunnuseid. Näiteks mõiste on ADRESSIOBJEKT ja selle tunnused on ADOB_LIIK, ADS_OID jne. Maa-amet soovib mõistete määratlusi andmekirjelduste koostamise failis mitte dubleerida, vaid viidata nende allikale. „Aadressiandmete käsiraamatule“ soovitatakse mitte viidata.
- Mõistete kirjeldamine aadressiandmetele juhul, kui aadressiandmed ei ole pärit ADSist peaks toimuma järgmiselt. Erinevate aadressikomponentide kohta on kasutusel ühine mõiste ADRESSIANDMED, et ei peaks iga detaili eraldi mõisteks kirjeldama. Ärisõnastiku mõiste aadressiandmed katab kõik asukohaga seotud andmed, nagu nt registreerimis-, juriidiline, asukohta, toimumiskoha aadress jm asukohta tähistavad andmed, sh sihtnumbrid, koordinaadid. Näiteks andmesõnastiku termini ÕPPEASUTUSE JURIIDILINE ADRESS puhul on põhimõisteks ÕPPEASUTUS ning seotud mõisteks ADRESSIANDMED.
- Isiku või ettevõtte/organisatsiooni **kontaktandmete** kirjeldamiseks (seotud mõistena) soovime kasutada ühist mõistet KONTAKTANDMED, et kõiki selliseid andmeelemente (näiteks isiku e-posti aadress või isiku kontakttelefon) ei peaks eraldi mõisteks kirjeldama. Ärisõnastiku mõiste KONTAKTANDMED hõlmab kõiki isiku või organisatsiooniga seotud kontaktandmeid, mis ei ole aadressiandmed, nagu TELEFONINUMBER, E-POSTI ADRESS, FAKSINUMBER, VEEBISAIDI ADRESS, SOTSIAALMEEDIA KONTO. Näiteks andmesõnastiku termini ISIKU E-POSTI ADRESS puhul on põhimõisteks ISIK ning seotud mõisteks KONTAKTANDMED.
- Üks ja sama mõiste võib esineda nii ärisõnastiku põhimõistena kui ka hierarhiliselt laiema mõistena sõltuvalt kontekstist. Näiteks laiem mõiste KÜTUS on hierarhiliselt seotud mõistega BIOKÜTUS. Kui andmed on biokütuse kohta, siis kirjeldatakse põhimõisteks biokütus ning luuakse sellele mõistele seos laiema mõistega kütus. Kui andmed on kütuse kohta, siis kirjeldatakse põhimõisteks kütus.
- **Tegevused**, mis on seotud andmekogu kontekstis oluliste objektide / sündmuste / protsesside staatuse või olekuga, võib formuleerida ärisõnastiku mõisteks. Näiteks võivad kindla andmekogu kontekstis olla põhimõisteks: LIIKMEKS ASTUMINE, LIIKMESUSE PEATAMINE, JÄÄTMETE KÄITLEMINE.
- Tegevused, mis on seotud sündmuste dokumenteerimisega või menetlemisega, ei moodusta üldreeglina ärisõnastiku mõisteid. Näiteks: dokumendi registreerimine, kinnitamine, allkirjastamine jne.
- Üldjuhul ei ole ärisõnastiku mõisted ka põhimõistet iseloomustavad üldised kirjeldused, tunnused, omadused ega üldlevinud mõisted. Näiteks mass, liik, tüüp, vanus, arv, nimi, sugu, värvus, tunnus, ajad/kuupäevad/kellaajad jmt.

2.4 Ärimõistete vormistamine ärisõnastikus

Ärisõnastiku ja andmesõnastiku mõistete ja terminite kirjelduselemendid põhinevad andmekirjelduse standardil.⁴ Ärisõnastiku mõistekirje koosneb alltoodud kirjelduselementidest.

Ärimõiste kirjelduselement TERMIN

- Ärimõistet tähistav termin sõnastatakse üldjuhul ainsuses. Konteksti ja vormi põhjal loodud terminigruppides võib vajaduse korral kasutada ka mitmuse vormi, kui grupeeriv termin rühmitab mõisteid ühiste tunnuste põhjal.
- Termin kirjutatakse väikese algustähega, välja arvatud juhul, kui algustähereegel nõuab suure algustähe kasutamist.
- Termin sõnastatakse eesti ja inglise keeles.
 - Terminis võib kasutada üldtuntud või andmekogu oskuskeeles käibel olevaid akronüüme. Näiteks: EPS-põld, MAH, EHAK, EMTAK. Akronüümide kasutamisel peaks lähtuma andmekogu oskussõnavarast. Kitsalt andmekogu kontekstis kasutusel olevad akronüümid, mis ei ole üldlevinud, soovitame ka lahti kirjutada, mõlemad mõisteteks kirjeldada ning omavahel siduda. Võimalusi on erinevaid: (1) VÕTA; Varasemate õpingute ja töökogemuse arvestamine; (2) Varasemate õpingute ja töökogemuse arvestamine (VÕTA); (3) VÕTA (Varasemate õpingute ja töökogemuse arvestamine).

Ärimõiste kirjelduselement MÄÄRATLUS

- Kirjelduselement sisaldab kas mõiste **määratlust** ehk definitsiooni või mõiste sisu avavat **selgitust**. Määratlus on lühike ja ühemõtteline mõistet kirjeldav ning kindlate vormistusreeglite järgi koostatud lausung, milles esitatakse definitsiooni tavatunnused.
 - Määratluste tüübid leiab Eesti Keele Instituudi „EKI teatmiku“ veebilehelt [Definitsiooni tüübid](#).
 - Mõistete määratlustes tuleks eelistada õigusaktidest pärit legaaldefinitsioone, kui need on olemas.
 - Määratluse puudumisel või selle mõistmise hõlbustamiseks võib lisada selgituse. Selgitus võib sisaldada termini vabas vormis kirjeldust, täpsustavaid ja/või täiendavaid kommentaare, kasutuskonteksti vm mõistet avavat lisateavet, mis põhineb õigusaktidel või andmekogu dokumentatsioonil. Nende puudumisel võib kokkuleppel asutusega kasutada ka muudest terminiallikatest pärit määratlusi (näiteks üld- või oskuskeele sõnastikud, terminibaasid).
 - Määratlusi võib ka ise koostada, kasutades erinevaid terminiallikaid nagu õigusaktid, andmekogu tehniline dokumentatsioon, juhised jms. Määratlust täiendavad ja selgitavad kommentaarid ning märkused võib lisada kirjelduselemendi märkused väljale.

⁴ [Andmekirjelduse juhise Lisa 1 andmekirjelduse standard. Ver 2.0 \(Statistikaamet, 2022\)](#)

- Mõiste määratlus algab üldjuhul väikese algustähega ning lõpeb ilma punktita. Kui mõistet kirjeldav selgitus koosneb ühest või mitmest täislauselst, siis kirjutatakse see suure algustähe ja punktiga lause lõpus:

mõiste	määratlus / selgitus
RATSUTAMISHALL	statsionaarne ruum ratsaspordiga tegelemiseks
EHITIS	inimtegevuse tulemusel loodud ja aluspinnasega ühendatud või sellele toetuv asi, mille kasutamise otstarve, eesmärk, kasutamise viis või kestvus võimaldab seda eristada teistest asjadest. Ehitised jagunevad hooneteks või rajatisteks.
SEIN	vertikaalne konstruktsioon, mis piirab või jagab ruumi ja täidab tavaliselt kande- või tugifunktsiooni
KINNISTU NUMBER	Kinnistusraamatusse kantakse kõik kinnisasjad, kui seaduses ei ole sätestatud teisiti. Iga kinnistusraamatusse kantud kinnisasja kohta avatakse iseseisev registriosa ja sellele antakse eraldi number (kinnistu number).

- Mõistet tähistavat terminit määratluse alguses üldjuhul ei korrata. Näiteks mitte ehitis on aluspinnasega kohtkindlalt ühendatud ja inimtegevuse tulemusena ehitatud terviklik asi, vaid aluspinnasega kohtkindlalt ühendatud ja inimtegevuse tulemusena ehitatud terviklik asi.
- Mõiste määratlus sõnastatakse eesti ja inglise keeles. Ühe mõistekirje eesti- ja ingliskeelne määratlus peavad omavahel sisuliselt kattuma ning pärinema samast allikast.
- Lühendatud sõnad kirjutatakse lahti. Näiteks: nr – number, tuh – tuhat, ik – isikukood, tn – tänav, kp – kuupäev, a – aasta, mln – miljon, mld – miljard.
- Üldlevinud akronüüme võib määratlustes kasutada. Näitesks: KOV, EHAK, ÜRO jt.

Ärimõiste kirjelduselement ALLIKAS

- Kirjelduselement allikas sisaldab **Riigi Teatajale** viidates linki õigusaktile. Õigusakti paragrahvi, lõiget ja punkti ei täpsustata. Üldjuhul viidatakse kirjeldamise hetkel kehtiva õigusakti terviktekstile. Lisatakse viide avaldamismärkele. Näiteks: Rahvastikuregistri seadus (RT I, 10.02.2023, 9)
- Lisaks õigusaktile võib mõiste allikaks olla ka:
 - Üldkeelesõnastik või oskussõnastik; näiteks [Eesti märksõnastik](#) (EMS) või mõni oskusõnastik nagu [Mitmekeelne Demograafiasõnastik](#) või [Kosmoseterminoloogia](#). Täpsem nimekiri on Eesti Keele Instituudi [terminiallikate](#) veebilehel. Allikaviiteks on viide sõnastikule.
 - Klassifikaator või klassifikaatori element: allikaviide on [Statistikaameti klassifikaatorite portaalis](#) avaldatud klassifikaator, näiteks [Haridus- ja koolitusvaldkondade liigitus](#); samuti võib allikaks olla mõni rahvusvaheline portaal

nagu [RAMON](#) (Reference And Management Of Nomenclatures) ja selles avaldatud standard, näiteks PRODCOM List 2021. Allikaviiteks on viide klassifikaatorile.

- Asutuse koduleht, andmekoguga seotud dokument või (kasutus)juhend. Allikaviiteks on teabeallikale võimalikult täpsel tasandil.
 - Juhend, väljaanne vms, millele ei ole elektroonilist viidet. Allikaviiteks on allika pealkiri ja ilmumisandmed.
 - Enda asutuse koostanud mõiste määratluse korral on allikaviiteks asutuse nimi.
- Õigusaktidest või asutuse dokumentatsioonist pärit grupeerivate terminite puhul on soovitatav lisada allikaviide vastavale dokumendile ka siis, kui terminil määratlus puudub.

Ärimõiste kirjelduselement MÄRKUSED

- Mõiste märkused on lühike mõistet avav lause, lõik või pikem selgitus. Kõik täiendavad kommentaarid mõiste määratluse, selgituse, koostamise või kasutamise kohta võib lisada märkustesse. Märkused vormistatakse suure algustähega, lause lõpus on punkt.
- Mõiste märkused sõnastatakse ainult eesti keeles.
- Lühendatud sõnad kirjutatakse lahti. Näiteks: number, tänav, kuupäev, aasta, miljon, miljard.
- Märkustes võib kasutada üldtuntud või andmekoguga oskuskeeles käibel olevaid akronüüme. Näiteks: EPS-põld, MAH, EHAK, EMTAK, ÜRO.

3 Andmesõnastiku terminite kirjeldamine

3.1 Mis on andmesõnastik?

Andmesõnastik on ühelt poolt terminite loetelu ning teiselt poolt esitab terminitena andmeelementide sisu. Andmeelement võib olla tähistatud selle sisu väljendava inimloetava sõna või fraasiga, aga selleks võib olla ka akronüüm, lühend või muu sõne või tähis, mis on masinloetav ega pole lisaselgituseta mõistetav. Andmesõnastiku terminid seotakse vahetult andmeelemendiga. Andmeelemendil on tähis ja üldjuhul ka kirjeldus, andmesõnastiku terminil on nimetus ja määratlus.

3.2 Andmesõnastiku termini sisuline kirjeldamine

Järgnevalt on esitatud andmesõnastikes olevate terminite praktilise kirjeldamise soovitused koos ärimõistetega sidumisega.

- Kirjeldama asudes toetu **andmekogu põhimääruses**, metoodikadokumentides jm ametlikes dokumentides kasutatud sõnavarale.
- Andmesõnastiku **termin** peaks olema andmeelemendi konteksti võimalikult täpselt avav, vajaduse korral mitmest sõnast koosnev liittermin või terminifraas. Näiteks: MÜÜDUD KÜTUSE KOGUS, SPORTLASESTIPENDIUMI SAAJA EES- JA PEREKONNANIMI, SPORDIOBJEKTI RIIETUSRUMIDE ARV.
- **Ajaga** seotud terminid soovitame sõnastada võimalikult täpselt, et terminist selguks, kas tegu on näiteks aasta, kuupäeva ja/või kellaajaga. Näiteks: ORIENTEERUMISKAARDI ESMASE KASUTUSELEVÕTU AASTA, TREENERIKUTSE KEHTIVUSE ALGUSKUUPÄEV, VANGLASSE SAABUMISE KUUPÄEV JA KELLAAEG.
- Andmesõnastiku terminid soovitame võimalikult detailselt lahti kirjutada, näiteks: E-POSTI ADDRESS (mitte e-post), VEEBILEHE ADDRESS (mitte veebileht).
- Andmesõnastiku **termin seotakse mõistega**, mis on kõige lähemal andmete kontekstile ja näitab, kelle või mille kohta on andmed. Kontekste võib olla mitu (vt ka ärimõistete sisulise kirjeldamise alateemat).
- Juhul kui andmeelement sisaldab mitut üksteist täiendavat konteksti, näiteks andmed on pärit erinevatest plankidest, vormidest või dokumendiliikidest, on vaja tähelepanu pöörata nii dokumendi koostamise eesmärgile kui ka andmeelemendi kontekstile. Näiteks: andmeelemendis METSATEATISE, EKSPERTIISI VÕI VÄLITÖÖ OBJEKTI OMANIKU ISIKUKOOD on andmed omaniku kohta, temaga seotud täiendavad kontekstid on METSATEATISE OBJEKT, EKSPERTIISI OBJEKT ja VÄLITÖÖ OBJEKT. Sellisel juhul tuleb andmekogu dokumentatsiooni põhjal või analüüsidest andmemudelit ning andmekoguga läbi rääkides, kindlaks teha, kas tegemist on ühe andmeelemendiga või kolme erinevaga. Kui kolm konteksti on kirjeldatud üheks andmeelemendiks, siis kirjeldatakse sellele üks andmesõnastiku termin. Kui mitte, siis kirjeldatakse kolm eraldi andmeelementi ja kolm andmesõnastiku terminit. Viimasel

juhul kirjeldatakse andmesõnastiku terminid: (1) METSATEATISE OBJEKTI OMANIKU ISIKUKOOD, (2) EKSPERTIISI OBJEKTI OMANIKU ISIKUKOOD ja (3) VÄLITÖÖ OBJEKTI OMANIKU ISIKUKOOD ning seotakse need vastava andmeelemendiga: (1) metsateatise objekti omaniku isikukood, (2) ekspertiisi objekti omaniku isikukood ja (3) välitöö objekti omaniku isikukood.

- Kui andmeelemendil on mitu konteksti ning vormistamine tehakse Excelis, siis lisatakse igale terminile/andmeelemendile uus rida.

3.3 Terminisõnastuse ühtlustamine, täiendamine ja täpsustamine

Andmesõnastiku termini sõnastust ühtlustatakse alljärgnevatel juhtudel.

- Andmeelemendi kirjeldus ei ava konteksti, näiteks andmeelement on KÕRGUS ning pole täpsustatud, kelle/mille kõrgus.
- Andmesõnastiku terminit täiendatakse, kui dokumentatsioonist selgub termini sisu täpsustav või täiendav kontekst. Näiteks: andmeobjektile ESINDUSÕIGUSE KEHTIVUSE ALGUS lisatakse terminis SPORDIORGANISATSIOONI ESINDAJA termin ESINDUSÕIGUSE KEHTIVUSE ALGUSKUUPÄEV.
- Andmekoosseisu kirjelduse või asutuse/andmekogu dokumentatsiooni põhjal võib kindlalt järeldada, et andmeelemendi nimetus ei väljenda selle korrektset sisu. Näiteks: andmeelement KÕRGUS tähistab hoopis KESKMIST KÕRGUST. Konteksti keskmine kõrgus siiski edasi ei anna, seega tuleb andmesõnastiku terminiks sel juhul kirjeldada PUISTU KESKMINE KÕRGUS.
- Andmesõnastiku terminit võib vajaduse korral lühendada, kui:
 - andmeelement sisaldab õigusaktist või vormist tulenevat konteksti; näiteks andmeelemendi puhul: ANKEETI ISELOOMUSTAV TUNNUS PANGAKONTO NUMBER piisab terminist PANGAKONTO NUMBER, andmete päritolu vormi, milleks on ankeet, andmesõnastiku terminis ei ole vaja täpsustada;
 - andmeelement sisaldab lisatäpsustust andmete kohta; näiteks andmeelemendi kirjeldus on ETTEVÖTTE MÜÜGI ANDMED: müüdnud kauba kogus. Sellisel juhul võib kirjeldada andmesõnastiku termini lühemaks, piisab liitterminist MÜÜDUD KAUBA KOGUS.
- Kui **aadressiandmed** on pärit aadressiandmete süsteemi infosüsteemist (ADSi infosüsteem), soovib Maa-amet kirjeldamisel kasutada ADSi mõisteid ja tunnuseid „[ADSi liidestumise juhendist](#)“, mille värskem dokument on kättesaadav [Geoportaalist](#). Juhendi peatükk „Põhisüsteemi mõisted“ sisaldab nii mõisteid kui ka nendega seotud tunnuseid. Näiteks mõiste AADRESSIOBJEKT tunnused on ADOB_LIIK, ADS_OID jne. Maa-amet soovib kasutada andmesõnastiku terminite sõnastamiseks tunnuste kirjeldusi, näiteks kui andmeelement on ADOB_LIIK, siis sellele vastav andmesõnastiku termin on ADSi spetsifikatsiooni järgi AADRESSIOBJEKTI LIIGI KOOD. Kui tunnuse kirjeldus koosneb mitmest lausest, tuleb seda lühendada andmesõnastiku terminile vastavaks.

3.4 Andmesõnastiku termini vormistamine

Andmesõnastiku terminite kirjelduselemendid põhinevad andmekirjelduse standardil.⁵ Andmesõnastiku terminikirje koosneb alltoodud kirjelduselementidest

Andmesõnastiku termini kirjelduselement TERMIN

- Termin kirjutatakse väikese algustähega, välja arvatud juhul, kui algustähereegel nõuab suure algustähe kasutamist.
- Andmesõnastiku termin sõnastatakse ainult eesti keeles, seevastu ärimõisted nii eesti kui inglise keeles.
- Andmesõnastiku termini kirjeldamisel välditakse kõrvallauseid, pikkades terminifraasides võib selle erinevad osad eristada mõttekriipsuga (Excelis sümbol – (EN Dash)).
- Üldjuhul terminis mõõtühikut ei kirjelda. Kui see on RIHA andmeobjekti kommentaaris olemas, siis on soovitatav see sinna alles jätta. Näiteks kui RIHA andmeobjekt (sellele vastab siinse andmekirjelduse juhise kontekstis andmeelement) on FOSSIILSE KÜTUSE KASVUHOONEGAASIDE (KHG) MAHUKUS (GCO2EKV/MJ), siis temaga seotud andmesõnastiku termin on samuti FOSSIILSE KÜTUSE KASVUHOONEGAASIDE (KHG) MAHUKUS (GCO2EKV/MJ).
- Kirjavahemärkidest kasutatakse pikkades terminifraasides ainult mõttekriipsu ja vajaduse korral ka koma; ei kasutata lauselõpumärke ega muid kirjavahemärke, nagu näiteks: hüüumärk, küsimärk, kaldkriips, koolon, mõttepunktid jt (? . ! / : ...).
- Lühendatud sõnad kirjutatakse terminis üldjuhul lahti (näiteks: *number, tänav, kuupäev, aasta*).
- Terminis võib kasutada üldtuntud või andmekogu oskuskeeles käibel olevaid akronüüme, nt EPS-põld, MAH, ÜRO, EHAK, EMTAK. Akronüümide kasutamisel peaks alati lähtuma asutuses käibel olevast oskussõnavarast.

Andmesõnastiku termini kirjelduselement MÄÄRATLUS

- Termini määratlus algab väikese algustähega ning lõpeb ilma punktita. Kui terminit kirjeldav selgitus koosneb täislause(te)st, siis kirjutatakse see suure algustähe ja punktiga lause lõpus.
- Lühendatud sõnad kirjutatakse lahti (näiteks: *number, tänav, kuupäev, aasta, telefoninumber, miljon, miljard*).
- Üldlevinud akronüüme võib andmesõnastiku määratlustes kasutada (näiteks: KOV, EHAK, ÜRO, EU, NATO).

⁵ [Andmekirjelduse juhise. Lisa 1 andmekirjelduse standard. Ver 2.0 \(Statistikaamet, 2022\)](#)

- Kas termini määratlus võib olla pärit üldisest sõnastikust ja/või märksõnastikust või entsüklopeediast sh wikipediast? Üldjuhul mitte, sest andmekogu kontekstis võib andmete tähendus erineda üldlevinud määratlustest.
- Kui on olemas Eesti standard, milles termin on määratletud, tuleb eelistada seda määratlust.

Andmesõnastiku termini kirjelduselement MÄRKUSED

- Mõiste märkused kirjutatakse suure algustähega, lause lõpus on punkt.
- Üldlevinud akronüüme võib andmesõnastikus kasutada; näiteks: KOV, EHAK, ÜRO.

Viited

Andmekirjelduse juhised. Lisa 1: Andmekirjelduse standard. Ver 2.0. Statistikaamet, 2022.

ADS-iga liidestumise juhend. Maa-amet, 2022

Chaki, S. Enterprise Information Management in Practice: Managing Data and Leveraging Profits in Today's Complex Business Environment (APress, 2016)

EVS-ISO 1087-1:2002 Terminoloogiatöö. Sõnastik. Osa 1: Teooria ja rakendus.

ISO 25964-1:2011 Informatsioon ja dokumentatsioon. Tesaaurused ja nende koostalitusvõime teiste sõnastikega. Osa 1: Infootsingu tesaaurused

Erelt, Tiiu. Terminiõpetus. (Tartu, 2007)

ISO 25964-1:2011 Informatsioon ja dokumentatsioon. Tesaaurused ja nende koostalitusvõime teiste sõnastikega. Osa 1: Infootsingu tesaaurused (eesti standardiks ülevõtmisel)