

ANDMEHALDUSE JUHISED

Andmekirjelduse juhis

Lisa 2: Sõnastike koostamine andmekirjeldustes. Praktiline juhis

Juuni 2024

Versioon 0.9

Dokumendi ajalugu

ver	muutuse sisu	autor	kuupäev
0.1	Senise praktika alusel	Aune Suitsmart, Reana Parve, Annika Uibopuu	
0.2	Kommentaari sisule ja struktuurile	Veiko Berendsen, Kuldar Aas	veebr 2023
0.3	Kommentaari sisseviimine, kujundus	Veiko Berendsen	märts 2023
0.8	Saatmine juhtrühmale	Veiko Berendsen	13.03.2023
0.9	Andmekirjelduse standardi versiooniga 3.0 vastavusse viimine	Piret Heinsaar, Aune Suitsmart, Reana Parve	06.06.2024
1.0	Lõppversioon		

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

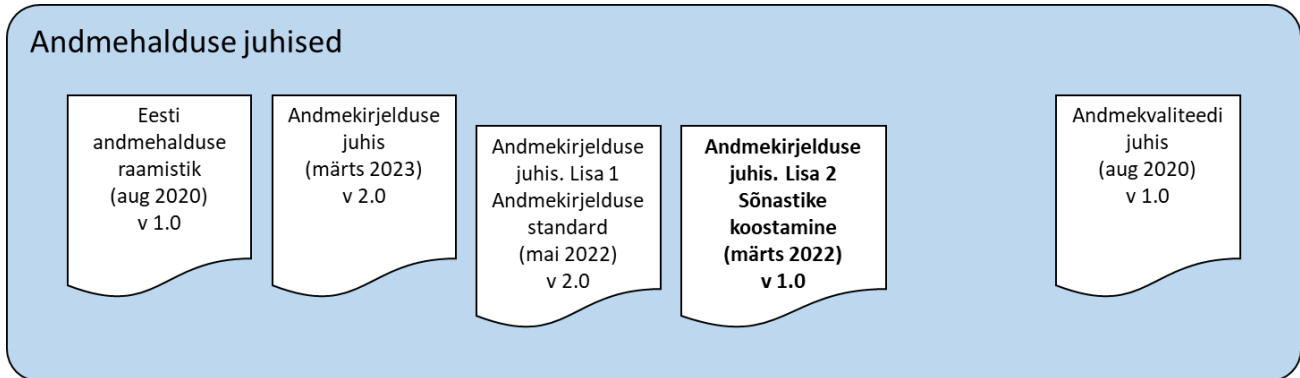
Sisukord

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	8
2.4	Ärimõistete vormistamine ärisõnastikus	9
3	Andmesõnastiku terminite kirjeldamine	12
3.1	Mis on andmesõnastik?	12
3.2	Andmesõnastiku termini sisuline kirjeldamine	12
3.3	Termini sõnastuse ühtlustamine, täiendamine ja täpsustamine	13
3.4	Andmesõnastiku termini vormistamine	13
4	Näited	15
	Ärimõistete sisulise kirjeldamise näited	15
	Andmesõnastiku termini sisulise kirjeldamise näited	17
	Kasutatud allikad	19

1 Sissejuhatus

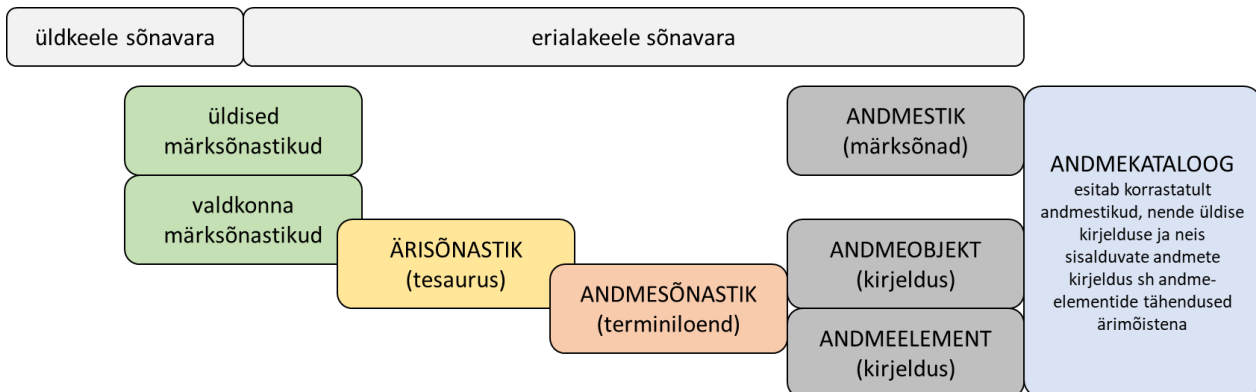
1.1 Üldist

Sõnastike koostamise praktiline juhised on osaks andmehalduse juhistest. Andmehalduse juhised (joonis 1) on mõeldud asutustele andmehalduse sisseviimiseks ja korraldamiseks, samuti abistamiseks andmehalduse sisuliste tegevustega.



Joonis 1: Andmehalduse juhised

Juhised on koostatud, lähtudes asutuste vajadustest saada sisuline ülevaade nende hallatavatest andmestikest ning kirjeldada nendes olevad andmed terminite ja mõistetena. Sõnastikud on osaks andmekirjeldusest. Juhise aluseks on terminitöö 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 sidujana³. Järgneval joonisel (joonis 2) on näidatud, kuidas sõnavara korrastamine sõnastikesse on seotud andmehalduse korrastusliku vahendiga, milleks on andmekataloog.



Joonis 2: Sõnastikud andmekirjelduses

¹ Tiiu Erelt, *Terminiõpetus* (Tartu, 2007).

² EVS-ISO 1087-1:2002 *Terminoloogiatöö. Sõnastik: Osa 1. Teooria ja rakendus = Terminology work. Vocabulary: Part 1. Theory and application*;
EVS-ISO 25964-1:2023 *Informatsioon ja dokumentatsioon. Tesaurused ja nende koostalitusvõime teiste sõnastikega. Osa 1: Infootsingu tesaurused*;
EVS-ISO 25964-2:2023 *Informatsioon ja dokumentatsioon. Tesaurused ja nende koostalitusvõime teiste sõnastikega. Osa 2: Koostalitusvõime teiste sõnastikega*

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

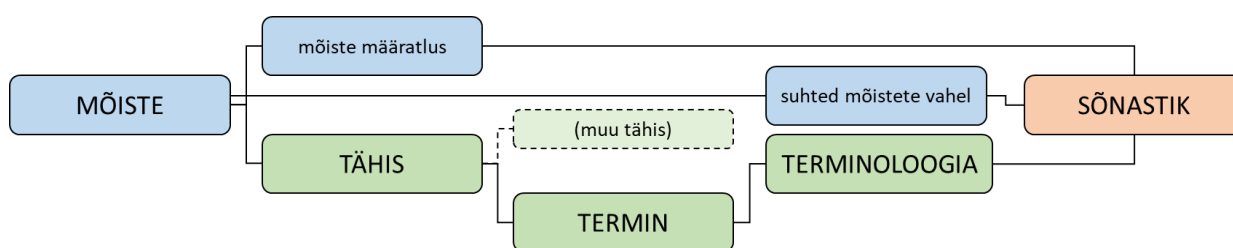
Samas on tegu praktilise juhisega, mis esitab konkreetsete elementide kirjeldamise ja vormistamise nõuded. Juhise on koostanud asutuste andmekogudes kasutatavate IT ja valdkonna terminoloogia korrastamise ja sõnastike koostamise praktika põhjal Statistikaameti töötajad.

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 ka üldiste või valdkonna märksõnastikega.

Eesti andmehalduses kehtib nii Eesti kui rahvusvahelistest standarditest lähtuv põhimõte, et andmed ja andmestikud on korrastatud ja kirjeldatud andmekataloogis. Andmekataloog on kindlat tüüpi liigitusskeem. Andmekataloog sisaldab erinevaid kirjeldatavaid objekte, millest olulisemad on andmestik ja andmeelement. 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 sõnastike sisuks mõisted ja terminid: andmesõnastiku puhul nimetatakse neid terminiteks ja ärisõnastikus on mõisted. Rangelt võttes tuleks lähtuda terminoloogia korrastamisel joonisel 3 kujutatud süsteemist. Igapäevatoos mõistetakse seda nii, et andmeelemendi tähised sõnastatakse andmesõnastikus terminitena ning ärisõnastikus, mis sisaldab ka mõistete ja terminite vahelisi seoseid, nimetatakse neid mõisteteks.

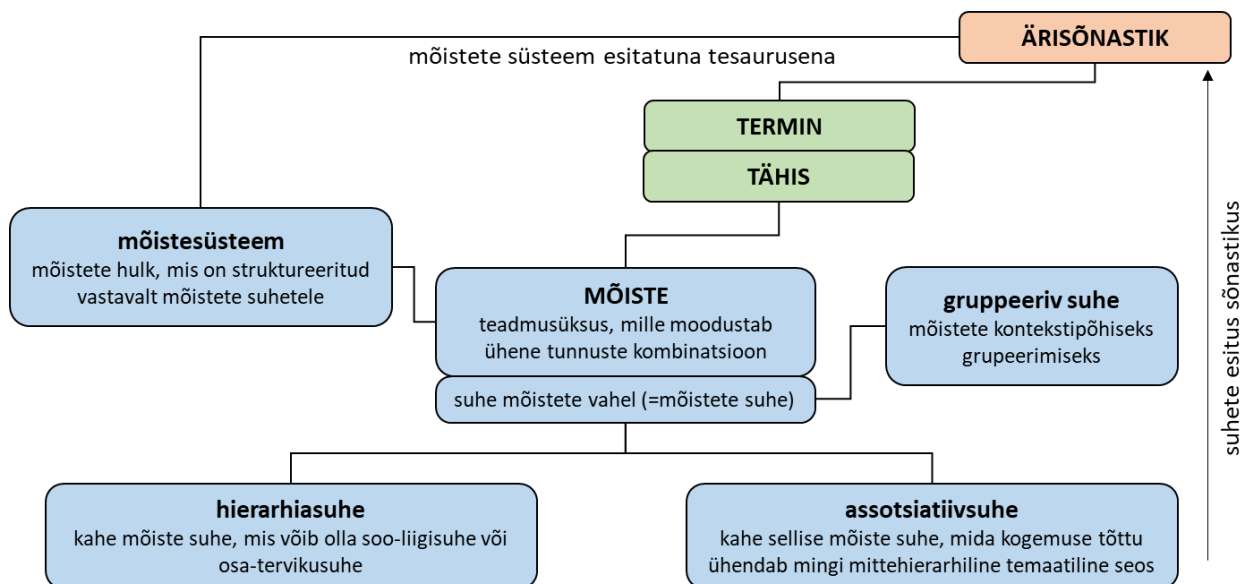
2 Ärisõnastiku mõiste kirjeldamine

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 üldisemaid kui ka detailsemaid mõisteid. Ärisõnastikke võib asutusel olla üks või mitu; nendes sisalduvad mõisted-terminid võivad pärineda nii valdkonna oskussõnastikest kui ka üldisest märksõnastikust.

2.2 Ärisõnastiku mõisted ja nendevahelised seosed

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



Joonis 4: Mõistetevahelised suhted ja ärisõnastik kui tesaurus

- 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*.
 - a. hierarhiasuhte kontrollimiseks saab teha *all-and-some* testi: *auto* on kitsam mõiste kui *sõiduk*, kuna kõik autod on sõidukid ja mõned sõidukid on autod;
 - b. hierarhiasuhte määramisel saab toetuda definitsioonidele, kus üks mõiste defineeritakse teise kaudu, nt mõistete *sõiduk*, *mootorsõiduk*, *auto* ja *sõiduauto* määratlused liikluseaduses näitavad nendevahelisi hierarhilisi suhteid. Olgu siin näiteks toodud vaid määratluste algused: *sõiduauto on sõitjate vedamiseks ettenähtud auto, milles...;* *auto on sõitjate või veose veoks või sõidukite haakes vedamiseks või eritööde tegemiseks ettenähtud vähemalt neljarattaline mootorsõiduk, mille...;* *mootorsõiduk on mootori jõul liikuv sõiduk, välja arvatud ...*
- 2. assotsiatiivsuhe** – väljendab seost semantiliselt või kontekstuaalselt (kasutuskonteksti kaudu) seotud mõistete vahel, mis ei ole omavahel hierarhilises suhtes. Näiteks mõiste *vigastatu* on seostatud mõistega *liiklusõnnetus*.
 - a. assotsiatiivsuhtes on mõisted, mille tähendusulatused osaliselt kattuvad, nt *paat* ja *laev*; protsess või tegevus ja selle tegija või vahend, nt *kuritegude uurimine* ja

kriminaaluuri; tegevus ja selle sihtmärk või saaja, nt *juhiloa taotlemine* ja *juhiluba*; objekt või materjal ja selle omadused, nt *mürgistus* ja *mürkaine*; põhjusliku seosega mõisted, nt *haigus* ja *patogeen*; objekt või protsess ja selle vastureaktsioon, nt *põletik* ja *põletikuvastane aine*; liittermin ja selle põhisõna (kui nende vahel puudub hierarhiasuhe), nt *mudellaev* ja *laev*;⁴

- b. sageli on üks mõistetest vajalik teise määratlemisel või selgitamisel, nt mõiste *hoone köetav pind* määratlus on *hoone kõigi sisekliima tagamisega ruumide suletud netopindade summa* ning on assotsiatiivsuhtes mõistega *sisekliima tagamisega ruum*;
- c. seoste tegemisel on kasu, kui mõisted on ärisõnastikus määratletud. Näiteks energiamärgise puhul võib ühel juhul määratlus olla seotud kodumasinatega ja teisel juhul hoopis hoonete vastava märgisega. Kui ärisõnastikus on mõiste *energiamärgis* puhul selgitatud, et energiamärgis annab infot projekteeritava või olemasoleva hoone projekteeritud energiavajaduse või tegeliku energiatarbimise kohta, siis kehtib ärisõnastikus mõistete *energiamärgis* ja *hoone* vahel assotsiatiivsuhe;

3. grupeeriv suhe on mõistete kontekstipõhiseks grupeerimiseks (1) temaatiliselt (näiteks mõiste *kindlustus* grupeerib mõisted nagu *kindlustusvõtja*, *kindlustusleping*, *kindlustusperiood*) või (2) vormi põhjal (näiteks dokumendi vormist tuletatud mõiste *metsateatis*, mis grupeerib metsateatise vormiga seotud mõisted *kinnistu*, *puit*, *metsaomanik* jne);

4. sünonüümsuhe väljendab samatähenduslikkust (nt *tervishoiuteenuse osutaja*, *tervishoiuteenuse pakkuja* ja *tervishoiuasutus*).

Iga kirjeldatav andmeelement seotakse ühe andmesõnastiku terminiga. Iga andmesõnastiku termin seotakse ühe või mitme ärisõnastiku mõistega.

Joonisel 5 ja 6 on toodud näited andmekirjelduse vormistamisest Excelis. Joonisel 5 on iga andmesõnastiku termini (veerg C) kohta kirjeldatud üks või mitut mõistet (veerg B). Joonisel 6 on ärisõnastiku mõistete (veerud B ja F) vahel määratud seosed (veerg D). Hierarhilise suhte puhul kasutatakse seosetüüpe LAIEM või KITSAM, assotsiatiivsuhte puhul SEOTUD, grupeeriva suhte puhul KUULUB GRUPPI või GRUPEERIV ning sünonüümsuhte puhul SÜNONÜÜM.

B	C
ANDMESÕNASTIKU TERMINIGA SEOTUD ÄRISÕNASTIKU MÕISTE(D)	ANDMESÕNASTIK ANDMESÕNASTIKU TERMIN
kliiniline uuring	kliinilise uuringu number
kliiniline uuring	kliinilise uuringu täielik pealkiri inglise keeles
kliiniline uuring	kliinilise uuringu uuringuplaani number
kliiniline uuring	kliinilise uuringu WHO number
vähesekkuv kliiniline uuring	põhjendus, miks on tegemist vähesekkuva uuringuga
kliiniline uuring; EudraCT	kliinilise uuringu EudraCT number
akadeemiline kliiniline uuring	märke, kas tegemist on akadeemilise uuringuga
kliiniline uuring	kliinilise uuringu algusaeg
kliiniline uuring	kliinilise uuringu lõpuaeg
kliiniline uuring	uuringu ajutise peatamise kuupäev
kliiniline uuring	uuringu taasalguse kuupäev

Joonis 5: Andmesõnastiku koostamise näidis Excelis

⁴ EVS-ISO 25964-2:2023 Informatsioon ja dokumentatsioon. Tesaurused ja nende koostalitlusvõime teiste sõnastikega. Osa 1: Infootsingu tesaurused

A	B	C	D	E	F
JRK NR	MÕISTE_ET	MÕISTE_EN	SEOSE TÜÜP	ON EELISTERMIN	SEOTUD MÕISTE
125	somaatiline rakk	<i>somatic cell</i>	SÜNUNÜÜM	JAH	keharakk
130	taotluse esitaja		SEOTUD	JAH	taotlus
135	toimeaine		KITSAM	JAH	koostisaine
137	toimeaine tootja		SEOTUD	JAH	toimeaine
147	uuringu lõpparuanne		SÜNUNÜÜM	JAH	kliinilise uuringu tulemuste kokkuvõte
154	vähesekkuv kliiniline uuring	<i>low-intervention clinical trial</i>	KITSAM	JAH	kliiniline uuring

Joonis 6: Ärisõnastiku koostamise näidis Excelis

2.3 Ärimõistete sisuline kirjeldamine

Järgnevalt on esitatud sõnastikes olevate ärimõistete praktilise kirjeldamise soovitusel.

- 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 meetodikadokument, kasuta nendes esitatud sõnavara.
- Ärisõnastiku mõisted tuletatakse andmesõnastiku terminitest, mis väljendavad andmeelementide sisu. Mõiste moodustatakse andmesõnastiku termini sellest osast, mis on kõige lähemal andmete kontekstile ja näitab, kelle või mille kohta on andmed. Näiteks andmesõnastiku termini *kodulindude arv* puhul mõisteks *kodulind*; andmesõnastiku termini *munejate kanade väljahautud tibude arv* puhul kirjeldatakse mõisteks *tibu*; andmesõnastiku termini *diislongivaguni kasutatava kütuse liik* puhul *kütus*. Andmesõnastiku termin seotakse ärisõnastiku mõistega.
- Kui andmesõnastiku termin sisaldab mitut olulist konteksti, kirjeldatakse ka need mõisteteks. Eelnevate näidete puhul vastavalt *muneja kana* ja *diislongivagun*.
- Kõik andmesõnastiku terminite alusel tuletatud mõisted lisatakse ärisõnastikku. Ärisõnastikku lisatakse töö käigus ka teisi andmekogu andmete mõistmiseks olulisi mõisteid. Mõistete vahel luuakse seosed. Mõistetevaheliste seoste tegemisel lähtutakse mõistete sisust, mitte esinemisest samas andmesõnastiku terminis.
- Eelnevate näidete puhul lisatakse ärisõnastikku nii mõiste *kodulind* kui ka selle laiem mõiste *lind*; nii mõiste *diislongivagun* kui ka selle laiem mõiste *raudteeveerem*.
- Kui andmesõnastiku termin sisaldab mitut üksteist täiendavat konteksti, siis lisatakse ärisõnastikku kõik vastavad mõisted ning seotakse need andmesõnastiku terminiga. Näiteks andmesõnastiku termin *metsateatise, ekspertiisi või välitöö objekti omaniku isikukood* seotakse mõistetega *metsateatise objekti omanik*, *ekspertiisi objekti omanik* ja *välitöö objekti omanik*.
- Soovitame asukohaga seotud andmesõnastiku terminite puhul kasutada üldistavat mõistet **aadressiandmed**, et ei peaks iga asukohadetaili eraldi mõisteks kirjeldama. Ärisõnastiku mõiste *aadressiandmed* katab kõik asukohaga seotud andmed, nagu nt registreerimis-, juriidilise, asukoha, toimumiskoha aadressi jm asukohta tähistavad andmed, sh sihtnumbrid, koordinaadid. Nt andmesõnastiku termin *õppeasutuse juriidiline aadress – postiindeks* on seotud mõistetega *õppeasutus* ja *aadressiandmed*.
- Kontaktandmetega seotud andmesõnastiku terminite puhul soovitame kasutada üldistavat mõistet **kontaktandmed**, et ei peaks iga detaili eraldi mõisteks kirjeldama. Ärisõnastiku mõiste *kontaktandmed* hõlmab kõiki isiku või organisatsiooniga seotud kontaktandmeid, mis ei ole aadressiandmed, nt nagu telefoninumber, e-posti aadress,

faksinumber, veebisaidi aadress, sotsiaalmeedia konto. Nt andmesõnastiku termin *isiku e-posti aadress* on seotud mõistetega *isik* ja *kontaktandmed*.

- **Tegevused**, mis on seotud andmekogu kontekstis oluliste objektide/sündmuste/protsesside staatuse või olekuga, võib formuleerida ärisõnastiku mõisteteks (nt *liikmeks astumine*, *liikmesuse peatamine*, *jäätmete käitlemine*).
- Tegevused, mis on seotud sündmuste dokumenteerimise või menetlemisega, ei moodusta üldreeglina ärisõnastiku mõisteid (nt dokumendi registreerimine, kinnitamine, allkirjastamine jmt). Näiteks andmesõnastiku termini *dokumendi registreerimise aeg* puhul ei moodustata ärimõistet dokumendi registreerimine, vaid *dokument*.
- Üldjuhul ei moodustata ärisõnastiku mõisteid terminit iseloomustavatest üldistest kirjeldustest, tunnustest, omadustest ega ka mitte sellistest üldlevinud mõistetest nagu 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 standardil⁵. Ärisõnastiku mõistekirje koosneb alltoodud kirjelduselementidest.

Ärimõiste kirjelduselement TERMIN

- Ärimõistet tähistav termin sõnastatakse üldjuhul ainsuses.
- 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, nt termin *CTISi teavitus*. Viimase näite puhul on soovitatav kirjeldada ärisõnastikus ärimõistena ka infosüsteemi nimi *Clinical Trials Information System (CTIS)* ning siduda see mõistega *CTISi teavitus*.

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 määratlustena tuleks eelistada õigusaktide legaaldefinitsioone, kui need on olemas.
 - Määratluse puudumisel või selle mõistmise hõlbustamiseks võib lisada selgituse. Selgitus võib sisaldada mõiste vabas vormis kirjeldust, täpsustavaid ja/või täiendavaid kommentaare, kasutuskonteksti vm mõistet avavat lisateavet, mis põhineb õigusaktidel või andmekogu dokumentatsioonil. Legaaldefinitsioonide

⁵ [Andmekirjelduse juhise Lisa 1: Andmekirjelduse standard](#). Ver 3.0 (Statistikaamet, 2024).

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 nt õigusaktid, andmekogu tehniline dokumentatsioon, juhised jms. Määratlust täiendavad ja selgitavad kommentaarid ning märkused võib lisada kirjelduselemendi **MÄRKUSED** väljale.
- Mõiste määratlus algab üldjuhul väikese algustähega ning on ilma lõpupunktita. Kui mõistet kirjeldav selgitus koosneb ühest või mitmest täislausel, siis kirjutatakse need 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 *aluspinnasega kohtkindlalt ühendatud ja inimtegevuse tulemusena ehitatud terviklik asi* (mitte *ehitis on aluspinnasega kohtkindlalt ühendatud ja inimtegevuse tulemusena ehitatud terviklik asi*).
- Mõiste määratlus sõnastatakse eesti ja inglise keeles. Ühe mõistekirje eesti- ja ingliskeelsed määratlused peavad omavahel sisuliselt kattuma.
- 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. Näiteks: 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 allikas 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: nt [Statistikaameti klassifikaatorite portaalis](#) avaldatud klassifikaator või [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 võimalikult täpne viide teabeallikale.
- Juhend, väljaanne vms, millele ei ole elektroonilist viidet. Allikaviiteks on allika pealkiri ja ilmumisandmed.
- Kui mõiste määratluse on koostanud oma asutus, lisatakse allikaviitena asutuse nimi.

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 andmekogu oskuskeeles käibel olevaid akronüüme. Näiteks: *EPS-põld, MAH, EHAK, EMTAK, ÜRO*.

Ärimõiste kirjelduselement ON EELISTERMIN

- Eelistermin on mõistet väljendav termin, mida eelistatakse selle sünonüümidele. Näiteks *tervishoiuteenuse osutaja, tervishoiuteenuse pakkuja* ja *tervishoiuasutus* on sünonüümsed terminid ja ärisõnastikus vastava seosetüübiga omavahel seotud. Kui *tervishoiuasutus* märgitakse ärisõnastikus eelisterminiks, siis see tähendab, et asutus eelistab neist kolmest kasutada just seda.

3 Andmesõnastiku terminite kirjeldamine

3.1 Mis on andmesõnastik?

Andmesõnastik on ühelt poolt terminite nimekiri ning teiselt poolt esitab terminitena andmeelementide nimetused (pealkirjad). 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.⁶ Iga kirjeldatav andmeelement seotakse ühe andmesõnastiku terminiga.

Andmekirjelduse mõttes vähemtähtsateks andmeteks võib üldjuhul lugeda süsteemi logi, andmebaasi kirjade loomise ja muutmise kuupäevi, failide räsiseid ja muud tehnilist tugiinfot, mis ei ole otseselt seotud organisatsiooni põhiülesannete täitmise käigus andmete töötlemise, andmepõhise juhtimise või andmevahetuse eesmärkidega.⁷ Seesuguste andmeelementide kohta ei ole vaja andmesõnastiku terminit moodustada. Klassifikaatorite ja koodiloenditega seotud andmeelemendid tuleks kirjeldada.

Igale andmeelemendile vastava andmesõnastiku termini kohta kirjeldatakse üks või mitu ärisõnastiku **mõiste(t)**.

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: *müüdü kütuse kogus, sportlasestipendiumi saaja ees- ja perekonnanimi, spordiobjekti riietusruumide 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 kellaeg*.
- Andmesõnastiku terminid soovitame võimalikult detailselt lahti kirjutada, näiteks: *e-posti aadress (mitte e-post), veebilehe aadress (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).

⁶ [Andmekirjelduse juhise Lisa 1: Andmekirjelduse standard](#). Ver 3.0 (Statistikaamet, 2024)

⁷ [Andmekirjelduse juhise](#). Ver 2.0 (Statistikaamet, 2023)

3.3 Termini sõnastuse 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: andmeelemendist *esindusõiguse kehtivuse algus* moodustatakse andmesõnastiku termin *spordiorganisatsiooni esindaja 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 *ankeeti iseloomustav tunnus pangakonto number* puhul 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 *ettevõtte müügi andmed: müüdüd kauba kogus*. Sellisel juhul võib kirjeldada andmesõnastiku termini lühemaks, piisab liitterminist *müüdüd kauba kogus*.

3.4 Andmesõnastiku termini vormistamine

Andmesõnastiku terminite kirjelduselemendid põhinevad andmekirjelduse standardil.⁸ 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.
- Üldjuhul terminis mõõtühikut ei kirjeldata. 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

⁸ [Andmekirjelduse juhise. Lisa 1: Andmekirjelduse standard](#). Ver 3.0 (Statistikaamet, 2024).

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 (? . ! / : ...). Ümarsulge soovitame vältida, kuid see ei ole alati võimalik.
- 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. Soovitav on akronüümid, mis ei ole üldtuntud, ärisõnastikus või andmesõnastiku termini määratluses lahti kirjutada.

Andmesõnastiku termini kirjelduselement MÄÄRATLUS

- Termini määratlus algab väikese algustähega ning on ilma lõpupunktita. Kui terminit kirjeldav selgitus koosneb täislause(te)st, siis kirjutatakse need 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*).
- Enamasti ei võeta termini määratlust üldsõnastikust (nt EMS, Esterm), 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 algustähe ja lauselõpupunktiga.
- Üldlevinud akronüüme võib andmesõnastiku märkustes kasutada, nt *KOV, EHAK, ÜRO*.

4 Näited

Ärimõistete sisulise kirjeldamise näited

Näited 1–3. Mõiste moodustamine. Mõiste moodustatakse andmesõnastiku termini sellest osast, mis on kõige lähemal andmete kontekstile ja näitab, kelle või mille kohta on andmed. Näiteks pangaandmete puhul on andmesõnastiku terminite *panga kood* ja *panga nimi* mõisteks *pank* (tabel 1). Mõiste seotakse andmesõnastiku terminitega. Ärisõnastikku lisatakse ka mõisted *pangatehing* ja *arveldus*. Isikuandmete puhul on andmesõnastiku termini *isiku e-maili aadress* mõisteks *isik* ja *kontaktandmed*. Mõlemad mõisted seotakse andmesõnastiku terminiga (tabel 1). Välismaalase andmete puhul on andmesõnastiku termini *välismaalase isikukood* või *muu identifitseeriv kood* mõisteks *välismaalane*. Mõiste seotakse andmesõnastiku terminiga (tabel 1).

Tabel 1. Ärisõnastiku mõisted ja andmesõnastiku terminid

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin
1	<i>pank</i>	<i>taotluse kohta tehtud makseotsuse arvelduse pangatehinguga seotud panga kood</i>
		<i>taotluse kohta tehtud makseotsuse arvelduse pangatehinguga seotud panga nimi</i>
2	<i>isik; kontaktandmed</i>	<i>isiku e-maili aadress</i>
3	<i>välismaalane</i>	<i>välismaalase isikukood või muu identifitseeriv kood</i>

Näited 4–8. Mõiste moodustatakse andmesõnastiku termini sellest osast, mis näitab, kelle või mille kohta on andmed (tabel 2).

Tabel 2. Ärisõnastiku mõisted ja andmesõnastiku terminid

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin
4	<i>käibemaks</i>	<i>arve käibemaksu summa</i>
5	<i>arve</i>	<i>arve summa käibemaksuta</i>
6	<i>müügitulu</i>	<i>ettevõtte müügitulu suurus</i>
7	<i>töövõime</i>	<i>eksperdiarvamus töövõime hindamise kohta</i>
8	<i>eksperdiarvamus</i>	<i>eksperdiarvamuse andmise kuupäev</i>

Näited 9-11. Mõistetevahelised seosed – hierarhiline suhe. Mõistete vahel kehtib hierarhiline suhe, kui kitsam mõiste mahub täielikult laiemasse mõistesse. Laiema mõiste lisamisel soovitame kontrollida, kas töötab *all-and-some* test. Näiteks kõik mitterahalised kohustised on kohustised, mõned kohustised on mitterahalised kohustised. Seega *kohustis* on *mitterahalise kohustise* laiem mõiste (tabel 3).

Tabel 3. Ärisõnastiku mõisted

Näite jrk nr	MÕISTE_ET	SEOSE TÜÜP	SEOTUD MÕISTE
9	<i>mitterahaline kohustis</i>	KITSAM	<i>kohustis</i>
10.	<i>rahaline kohustis</i>	KITSAM	<i>kohustis</i>
11.	<i>kohustis</i>		

Näited 12-15. Mõistetevahelised seosed – hierarhiline suhe. Kui mõistet *isik* kasutatakse andmekogus nii füüsiliste kui juriidiliste isikute kohta, on *isik* laiem mõiste ärisõnastiku mõistetele *füüsiline isik* ja *juriidiline isik*. Alltoodud näites eristatakse juriidilise isikuna ka süsteemseid asutusi, sel juhul on *juriidiline isik* omakorda mõiste *süsteemne asutus* laiem mõiste (tabel 4).

Tabel 4. Ärisõnastiku mõisted

Näite jrk nr	MÕISTE_ET	SEOSE TÜÜP	SEOTUD MÕISTE
12	<i>füüsiline isik</i>	KITSAM	<i>isik</i>
13	<i>juriidiline isik</i>	KITSAM	<i>isik</i>
14	<i>süsteemne asutus</i>	KITSAM	<i>juriidiline isik</i>
15	<i>isik</i>		

Näited 16–18. Hierarhiasuhte määramise allikad. Hierarhiasuhte määramisel saab toetuda mõistete määratlustele ning asjakohastes materjalides toodud liigitustele. Näiteks eristab karistusseadustik süütegude liikidena kuriteod ja väärteod, kuriteod omakorda on esimese ja teise astme kuriteod. Ärisõnastikus loome vastavad seosed nende mõistete vahel (tabel 5).

Tabel 5. Ärisõnastiku mõisted

Näite jrk nr	MÕISTE_ET	SEOSE TÜÜP	SEOTUD MÕISTE
16	<i>kuritegu</i>	KITSAM	<i>süütegu</i>
17	<i>väärtegu</i>	KITSAM	<i>süütegu</i>
18	<i>süütegu</i>		
	<i>esimese astme kuritegu</i>	KITSAM	<i>kuritegu</i>
	<i>teise astme kuritegu</i>	KITSAM	<i>kuritegu</i>

Näited 19–20. Mõiste moodustamine andmeelemendi loendi põhjal. Andmesõnastiku terminiga *menetluse liik* seotud andmeelemendi väärtused on seotud loendiga. Ärisõnastikku on lisatud mõisted erinevate menetlusliikide kohta ning nende ühine laiem mõiste on *menetlus*. Mõisted *kiirmenetlus* ja *käskmenetlus* ise ei pruugi ühegi andmesõnastiku terminiga seotud olla (tabel 6).

Tabel 6. Ärisõnastiku mõisted

Näite jrk nr	MÕISTE_ET	SEOSE TÜÜP	SEOTUD MÕISTE
19	<i>kiirmenetlus</i>	KITSAM	<i>menetlus</i>
20	<i>käskmenetlus</i>	KITSAM	<i>menetlus</i>

Näited 21–22. Mõiste olulisus. Andmesõnastiku termini sidumisel mõistega arvestatakse ärisõnastiku vajadustega. Ärisõnastikku ei ole vaja lisada andmekogu kontekstis väheolulisi mõisteid (näiteks *isikut tõendav dokument*) või mõisteid liialt killustada (näiteks *silla kaldasammas*, *silla vahesammas*) (tabel 7).

Tabel 7. Ärisõnastiku mõisted ja andmesõnastiku terminid

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin
21	<i>sillasammas</i>	<i>silla kaldasamba materjal</i>
		<i>silla vahesamba materjal</i>
22	<i>eksami sooritaja</i>	<i>eksami sooritaja isikut tõendava dokumendi number</i>

Näited 23–25. Klassifikaatoriga seotud terminitele ühise mõiste kirjeldamine. Mõiste *diagnoos* grupeerib andmesõnastiku terminites esitatud diagnoosid, mis on seotud rahvusvahelise haiguste ja terviseprobleemide statistilise klassifikatsiooniga (RHK-10) (tabel 8). Mõiste *toimingud lapsele* või *toimingud naisele* on ühine mõiste nendele terminitele, mis täpsustavad isikuga seotud detailseid toiminguid. Andmeelementide väärtused tulevad mõlemal juhul loenditest.

Tabel 8. Ärisõnastiku mõisted, andmesõnastiku terminid ja andmeelemendi loendid

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin	Andmeelemendi loend
23	<i>diagnoos</i>	<i>hingamishäired</i>	sr_jah_ei
		<i>ajuinfarkt</i>	sr_jah_ei
		<i>sepsis</i>	sr_jah_ei
24	<i>toimingud lapsele</i>	<i>K-vitamiin</i>	sr_jah_ei_vanemkeeldus
		<i>surfaktantravi</i>	sr_jah_ei
		<i>antibakteriaalne ravi</i>	sr_alla_ule_72h
25	<i>toimingud naisele</i>	<i>neuroproteksioon</i>	sr_jah_ei
		<i>neuroproteksiooni annus</i>	sr_taielik_mittetaielik
		<i>loote kopsude ettevalmistus</i>	sr_jah_ei

Andmesõnastiku termini sisulise kirjeldamise näited

Näited 26–28. Andmesõnastiku terminite sõnastuse ühtlustamine. Kui andmeelement viitab kirje unikaalsele identifikaatorile, siis tuleb ühtlustada vastavate terminite sõnastused kogu andmestiku piires. Nt andmesõnastiku termin kujul *viide ... kirjele* peaks olema kogu andmestikus läbivalt kasutusel (tabel 9).

Tabel 9. Ärisõnastiku mõisted, andmesõnastiku terminid ja andmeelemendi tähised

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin	Andmeelemendi tähis
26	<i>dokument</i>	<i>viide dokumendi kirjele</i>	document_id
27	<i>nõue</i>	<i>viide nõudele</i>	NOUDED_ID
28	<i>isik</i>	<i>isiku viide</i>	ISIKUD_ID

Näide 29. Kirje identifikaatorite puhul piisab ühest ühisest terminist *kirje identifikaator*. Termin seotakse mõistega *kirje* (tabel 10).

Tabel 10. Ärisõnastiku mõiste, andmesõnastiku termin

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin	Andmeelemendi tähis
29	<i>kirje</i>	<i>kirje identifikaator</i>	id

Näited 30-32. Boolean-andmetüüp. Boolean-andmetüübiga andmeelementide puhul on termin kirjeldatud kujul *märke, kas...* või tõese väärtuse vastena, nt küsimuse „kas tegu on kinnipidamisasutusega“ vastus (termin) on *kinnipidamisasutus*. Termini *märke, kas on korduv kõne või esmane kõne* puhul on täpsem, kui terminis on toodud vastuse mõlemad variandid, korduv kõne ja esmane kõne. Määratluse alla saab lisada selgituse (tabel 11).

Tabel 11. Ärisõnastiku mõisted, andmesõnastiku terminid ja andmetüüp

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin	Andmetüüp	Andme-elementi tähis
30	<i>eeltest</i>	<i>märke, kas sooritaja võib testimiskorruga testis tagasisidet näha kohe peale sooritamist, enne koondtulemuse avaldamist</i>	boolean (JAH/EI)	is_pre_test
31	<i>kinnipidamisasutus</i>	<i>kinnipidamisasutus</i>	boolean (JAH/EI)	on_kinnipidamisasutus
32	<i>mürgistuskõne</i>	<i>märke, kas on tagasi helistatud</i>	boolean (JAH/EI)	called_back

Näide 33. Termini pikkus. Sagedased ja ülemäära pikad terminid võiksid olla konkreetset ja lühemad. Nt termin *töötaja koolitamine* andmeelemendi *tööle asumisele eelnevalt vähemalt 12 kuud järjest töötuna arvel olnud töötaja koolitamine* puhul (tabel 12).

Tabel 12. Ärisõnastiku mõisted ja andmesõnastiku terminid

Näite jrk nr	Andmesõnastiku terminiga seotud ärisõnastiku mõiste(d)	Andmesõnastiku termin
33	<i>töötaja</i>	<i>tööle asumisele eelnevalt vähemalt 12 kuud järjest töötuna arvel olnud töötaja koolitamine</i>

Näited 34–36. Termini täpsus. Andmesõnastiku termin peaks andmeelemendi konteksti võimalikult täpselt avama, vajadusel mitmest sõnast koosneva liittermini või terminifraasina. Nt erinevate piirete materjalide puhul on terminis täpsustatud, millise piirdega on konkreetse andmeelemendi puhul tegu (tabel 13).

Tabel 13. Andmesõnastiku terminid

Näite jrk nr	Andmesõnastiku termin
34	<i>silla pealesõidu pörkepiirde materjal</i>
35	<i>silla pörkepiirde materjal</i>
36	<i>loomatõkkeaia piirde materjal</i>

Kasutatud allikad

Andmekirjelduse juhise juhise. Versioon 2.0. Statistikaamet, 2023,
https://www.kratid.ee/_files/ugd/7df26f_aa72d2833b514f20a07531d3b5bf3d9c.pdf

Andmekirjelduse juhise. Lisa 1: Andmekirjelduse standard. Versioon 3.0. Statistikaamet, 2024,
https://stat.ee/sites/default/files/2024-08/AH_juhise_andmekirjeldus_standard_3.0.1.pdf

Chaki, Saumya. *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 = Terminology work. Vocabulary: Part 1. Theory and application. Eesti Standardikeskus, 2002,
<https://www.evs.ee/et/evs-iso-1087-1-2002>

EVS-ISO 25964-1:2023 Informatsioon ja dokumentatsioon. Tesaurred ja nende koostalitusvõime teiste sõnastikega. Osa 1: Infootsingu tesaurred. Eesti Standardikeskus, 2023,
<https://www.evs.ee/et/evs-iso-25964-1-2023>

EVS-ISO 25964-2:2023 Informatsioon ja dokumentatsioon. Tesaurred ja nende koostalitusvõime teiste sõnastikega. Osa 2: Koostalitusvõime teiste sõnastikega. Eesti Standardikeskus, 2023,
<https://www.evs.ee/et/evs-iso-25964-2-2023>

Erelt, Tiiu. *Terminiõpetus*. Tartu, 2007.