

# Andmekirjelduse kohta korduma kippuvad küsimused

Statistikaamet  
andmehaldus@stat.ee  
mai, 2023

## ANDMEKIRJELDUSED

### Mis on andmekirjeldus?

Andmestiku või tema osade (näiteks andmeelementide) kirjeldused.

### Miks on vaja andmeid kirjeldada?

Andmeid on vaja kirjeldada sellepärast, et andmete sisu ja seosed teiste andmetega oleksid andmete kasutajaile paremini mõistetavad. Andmekirjeldustest saadav kasu on suurim siis, kui andmeid kirjeldatakse kokkulepitud viisil kindla meetodika järgi ja seda tehakse järjepidevalt.

### Kes kasutavad andmekirjeldusi?

Andmekirjeldusi kasutavad andmete kasutajad ja andmekirjelduste koostajad. Andmekirjeldus on nn dokument andmete paremaks mõistmiseks ja seoste loomiseks teiste andmetega.

### Kust ma leian vajalikud materjalid andmekirjelduste koostamisega alustamiseks?

Andmekirjelduse standardi leiad Statistikaameti kodulehe rubriigis *Andmehaldus* ([https://www.stat.ee/sites/default/files/2022-06/AH\\_juhis\\_andmekirjeldus\\_standard\\_2.0.pdf](https://www.stat.ee/sites/default/files/2022-06/AH_juhis_andmekirjeldus_standard_2.0.pdf)). See dokument esitab nii vajalikud andmeväljad kui ka näited nende sisustamiseks. Ideede kogumiseks ja terminite kasutamiseks oma andmestikus võib kasutada ka juba koostatud andmekirjeldusi (sõnastikke), mis on esitatud andmestike kaupa Statistikaameti pilveteenuses ([pilv.stat.ee](http://pilv.stat.ee)). Juurdepääsu pilveteenusele saad tellida kontaktaadressilt [andmehaldus@stat.ee](mailto:andmehaldus@stat.ee).

### Kust ma saan tuge ja abi andmekirjelduste koostamisega alustamiseks?

Statistikaameti andmehalduse tiim pakub abi andmekogudele andmekirjelduste koostamisel ning nõustab asutusi andmekirjelduse sisu lahtimõtestamisel. Võta meiega ühendust kontaktaadressil [andmehaldus@stat.ee](mailto:andmehaldus@stat.ee).

## ANDMESTIKUD

### Mis on andmestik?

Andmestik on andmete komplekt, kus andmed moodustavad seoste kaudu ühe loogilise terviku.

### Mis on RIHAKE?

RIHAKES põhieesmärgiks on võimaldada asutusel kirjeldada oma andmestikke. Andmestike kirjeldamise aluseks on ühelt poolt füüsilise andmehoidla, näiteks andmebaasi struktuur ja teiselt poolt sõnastiku mõisted. Nende omavahelisel seostamisel tekib erialaterminites andmetest arusaamine.

## ANDMETE KVALITEET JA STANDARDID

### Mis on kvaliteetse andmekirjelduse tunnused?

Kvaliteetne kirjeldus on standardiseeritud ja ühtlustatud andmekirjeldus, mis on koostatud kokkulepitud viisil ja andmestike kaupa. Kokkulepeteks on tavaliselt standardid ja juhised.

### Kes vastutab andmekirjelduste kvaliteedi eest?

Kvaliteedi eest vastutab nii andmekirjelduste tegija kui ka andmete omanik. Andmekirjelduste koostaja jälgib, et kirjeldused vastaksid etteantud nõuetele nii vormilt kui ka sisult, ning andmete omanik jälgib kirjelduste vastavust andmete sisule.

### Mis on andmekirjelduse standard?

Andmekirjelduse standard on juhised, mis määrab andmekirjelduse koostamise nõuded.

## SÕNASTIKUD, TERMINID, MÕISTED

### **Mis on andmesõnastik?**

Andmesõnastik on ühelt poolt terminite loetelu ning teiselt poolt esitab terminitena andmeelementide sisu.

### **Mis on ärisõnastik?**

Ärisõnastikus on mõisted, need on omavahel seostatud, näidatud on mõiste eelistermin ja seosed ning üldjuhul on antud mõiste määratlus või tähenduse selgitus. Ärisõnastikus on mõistete vahelised seosed hierarhilised, sisulised (assotsiatiivsed) või grupeerivad.

### **Mis on andmesõnastiku termin?**

Andmesõnastiku termin on mõistet tähistav (liit)sõna või mitmest sõnast koosnev liittermin. Andmesõnastiku terminid ja mõisted on seotud ärisõnastiku terminite ja mõistetega ning kasutavad terminite tähistamisel ärisõnastiku sõnavara. Näiteks töötajate arv, isiku vanus, ettevõtte tegevusala jne.

### **Mis on ärisõnastiku termin?**

Ärisõnastiku termin on kasutusel olevat mõistet tähistav termin, tavaliselt üks või mitu selget erialast tähendust omavat sõna. Näiteks litsents, taotlus, dokumendi menetleja jne.

### **Miks nimetatakse ärisõnastiku ja andmesõnastiku termineid mõlemaid terminiteks, kuigi need on sisult erinevad?**

Nii ärisõnastiku kui andmesõnastiku terminid on terminid. Kirjapildilt võivad mõlema sõnastiku terminid kas kattuda või olla väga erinevad. Näiteks üks ja sama termin „verekaotus“ väljendab andmesõnastikus kogust milliliitrites ning ärisõnastikus sellega seotud nähtust või seisundit. Ärisõnastiku termin kannab andmesõnastiku termini olulisimat sisu. Andmesõnastiku termin aga väljendab ärisõnastiku terminiga tähistatud objekti, nähtuse, sündmuse või olukorra tunnuseid või seda, kuidas vaadeldavat objekti/nähtust/sündmust mõõdetakse. Ärisõnastiku termin on näiteks „isik“ ja talle vastav andmesõnastiku termin on „isiku vanus“; ärisõnastiku terminile „ettevõtte“ vastab andmesõnastiku termin „ettevõtte registrikood“.

### **Mis on mõiste määratlus?**

Määratlus on mõiste definitsioon ehk mõiste sisu lühike ja täpne esitus oluliste tunnuste abil, mis eristavad seda teistest mõistetest. Määratlused võivad olla näiteks sisumääratlused või mahumääratlused. Näiteks mõiste „dokument“ (sisu)määratlus Sõnaveebis on „ametlik paber, kirjalikus vormis teade millegi tõendamiseks, õiguste kinnitamiseks vms“. Mahumääratluse puhul loetletakse mõiste kõik alammõisted ühe jaotuskriteeriumi põhjal, näiteks „lastiveoüksus“ – „kaubaveoauto, raudtee kaubavagun, konteiner, paakauto, raudteevagun või teisaldatav paak“.

### **Kui üht mõistet väljendatakse eesti keeles mitme erineva terminiga ja inglise keeles on kasutusel üks ja sama tõlkevaste, kas siis tuleb mõisted ühtlustada ja kirjeldada vastav termin ärisõnastikus ühekordselt?**

Ärisõnastikku kirjeldatakse mõisted asutuse terminoloogiat kasutades. Kui asutus kasutab sarnase kirjapildiga termineid, mis on erinevalt defineeritud, siis kirjeldatakse ärisõnastikus mitu terminit. Terminid ühtlustatakse ainult siis, kui nende määratlus on täpselt ühesugune (nt sel juhul võetakse üks neist kasutusele eelisterminina). Nt ärisõnastiku eestikeelsetele terminitele „algdokument“ ja „alusdokument“ vastab mõlemal juhul sama ingliskeelne termin „source document“; sama kehtib ka vastupidisel juhul, nt ingliskeelsete terminite „enterprise“ ja „company“ vaste on „ettevõtja“.

### **Kui sõnastikud kirjeldatakse RIHAKESse, kas siis Excelisse kirjeldamise vajadus kaob ära.**

Excelit on soovitatav kasutada ettevalmistava tööna terminite kirjeldamiseks. Terminite kokkuotsimine erinevatest allikatest ning määratluste võrdlemine ning ühtlustamine on suuremahuline töö, mis on vaja enne andmekataloogi kandmist ära teha. Kui sõnastikud on koostatud ja muudatusi on pigem harva, näiteks on vaja defineerida mõni uus termin või definitsiooni täpsustada, siis ei ole vaja ilmingimata alustada Excelist, vaid kirjeldused võib teha kohe andmekataloogi.