



TARTU ÜLIKOOL
RAKE

EESTI ANDMEMAJANDUSE TURUVÄÄRTUSE HINDAMISE METOODIKA KOOSTAMINE



2024

Juuni



EESTI ANDMEMAJANDUSE TURUVÄÄRTUSE HINDAMISE METOODIKA KOOSTAMINE

Tellinud Majandus- ja Kommunikatsiooniministeerium

Lõpparuanne

Tartus 2024. aastal

Uuringu tellis Majandus- ja Kommunikatsiooni-
ministeerium ning selle teostavad Tartu Ülikooli
sotsiaalteaduslike rakendusuringute keskus ja STACC
OÜ.

UURINGU AUTORID:

Mati Mõtte, TÜ

Siim Espenberg, TÜ

Andres Võrk, TÜ

Lisanna Lehes, TÜ

Tiit Sepp, STACC OÜ

Anneli Saaroja, TÜ

RAKE

RAKE on võrgustikutüüpi rakendusuringute keskus. Meie missioon on edendada teadmisel põhinevat otsustamist Eesti ühiskonnas. Lisaks RAKE meeskonnale kaasame kõrgeima kvaliteedi tagamiseks oma uuringutesse valdkondlikke eksperte nii Tartu Ülikoolist kui vajadusel ka väljastpoolt. RAKE võrgustikust leiab nii sotsiaalteadlasi kui meditsiini-, loodus-, tehnika- ja humanitaarteaduste valdkonna esindajaid.

KONTAKTANDMED:

Lossi 36-303, 51003, Tartu
rake@ut.ee
<http://rake.ut.ee>

ISBN: 978-9985-4-1432-3

SISUKORD

SISUKORD	5
SISSEJUHATUS	7
1. UURIMISÜLESANDED	8
2. ANDMEMAJANDUSE VÄÄRTUS	9
2.1. TERMINOLOOGIA	9
2.2. METOODILISED KÄSITLUSED	13
2.3. METOODIKATE VÕRDLUS	16
2.3.1. EUROOPA ANDMETURU SEIRETÖÖLAUD.....	16
2.3.2. ANDMETE VÄÄRTUSE KULUPÕHINE HINDAMINE	21
3. TEISTE RIIKIDE KOGEMUS	26
3.1. ÜHENDKUNINGRIIGID.....	26
3.2. HOLLAND.....	27
3.3. SOOME.....	28
3.4. ROOTSI	30
4. JUHTIMISLAUD JA ANDMEMAJANDUSE VÄÄRTUS	31
4.1. ANDMEMAJANDUSE VÄÄRTUSE HINDAMISE ARVUTUSKÄIK KULUPÕHISE MEETODIGA	31
4.1.1. TÖÖTAJATE ARVU JA TÖÖTASU LEIDMINE	31
4.1.2. ANDMETEGA TÖÖTAMISE OSAKAAL.....	34
4.1.3. ANDMETEGA TÖÖTAVATE INIMESTE NÄITAJAD.....	36
4.1.4. ANDMETEGA TEGELEMISELE KULUV TÖÖJÕUKULU	38
4.1.5. MUUDE KULUDE LISAMINE.....	39
4.1.6. KOONDTULEMUS.....	41
4.2. ETTEVÕTETE INDIKAATORID JUHTIMISLAUAL	42
4.3. MUUD INDIKAATORID	44
4.4. TÖÖLAUA TEHNILINE RAKENDUS JA SISENDANDMED	45
4.4.1. ANDMEMAJANDUSE KOGUVÄÄRTUS	45
4.4.2. TÖÖHÕIVE.....	48
4.4.3. ETTEVÕTTED.....	49
4.4.4. VÄLISLINGID.....	50
4.4.5. TAUSTAINFORMATSIOON.....	51
4.5. TÖÖLAUA VALIDEERIMINE	52
KOKKUVÕTE JA JÄRELDUSED	55
KASUTATUD KIRJANDUS	59
LISAD	62
LISA 1. DIMENSIONAALSE ANDMETE HINDAMISE MUDELI RAAMISTIK	62
LISA 2. ANDMEMAJANDUSE VÄÄRTUSE MUDELITE ANDMETE KAARDISTUS.....	63

LISA 3. ANDMEPÄRING STATISTIKAAMETILE.....	64
LISA 4. MUUDE KULUDE SKALEERIMISFAKTORI SISENDFAIL.....	67
LISA 5. TÖÖLAUA NÄITAJATE UUENDAMISE PROTSESS.....	69
LISA 6. TÖÖLAUA PROTOTÜÜBI KOOD.....	70

SISSEJUHATUS

Andmed muutuvad ühiskonnas ja poliitilises otsustusprotsessis järjest olulisemaks. Majandus- ja kommunikatsiooniministeeriumi juhtimisel on koostatud Eesti digiühiskonna 2030 arengukava (edaspidi digiühiskond 2030), milles nähakse võimalust 2030. aastal teha kõik riigi otsused kvaliteetsete andmete abil ning sealjuures rakendada tehisintellekti ehk kratte, et saavutada uus tõhusushüppe. Teadus- ja arendustegevuse, innovatsiooni ning ettevõtluse (TAIE) arengukava 2021–2035 Eestis seab omakorda eesmärgiks tõsta ettevõtluse rahvusvahelist konkurentsivõimet, luua kõrgemat lisandväärtust ning suurendada ekspordivõimekust. TAIE eesmärkide hulgas on andmevaldkonna arendamine, millega nähakse ette, et andmed aitaksid lisaks paremale otsustusprotsessile hoida kokku ressursse (aeg, töömaht, raha) ning parendada suhtlust ja koostööd osapoolte vahel.

Eeltoodud eesmärkide saavutamiseks on oluline tagada andmete kättesaadavus ja kvaliteet. Samas kvaliteetsete andmete kättesaadavus nõuab investeringuid ja ühiskonna digivõimekust, sh nii halduse, aga ka oskuste poolelt. Käesoleva uuringu autorite arvates on digimuutuste võimendamine võimalik, aga sealjuures on oluline mõiste tarbija vaadet, ärimudeleid ja ressursside kasutust (tõuketegureid). Andmete väärtus sõltub paljudest teguritest, mida esindavad n-ö kasutaja kui looja poolne panus. Digivõimenduse ressursside efektiivset kasutamist on oluline meetoodiliselt mõõta, mille tulemuseks on andmete kasutamisest tekkiv lisaväärtust.

Selles uuringus vaatleme olulisemaid meetoodikaid, mis aitavad hinnata andmete väärtust ning seda läbipaistvalt mõõta, mis on omakorda vajalik selleks, et hinnata andmete valdkonnaga seotud meetmete ja tegevuste kasusid ja/või kahjusid majanduses tervikuna. Euroopa andmeportaali avaldatud mõjuhinnangu kohaselt on hinnatud, et üksnes avaandmete mõju Eesti majandusele aastal 2021 oli 334 miljonit eurot. Too uuring viidi läbi Eesti-väliste partnerite poolt, mistõttu jääb turuväärtuse täpne suurus ja kasv või kahanemine (ning nende hinnangute põhjused) teadmata. Kuna eelnevalt läbiviidud andmemajanduse alastes uuringutes Eesti kohta on meetoodilised alused läbipaistmatud, siis käesolev uuring püüab pakkuda läbipaistvama meetodi andmete väärtuse hindamiseks.

Uuringu olulisus seisneb veel selles, et sõltuvalt väärtuse hindamise mudelarvutustest võib olla vajalik spetsiifiline valdkondlik informatsioon ning andmed, mille efektiivset kogumist tuleks edaspidi Eestis rakendada. Piltlikult öeldes ka andmete väärtuse hindamine vajab sisendandmeid, mis peavad olema efektiivselt kogutavad. Arvestades võimalikku andmemajanduse arengut, siis hindamiseks vajalikud muutujad on soovitatav kindlaks määrata võimalikult kiiresti ning rakendada vastava seiresüsteem (nt riikliku statistika osana). Vajalike andmete olemasolu järgselt saab teha täiendavaid arvutusi ja hinnata digitehnoloogiate mõju majandusele ja ühiskonnale tervikuna. Käesoleva uuringu läbiviimise idee sai alguse 2022. aasta Arenguseire Keskuse raportis „Andmeühiskonna tulevik. Stsenariumid aastani 2035”¹, mis tõi esile vajaduse luua meetodika andmemajanduse turuväärtuse ja hetkeolukorra hindamiseks Eestis. Meetodika aitaks kaasa poliitikakujundamise mõju hindamisele vastavalt andmete valdkonna hetkeseisule Eestis majanduses nii tervikuna kui ka sektoripõhiselt.

¹ Arenguseire Keskus (2022). Andmeühiskonna tulevik. Stsenariumid aastani 2035. Raport. Tallinn: Arenguseire Keskus.

1. UURIMISÜLESANDED

Uuringu eesmärk on anda ülevaade andmevaldkonna hetkeseisust ja pakkuda välja metoodika Eesti andmemajanduse mahu ning andmete turuväärtuse hindamiseks. Ühtlasi on vajalik metoodika valideerimine ja sobivuse katsetamine Eestis andmemajanduse turuväärtuse hindamiseks. Metoodika koostamisel lähtusime vajadusest hinnata andmemajandust sellisel viisil, mis võimaldaks saada tervikpildi Eesti andmemajandusest. Lisaks pidasime oluliseks seda, et loodud metoodikat oleks võimalikult lihtne mõista ning valideerida olemasolevale Eesti andmemajanduse mahu näitajatele tehtavate teoreetiliste muudatuste kaudu.

Uuringu esimese etapi ülesanded olid järgmised:

- analüüsida, kas ja mis tingimustel oleks võimalik rakendada Eestis Statistics Canada (2019a;b) ja selle OECD ekspertide edasiarendust (Corrado, Haskel *et al* (2022)) mudeleid;
- töötada läbi Euroopa Komisjoni andmemajanduse mõõtmise töölaual kasutatavad indikaatorid (European Commission, 2020, 2023), nende publitseeritud metoodika ja analüüsitakse, kas see metoodika on kohandatav Eesti avalikele andmetele või kergesti registritest päritavatele andmetele;
- tutvustada metoodikat Eesti ekspertidele ja täiendada metoodikat vastavalt saadud tagasisidele;
- tutvustada mõlema analüüsi tulemusi tellijale;
- otsustada koostöös tellijaga, millise lähenemisega liikuda edasi teise etappi.

Märgitud ülesannete täitmisel on oluline esitada metoodikas erinevate mõõdikute eeliseid ja puudusi. Ühtlasi on olulised ettepanekud, mida peaks mõõtma ja mis andmeid peaks koguma, mida praeguse seisuga Eestis kogutud/kättesaadavad ei ole.

Vastavalt uuringu I etapi tulemustele ja tellija soovidele rakendati uuringu II etapis korraga mõlemat lähenemist: andmete väärtuse rahalist mõõtmist ja andmeturu ning andmemajanduse indikaatorite komplekti väljatöötamist. Selleks toimus II etapis järgmiste ülesannete täitmine:

- andmete väärtuse metoodika väljapakkumine ja näitearvutuste tegemine Eesti registriandmete põhjal koostöös Statistikaametiga;
- andmemajandust iseloomustavate indikaatorite valik avalikest andmetest;
- juhtimislaua kuvatavate indikaatorite valik ja nende esitamise viisi (graafiline, arvuline jm) valik;
- juhtimislaua prototüübi väljatöötamine ja selle tellijaga kooskõlastamine;
- juhtimislaua andmete uuendamise skeemi välja pakkumine.

Ülesannete täitmisel jälgisime andmemajanduse vaate andmist nii majanduse kui terviku, aga ka tegevusalade lõikes. Käesolevas rakendusuuringu oli teadus- ja arendustegevuse põhitingimustest olulisel kohal uudsuse kriteerium – rakendusuuringu raames loodi metoodika Eesti andmemajanduse turuväärtuse hindamiseks.

2. ANDMEMAJANDUSE VÄÄRTUS

2.1. Terminoloogia

Metoodiliste valikute tegemisel on oluline mõista eelnevate autorite poolt koostatud analüüse ja definitsioone sealjuures kasutatud andmeid. Selles alapeatükis anname ülevaate olulistest mõistetest, mis on seotud andmemajanduse, sh andmete väärtusahela defineerimisega. Mõistete koondamisel oleme kasutanud eelkõige Euroopas kasutusele võetud termineid, mille juurde saab tuua mõistete erinevuse mujalt maailmast (eelkõige OECD riikide näitel). Juhul, kui ei ole viidatud teisiti, siis peamised allikad terminoloogia määramisel on

- *European DATA Market Study 2021–2023 (The European Data Market (EDM) Monitoring Tool; European Commission 2021)*²,
- *Survey of Businesses on the Data Economy*³ ja
- *Measuring the economic value of data (Going Digital Toolkit)*⁴.

Andmemajanduse (Data Economy) all mõistame majandussüsteemi, kus andmed on oluline majanduskasvu ressurss ja kus andmetega seotud tooted ja teenused on olulised majandusse panustajad. Euroopa andmeturu uuringus 2021–2023 defineeritakse seda kui digitaalset ökosüsteemi, kus andmeid kogub, korraldab ja vahetab ettevõtete, üksikisikute ja institutsioonide võrgustik, et luua majanduslikku väärtust.

Selles süsteemis kogutakse, salvestatakse, analüüsitakse ja vahetatakse suuri andmete koguseid, sageli kasutades kaugemale arenenud tehnoloogiaid nagu tehisintellekt, masinõpe ja pilvearvutus. See hõlmab paljusid majandusharusid, sealhulgas infotehnoloogiat, tööstust, tervishoidu, rahandust, meelelahutust ja palju muud, mis tähendab andmeturu kaudu otseseid, kaudseid ja indutseeritud mõju majandusele.

Andmemajandust on praktikas käsitletud hierarhiliselt:

- 1) andmete ja nendega seotud investeeringute väärtuse leidmine (*Economic Value of Data*),
- 2) andmeturu (*Data Market*) suuruse leidmine,
- 3) andmemajanduse (*Data Economy*) suuruse leidmine.

Andmemajanduse fookuses on peamiselt andmed ja nende väärtustamine, mis võrreldes mõistega digimajandus (vt mõiste digimajandus – *Digital Economy*) on kitsam. Digimajanduse keskmes on digitaalne tehnoloogia ja automatiseerimine, aga ka e-kaubandus.

² <https://digital-strategy.ec.europa.eu/en/library/results-new-european-data-market-study-2021-2023>

³ <https://digital-strategy.ec.europa.eu/en/library/survey-businesses-data-economy-2022>

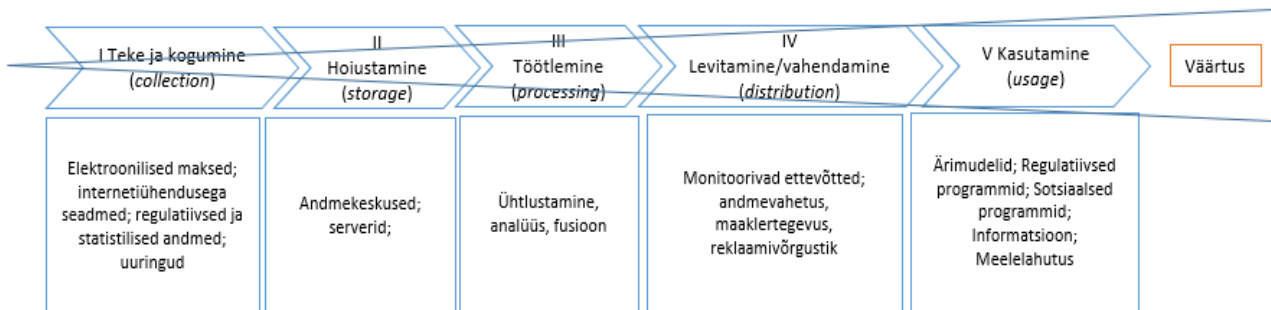
⁴ <https://goingdigital.oecd.org/en/theme/7#measurement-roadmap>

- **Andmed (Data)** defineeritakse tavaliselt kui kvalitatiivseid või kvantitatiivseid väiteid või teavet, mida saab kodeerida ja mille eelduslik olemus on faktiline. Andmed on teabe taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötluks⁵.

Andmemajanduses käsitletakse eelkõige andmeid, mida kogutakse, töödeldakse, salvestatakse ja edastatakse digitaalse teabe infrastruktuuri kaudu ja/või digitaaltehnoloogia abil. See määratlus hõlmab multimeediaobjekte, mida kogutakse, salvestatakse, töödeldakse ja edastatakse kasutamiseks digitaaltehnoloogiate (näiteks piltide andmebaaside) kaudu.

Andmeid saab klassifitseerida ka tüübi järgi, mis tuleneb nii allikast, aga ka andmettevõtte või -valdaja tüübist. Näiteks avaliku sektori andmed, kasutajate loodud sisu (B2C), äriandmed (B2B), uuringud ja teaduslikud andmed, organisatsioonide ja ettevõtete tegevuse andmed.

- **Andmebaas⁶ (Database)** on süsteem, mis hoiab mistahes omavahel seotud informatsiooni, olgu see teksti, pildi, audio või mõnel muul elektroonselt säilitataval kujul. Andmebaasi tüüpe on erinevaid, nt relatsiooniline andmebaasisüsteem (RDBMS) ja NoSQL-andmebaasisüsteem.
- **Andmete ökosüsteem (kasutusel Euroopas) (Data Ecosystem)** hõlmab terviklikult andmete väärtusahela, arenguraamistiku (sh andmeturu ja -majanduse regulatiivsed alused), sidusrühmade kategooriate ja rollide jaotust andmete väärtusahelas ning sotsiaalmajanduslike mõjutegurite raamistik (sh sotsiaalsed, majanduslikud ja keskkonnapõhised näitajad).
- **Andmete väärtusahel (Data Value Chain)** kirjeldab andmete loomise ja eraldamise ning töötlemise peamisi faase, mis võimaldavad andmete kasutamist, vahetamist ja jagamist (joonis 1). Need on spetsialiseeritud tegevused, mis moodustavad sidusrühmad, kes otseselt või kaudselt osalevad andmeturul ja -majanduses. Andmete väärtusahel on valdavalt väga keerulise reaalsuse lihtsustatud esitus, kuna andmevood on integreeritud erinevate tegevusalade ja majandus-süsteemide vahel, mis ühendab organisatsioone, tööstusharusid ja kogukondi.



Joonis 1. Andmete väärtusahel (Autorite koostatud)

- **Andmete väärtus (Value of Data)** hõlmab andmete sisemist väärtust, mida kasutatakse organisatsioonis ja mida kasutatakse näiteks tõhususe parandamiseks, kulude vähendamiseks, toimivuse hindamiseks või ärimõõdikute juhtimiseks. Antud mõistes andmete väärtus ei seisne turul rahaliste tehingute tulemusel ja seetõttu ei mõõdetata andmeturu komponendina. Antud kontekstis tähendab see kaudset kasu, mida andmekasutajad saavad tänu andmete jagamisele ja taaskasutamisele, mis liigitatakse andmemajanduse väärtuse alla kui "kaudsete järgnevate mõjude" üheks komponendiks. Kuid seda andmete väärtust ei ole võimalik indikaatorina eraldi mõõta.
- **Andmettevõtted (Data Supplier Companies)** on ettevõtted eelkõige andmetarnija rollis, mis tähendab, et nende põhitegevuseks on andmete ja andmetega seotud toodete, teenuste ja

⁵ <https://akit.cyber.ee/term/349-data-1>

⁶ https://courses.cs.ut.ee/MTAT.03.264/2013_spring/uploads/Main/L01_Andmebaasi_definitsioonid.pdf

tehnoloogiate tootmine ja tarnimine. Need ettevõtted moodustavad areneva andmetööstuse. Andmettevõtted võivad olla idufirmad, uuenduslikud VKEd või suuremate ettevõtete tütarettevõtted. Enamik neist pärineb IKT-tegevusvaldkonnast või on hetkel klassifitseeritud sellesse, kuna nende kasutatav põhitehnoloogia on suurandmete tehnoloogia (*big data technology*).

- **Andmeid kasutavad ettevõtted (*Data User Companies*)** on ettevõtted ja organisatsioonid, kes toetuvad andmetele oma käimasolevate tegevuste ja missiooni täitmiseks: see tähendab, et nad loovad ja kasutavad ise oma andmeid, koguvad veebipõhiseid kliendiandmeid, allutavad neid andmeid keerukatele analüüsidele (nt kontrollitud katsetused ja andme- ja tekstikaeve) ning kasutavad õpitut oma äri parandamiseks. Andmekasutajaettevõtted ostavad ja hangivad andmeid ka väljastpoolt oma organisatsiooni, andmeturgudelt, et laiendada ja tõhustada oma tõhusust ja suhtlust oma praeguste ja potentsiaalsete klientidega ning laiendada ja suurendada oma sisemist tõhusust.
- **Andmete valdajad (*Data Holders*)** on avalik-õiguslikud või eraorganisatsioonid, kes omavad või loovad andmeid organisatsioonisiseseks kasutamiseks või väljaspool organisatsiooni taaskasutamiseks, mida levitatakse andmeturgude, andmeplatvormide või andmevahendajate kaudu. Need võivad olla vertikaalselt integreeritud (hõlmavad kogu väärtusahelat alates andmete kogumisest kuni kasutamiseni) või toetuda väärtusahela muude etappide sisseostmiseks spetsiaalsetele vahendajatele.
- **Andmetarbijad (*Data Consumers*)** on toote-, protsessi-, inimressursi-, vara- või osakonna töötajad ja juhid, kes vastutavad muutuste juhtimise või positsiooni säilitamise eest, kus otsuste langetamine sõltub suurel hulgal andmetest ja arusaamadest. Andmetarbijad töötavad otse andmetega ainult osalise tööajaga; nad tuginevad äri- ja tehniliste andmete spetsialistide arusaamadele ja andmete kvaliteedile, mida sageli esitatakse armatuurlaual. Nad osalevad asjakohaste andmeallikate kasutamise ja andmete ootuspärase kvaliteedi tagamises. Nad on otsustajad või sidusrühmad otsustusprotsessis, kus esitatud andmed ja arusaamad määravad tehtud otsuste kvaliteedi. Andmetarbija juhib ettevõtet andmetarneahela kaudu saadud andmete ja arusaamade põhjal
- **Andmekogu (*Data storage*)** on mäluseade andmete talletamiseks, aga ka talletatavate andmekandjate hoiukoht või -seadmestik⁷
- **Andmetöötlus (*Data processing*)** on andmete manipuleerimine arvutiga, mis hõlmab töötlemata andmete teisendamist masinloetavale kujule, andmevoogu läbi protsessori ja mälu väljundseadmesse ning väljundi vormindamist või teisendamist.
- **Andmeanalüüs (*Data analysis*)** on andmete kontrollimise, puhastamise, teisendamise ja modelleerimise protsess, mille eesmärk on leida kasulikku teavet, teha järeldusi ja toetada otsuste tegemist. Tavaliselt hõlmab see erinevatest allikatest (avalikest, omandiõigusega kaitstud või institutsionaalsetest uuringutest) pärinevate andmete ühendamist ning analüütikale tuginemist arusaamade ja väärtuse saamiseks. Andmete valdajad/kasutajad täidavad seda ülesannet või usaldavad väliseid andmemaaklereid ja -pakkujaid.
- **Andmeanalüütika (*Data Analytics*)** on andmete uurimise, puhastamise, teisendamise ja modelleerimise protsess eesmärgiga leida kasulikku teavet, teha järeldusi ja toetada otsuste tegemist. Sh sisaldab tegevust nagu andmekaeve ja analüüsi tehnoloogiate ja tööriistade kogumit, näiteks arvutuslingvistiline, semantiline otsing, loomuliku keele töötlemine, tehisintellekti kasutamine jms. Suurandmete kontekstis viitab andmeanalüütika väga suurte ja keeruliste andmekogumite, mis on sageli pärit mitmest allikast ja erinevates vormingutes,

⁷ <https://akit.cyber.ee/term/3899-data-storage-2>

analüüsile. Suurandmete analüüsis kasutatakse andmetest teadmiste saamiseks masinõppe algoritme, statistilist analüüsi ja andmete visualiseerimise tehnikaid⁸.

- **Andmeteave** (*Data Intelligence*) on immateriaalne vara, mis koosneb kvantitatiivsetest omavahel integreeritud tööriistadest, mis seovad andmed ja andmepõhised mudelid (nt masinõppe algoritmid, mis tuginevad andmebaasidele) selleks, et pakkuda tuge juhtimisotsuste tegemiseks. (Seda mõistet kasutavad Corrado et al (2022)).
- **Andmevahetus** (*Data exchange*) on andmete edastamine erinevate süsteemide, rakenduste või organisatsioonide vahel. See hõlmab teabevahetust kahe või enama üksuse vahel ja võib toimuda eri vormides, näiteks failiedastuse, sõnumipõhise suhtluse või API-de abil⁹.
- **Andmete monetiseerimine** (*Data Monetisation*) on rahalise väärtuse (tulu) määratlemine, mis tekib andmete (andmekogudena või andmepõhiste teenustega ühendatud) kommertsialiseerimisel välistele osapooltele, näiteks andmeturu kaudu. See hõlmab muuhulgas kliendi tegevuse või võimaliku pettuse mõõtmist. Andmete monetiseerimine sisaldub andmeturu väärtuses.
- **Andmeturg (-turud)** (*Data market*) on turg, kus digitaalseid andmeid, andmetooteid ja -teenuseid, andmetega seotud seadmeid ja võrke vahetatakse tootena või struktureeritud andmetest tuletatud teenusena. Mõeldakse ka internetipõhist juurdepääsu pakkumist keerukatele ja mitmekesistele andmeallikatele.
- **Asjade internet** (*IoT, Internet of Things*) on erinevate seadmete kogum, mis on võrku ühendatud ning mis täidavad seal kindlat ülesannet. Seejuures ei ole vaja otseselt inimese sekkumist, kuna seadmed saavad oma ülesande jaoks olulise info kogumise ja vahetamisega iseseisvalt hakkama. Inimese hooleks jääb vaid kogutud andmete põhjal õigete järelduste tegemine ning seeläbi vastavate otsuste vastuvõtmine¹⁰.
- **Digimajandus (digitaalne majandus), e-majandus** (*Digital Economy*) on majandustegevus, mis tuleneb miljarditest igapäevastest võrguühendustest inimeste, ettevõtete, seadmete, andmete ja protsesside vahel. Digitaalrajanduse selgroog on hüperühendus, mis tähendab inimeste, organisatsioonide ja masinate kasvavat omavahelist seotust, mis tuleneb Internetist, mobiiltehnoloogiast ja asjade internetist (IoT)¹¹.

Digimajandus on seotud ka kontseptsiooniga "Tööstus 4.0", mida peetakse laialdaselt tootmissektori digitaliseerimise neljandaks ja järgmiseks etapiks. Tööstus 4.0 on termin, mis antakse tootmisprotsessidele täielikult integreeritud automatiseeritud seadmetega, mis suhtlevad üksteisega. Euroopa Komisjon on välja töötanud digitaalrajanduse ja ühiskonna indeksi (DESI), et jälgida liikmesriikide edusamme digitaalrajandusega tegelemisel ja selle mõju ühiskonnale. See hõlmab viit mõõdet: ühenduvus, inimkapital, interneti kasutamine, digitaaltehnoloogia integreerimine; digitaalsete avalike teenuste integreerimine.

- **Horisontaalne andmevahetus (jagamine)** (*Horizontal Data sharing*) on väärtusahela samas punktis tegutsevate ja sarnaste rollidega organisatsioonide vahel, kes võivad olla nii konkurendid, aga ka mitte: näiteks ettevõtted, kes müüvad sama andmeturu kaudu; või tootjad, kes jagavad andmeid teatud tootmismasinate kasutamise kohta, et parandada võrdlusuuringuid ja prognoositavat hooldust; haiglad või linnavalitsused jagavad teenusetaseme võrdlemiseks andmeid.

⁸ <https://learning.oreilly.com/library/view/big-data-analytics/9781785884696/>

⁹ <https://akit.cyber.ee/term/7037-data-exchange>

¹⁰ <https://digitark.telia.ee/era/mida-asjade-internet-endast-oigupoolest-kujutab/>

¹¹ <https://www2.deloitte.com/mt/en/pages/technology/articles/mt-what-is-digital-economy.html>

- **Informatsioon/ teave** (*Information*) on protsesside väljund, mis võtavad kokku, tõlgendavad või muul viisil esindavad sõnumi sisu, et edastada tähendust. Seetõttu ei ole teave pelgalt andmete sünonüüm.
- **Suurandmed** (*Big Data*) on mahukad andmehulgad, mida ettevõtted ja organisatsioonid praeguses digitaalmaailmas loovad ja kasutavad. Need andmekogumid on suured ja keerulised, mistõttu traditsioonilistest andmetöötlusvahenditest ja -tehnikatest nende käsitlemiseks sageli ei piisa. Suurandmed hõlmavad nii struktureeritud andmeid (nt tehinguandmed andmebaasis), struktureerimata andmeid (nt tekst, pildid ja videod) kui ka poolstruktureeritud andmeid (nt sotsiaalmeedia postitused).
- **Teadmusmajandus** (*Knowledge Economy*) on toodete ja teenuste tootmine, mis põhinevad teadmistemahukatel tegevustel, ning mis aitavad kaasa tehnika ja teaduse kiirenenud arengutempole ning kiirele vananemisele. Teadmispõhise majanduse põhikomponent on suurem toetumine intellektuaalsetele võimetele kui füüsilistele sisenditele või loodusvaradele.
- **Vertikaalne andmevahetus (jagamine)** (*Vertical Data sharing*) – organisatsioonide vahel, kellel on vertikaalses tarneahelas kliendi- või tarnijasuhe, et luua või täiustada tooteid või teenuseid. Näiteks tootjad, kes jagavad andmeid logistikaoperaatorite või jaemüüjatega. Näiteks on kodumasinatootjal huvi jagada andmeid tooteosade rikete kohta oma edasimüüjate ja hooldustarnijatega, et hallata tõhusamalt varuosade laoseisu ja nende logistikat.
- **Väline andmete jagamine** (*External data sharing*) – toimub kolmandate osapooltega, näiteks jagatakse mitme osapoole koostatud koondandmekogumeid väliste sidusrühmadega erinevatel eesmärkidel. See kehtib näiteks privaatsete andmete jagamise kohta sotsiaalsete hüvede või avalike eesmärkide saavutamiseks, näiteks linna ametiasutustega jagatud transpordi- või konnaalteenuste andmestikud, et mõõta liiklus- või heitkoguste andmeid.

2.2. Metoodilised käsitlused

Andmete rahalise väärtuse (monetiseeritud väärtuse, majandusliku väärtuse) ning andmemajanduse mõõtmiseks ja kvantitatiivseks hindamiseks on rahvusvaheliselt tehtud mitmeid empiirilisi uurimusi, mis erinevad nii uuritava objekti, metoodika kui kasutatud andmete vaates. Üldistatult võib öelda, et on kasutatud kolme lähemist:

- 1) **luuakse ülevaade andmeturust ja andmete rahalisest väärtusest (turupõhine ja kulupõhine lähenemine).**
- 2) **hinnatakse andmemajanduse suurust (osalevaid ettevõtteid ja organisatsioone) ja loodavat lisandväärtust sisemajanduse koguprodukti (osatähtsus SKP-st).**
- 3) **kahe eelneva kombineeritud süsteem, kaasates täiendavaid sotsiaalset mõõdet esitavaid indikaatoreid.**

Esimene suund püüab seega hinnata komplekselt andmete väärtust andmeturgudel, kasutades selleks suhteliselt palju hinnanguid, aga ka valitud sektorite ja tegevusalade piires andmettevõtete müügitulu näitajaid. Siinkohal teeb hindamise keerukaks asjaolu, et piiritlemiseks kasutatavad klassifikaatorid ei ole väga täpsed ja andmevahetusel toimub samuti nn vahetarbimine, mille täpset väärtust ei ole teada. Andmete väärtuse hindamiseks on üldistades kolm võimalust:

- turupõhine lähenemine – vaadatakse andmete ostmist ja müümist ning tehingute hindu;
- kulupõhine lähenemine – andmete väärtus leitakse andmete ja informatsiooni tootmiseks

- kuluvate ressursside summana;
- tulupõhine lähenemine – andmete väärtus leitakse, hinnates andmetest tingitud praeguste ja tulevaste rahavoogude nüüdisväärtust (OECD 2019).

Praktikas on kasutatud kõige enam teist ehk **kulupõhist lähenemist**. Selle lähenemise loogiliseks aluseks on see, et ettevõtted ei teeks investeeringuid andmete kogumiseks ja säilitamiseks, kui neil ei ole väärtust. Seega annavad andmete soetamise, töötlemise ja säilitamise kulud alumise piiri andmete väärtusele ehk potentsiaalsele rahavoole, mida ettevõtted prognoosivad, et neist andmetest saab. Taoliselt on andmete rahalist väärtust hinnanud näiteks Kanada statistikaamet (Statistics Canada 2019a; 2019b) või OECD uuringud (Mitchell, Ker, Leshner 2021), (Corrado, Haskel et al 2022). Selle suuna eeliseks on läbipaistev ja reprodutseeritav meetodika, kus hinnatakse andmete rahalist väärtust läbi ettevõtete ja avaliku sektori majandustegevuse (tööjõukulu, investeeringud) ning antakse lõpptulemusena nii hinnang agregeeritult valdkondade kui ka sektorite lõikes. Selle lähenemise puuduseks on aga see, et see ei kata andmete, eriti avaandmete, kõiki sotsiaalmajanduslikke mõjusid, nt mõju inimese elukvaliteedile, haridusele, riigivalitsemisele. Samuti on andmete kajastamine SNA süsteemis ühe vara liigina alles algusjärgus (ISWGNA 2020), mistõttu kokkulepitud meetodika veel puudub.

Teine lähenemine püüab andmemajanduse väärtust leida ja hinnata komplekselt rahvamajanduse arvepidamise süsteemi (SNA) raamistikus. Lisaks võidakse laiemalt kombineerida *ad hoc* viisil erinevaid numbrilisi indikaatoreid, millega mõõdetakse andmete panust majandusse ja ühiskonda, nt mõju hõivele, laste haridusele, tootmisele efektiivsusele, ajakasutusele, tervisele jmt. Siia alla kuuluvad näiteks ka avaandmete väärtuse hindamise raportid, nt McKinsey (2013) või European Data Portal (2020) ja ka Euroopa Liidu kohta tehtud andmeturu ja andmemajanduse hindamise põhjalikud raportid (European Commission 2020, European Commission 2023). Lähenemise puuduseks on sageli see, et analüüsitulemus jääbki mitmemõõtmeliseks, hinnatavate dimensioonide valik on subjektiivne ja lähtub tihti kättesaadavatest andmetest ning selle reprodutseeritavus ja regulaarselt uuendamine on keeruline, sest kasutatakse sageli ühekordsete uuringute andmeid ja subjektiivseid eeldusi.

Eelnevaid märgitud andmemajanduse ja andmete rahalise väärtuse mõõdikute võimalikke süsteeme on kokku võtnud Fleckenstein, Obaidi ja Tryfona (2023) ülevaateartiklis. Tähelepanekutele tuginedes liigitavad eelmärgitud autorid andmemajanduse mudelid kolme rühma:

- turupõhised mudelid, kus arvutatakse andmete väärtus kulude ja tulude alusel;
- majandusmudelid, kus hinnatakse väärtust majandusliku ja avaliku kasu näitajatega;
- kategooriatel ja dimensioonidel baseeruvad andmete rahalise väärtuse mudelid.

Autorid võtavad mudelid kokku tabel 1 kohaselt ja viitavad asjaolule, et iga mudelit saab kasutada erinevates olukordades ja ükski neist lähenemisviisidest ei tööta iga juhtumi puhul. Lisaks märgivad autorid, et kõik mudelid on spekulatiivsed, üksteisega mingil määral kattuvad ja sõltuvad andmete välisest kontekstist. Viimane tähendab, et näiteks on kõik mudelid mingil määral mõjutatud õigusnormidest või valdkonna poliitikast.

Tabel 1. Andmete rahalise väärtuse hindamise raamistikud (Allikas: Fleckenstein, Obaidi ja Tryfona, 2023)

Mudel	Kirjeldus
Turupõhine mudel	Tulu ja kulupõhised andmed: ostu- ja müügiandmed; andmed toodete ja teenuste täiustamiseks; kliendikogemusega seotud andmed toodete ja teenuste täiustamiseks; andmete väärtus kaotaminekul; litsentside väärtus. Ettevõtete väärtus, aktsiaturul põhinev väärtus: ühinemisel, ülevõtmisel, avalikel pakkumistel hinnatud väärtus.
Majanduslik mudel	Majandusliku kasu finantshinnangud; Andmete väärtus avalikuks hüvanguks; Politiika ja õigusliku regulatsiooni mõju
Dimensionaalne	Andmekogumite sisuline kategoriseerimine; Arifunktsioonide hindamine.

Fleckenstein, Obaidi ja Tryfona (2023) kirjeldavad ka võimalikku dimensionaalset andmete väärtuse hindamise mudelit, mis põhineb küsimustel omandiõiguse, kulude, kasulikkuse, vanuse, privaatsuse, andmete kvaliteedi, mahu ja mitmekesisuse vaates. Eesmärgiks oli luua süsteem, mis aitab organisatsioonidel hinnata konkreetsete andmekogumite väärtust konkreetsete kasutusjuhtude jaoks, kasutades arusaadavat dimensioonide kogumit. Dimensiooniline mudel sobib kõige paremini ühe andmekogumi võrdlemiseks teisega, sh avaandmete võrdlemiseks. Sealjuures osutavad autorid tähelepanu sellele, et kuidas erineva vaatenurgaga sidusrühmad võivad dimensioone erinevalt kaaluda ja väärtustada (st teadusasutus, kodanik, valitsus jne). Dimensionaalse andmete hindamise mudeli raamistik eelnimetatud autorite näitel on esitatud Lisas 1.

Lisaks on olemas palju majandusteoreetilisi teadusartikleid, mis püüavad selgitada andmete väärtust ettevõtete jaoks kasutades formaalseid teooriaid (nt mikroökonomikale tuginevad mudelid, kus ettevõtted optimeerivad oma pikaajalist kasulikkust, ja dünaamilised üldise tasakaalu mudelid, mis arvestavad laiemat mõju heaolule). Taolisi mudeleid püütakse kalibreerida konkreetsele majandusele. Need mudelid annavad andmetele küll väärtuse, kas osakaaluna lisandväärtusest või lausa ühiskonna heaolust, kuid praktiliseks kasutamiseks on need siiski liiga abstraktsed. Küll aitavad need mudelid mõista, millisel viisil andmeid saab kasutada (nt ettevõtete täpsemad prognoosid, efektiivsem tootmine, tarbijate ja toodete parem sobitamine, innovatsiooni soodustamine jm) ning lähtuvalt sellest annavad suuna, milliseid aspekte andmemajandusest peaks mõõtma. Näiteid kirjandusest esindavad Farboodi & Veldkamp (2021) toodud mudelid.

Olulisematest allikatest toome esile veel Hanna (2020) digimajanduse hindamine ülevaateartikli. Artiklis käsitletakse digitaalrajanduse tervikliku hindamise ajendeid. Selles kirjeldatakse Maailmapanga algatatud piloothindamise programmi ning kirjeldatakse valitud pilootriikides kasutatavaid hindamisraamistikke, tööriistu ja protsesse. Artikkel annab seega laiema vaate digimajandusest, mis aitab mõiste andmemajanduse hindamise erinevusi.

Eelneva kokkuvõtteks rõhutame, et metoodika loomisel tuleb arvestada Eesti konteksti andmeturu suurust ja omapärasid, aga ka andmevahetuse ja avaandmete olemasolu, siinjuures andmevahetuskivi X-tee ja selle standardite olemasolu. Metoodika loomisel tuleb pöörata erilist tähelepanu selle usaldusväärsusele, võrreldavusele ja korratavusele. Eelpool märgitud allikates jääb detailsuse tase ja seetõttu ka korratavuse võimalus tagasihoidlikuks. Leiame, et erinevatele osapooltele mõistetav andmemajanduse hindamise metoodika peab hõlmama selget dokumentatsiooni ja juhiseid

andmete nii andmete kogumiseks, analüüsimiseks ja arvutamiseks. Metoodika peaks võimaldama võrrelda Eesti andmemajandust teiste riikidega, mis omakorda võib seada piiranguid.

2.3. Metoodikate võrdlus

Järgnevalt viime läbi andmemajanduse uuringutes kasutatud metoodikate kaardistuse ja võrdlemise. Eesmärgiks on saada detailselt ülevaade kasutatud mõõdikutest ja nende kattuvusest koos allikatega erinevates andmete rahalist väärtust ja andmemajandust käsitletavates mudelites.

2.3.1. Euroopa andmeturu seiretöölaud

Kooskõlas esimeste Euroopa andmeturu uuringutega (*The European Data Market (EDM) Monitoring Tool*), nt vastavalt 2017. aasta veebruaris (SMART 2016/0063) või 2020. aastal esitatud Euroopa andmeturu seiretöölaud tulemustega (D2.9 Final Study Report; IDC & Lisbon Council, 2020) on 2021. aasta D1 alusaruandes esitatud täiendatud metoodika. Euroopa andmeturu uuringu 2021–2023 alusaruanne (D1, IDC & Lisbon Council, 2021) on kättesaadav Euroopa Komisjoni andmekogude registreeritud kasutajale.

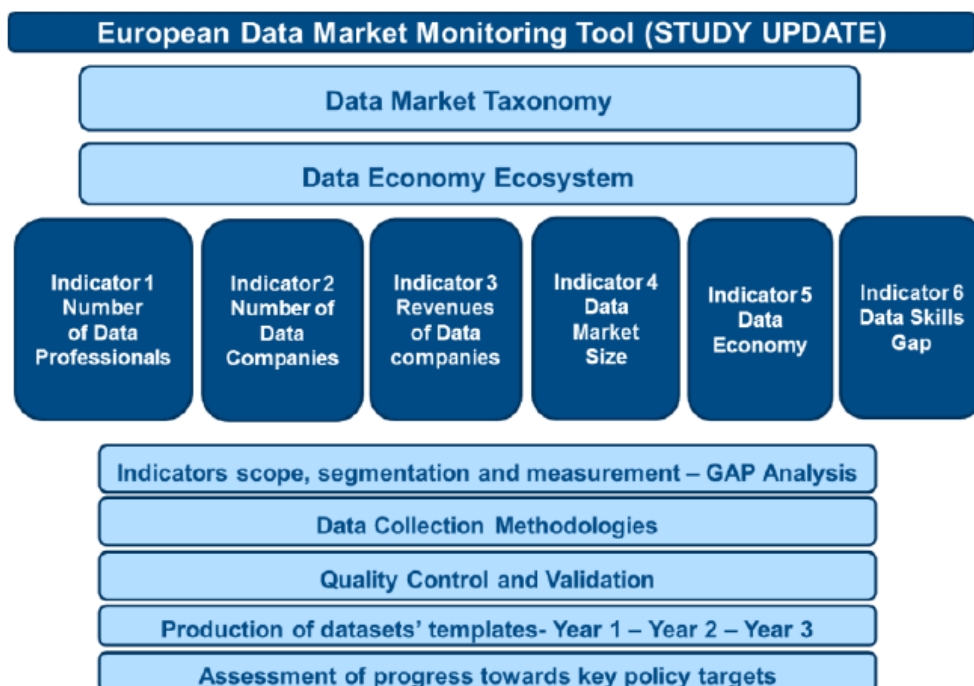
Eelnevalt on koostatud kolm EL-i andmeturu uuringut ja vastavat aruannet ning lisaks kolm EL-i andmemaastiku uuringut (*EU Data Landscape*), kus kolmandas (D4.3) aruandes antakse ülevaade EL-i andmemaastiku andmebaasi versioonist 2020. aasta jaanuari seisuga. Uuringusse on kaasatud kokku 42 riiki (Euroopa Liit-28, Valgevene, Bosnia ja Hertsegoviina, Gruusia, Island, Iisrael, Keenia, Moldova, Norra, Venemaa, Serbia, Šveits, Türgi, Ukraina ja Ameerika Ühendriigid) ja 1556 ettevõtet, millest 311 on määratletud kui võtmetähtsusega ettevõtted andmemaastikul.

IDC (*International Data Corporation*) loodud ja uuendatud Euroopa andmeturu jälgimise tööriista põhikomponendid (edaspidi EDM tööriist) on esitatud alljärgneval joonisel 2 ja selle põhikomponente kirjeldame lisas 2 (tööleht Excelis) ja lähemalt järgmistes osades.

Tööriista aluseks kirjeldatakse andmeturu taksonoomiat järgmiselt:

- andmete ökosüsteem;
- andmeturg;
- andmemajandus;
- andmeoskused;
- sidusrühmade kategooriad.

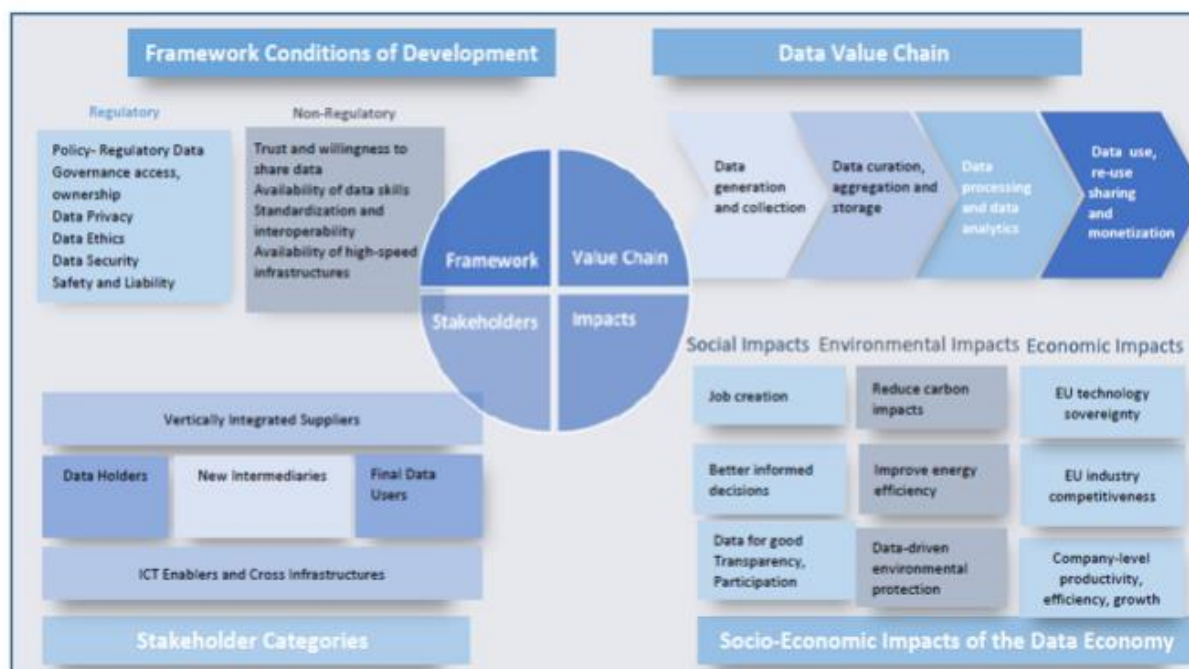
Kuigi mudel keskendub peamiselt indikaatoritele, millega kirjeldatakse andmemajandust, siis väga oluline on mõista mudelis eelnevalt kirjeldatud **andmeökosüsteemi**. Andmeökosüsteemi kaks peamist komponenti on andmete väärtusahel ja sidusrühmade kategooriad. Vaadeldes täpsemalt EDM tööriista valitud indikaatoreid, siis enamus neist on seotud just väärtusahela ja sidusrühmadega.



Joonis 2. EDM tööriista taksonoomia (Allikas: IDC & Lisbon Council, 2021)

Andmeökosüsteem on mudelis defineeritud nelja põhikomponendiga (joonis 3), milleks on

- andmete väärtusahel;
- väärtusahelaga kaasnevad sotsiaalmajanduslikud ja keskkonnamõjude tüpoloogiad (möödikud/indikaatorid);
- sidusrühmade kategooriad;
- raamtingimused ehk keskkonda kujundavad aspektid.



Joonis 3. EDM aluseks olev andmeökosüsteem (Allikas: IDC & Lisbon Council, 2021)

Vaadeldes EDM tööriistas kasutatud indikaatoreid, siis sidusrühmade kategooriatega on seotud andmespetsialistide ja andmeettevõtete mõõdikud. Andmespetsialistid on defineeritud 11 tegevusala järgi (majandustegevuse klassifikaator (NACE rev2) pearühmad) ja ametite klassifikaatori (ISCO-08) koodid alusel. Siinjuures on oluline, et viimase meetodika täiendamise käigus kaasati 245 ametit ja 11 tegevusvaldkonda (A, C, D, E, G, H, J, K, M, P, Q). Ametite klassifikaatori detailne andmestik meile teadaolevalt puudub, kuid meetodikas tuuakse välja, et andmespetsialistide määratluses on kolm põhirühma:

- andmetehnilised spetsialistid**,
- andmeäri spetsialistid** ja
- andmetarbijad**.

Seega antud kolm rühma agregeeritakse **indikaator 1: Andmespetsialistid** alla.

Andmespetsialistid on töötajad, kes koguvad, salvestavad, haldavad ja/või analüüsivad, tõlgendavad ja visualiseerivad andmeid oma tegevuse peamise või asjakohase osana. Andmespetsialistid peavad oskama kasutada struktureeritud ja struktureerimata andmeid, suutma töötada suure hulga andmetega ja olema kursis uute andmebaasitehnoloogiatega.

IDC taksonoomia järgi **andmetehnilised spetsialistid on järgmistes rollides:**

- Andmeadministraator (osaliselt, vt allpool),
- *Business Intelligence*'i arhitekt/arendaja,
- andmebaasi arhitekt,
- andmehoidla spetsialist,
- andmeanalüütik ja
- andmeinsener.

Valdavalt töötavad IT osakondades.

Andmeäri spetsialistid on järgmistes rollides: *Business Intelligence'i* analüütikud ja andmeteادلased, töötades erinevates valdkondades finants- ja planeerimisalaste andmetega, sh keskendudes juhtidele vajalike andmete esitamisega (eelkõige *dashboardi* vormis).

Andmete tarbijad on järgmistes rollides: toote-, protsessi-, inimressursi-, vara- või osakonna töötajad ja juhid, kes vastutavad muutuste juhtimise või positsiooni säilitamise eest, kus otsuste langetamine sõltub suurel hulgal andmetest ja arusaamadest. Andmetarbijad töötavad otse andmetega ainult osalise tööajaga; nad tuginevad äri- ja tehniliste andmete spetsialistide arusaamadele ja andmete kvaliteedile, mida sageli esitatakse armatuurlaual

Euroopa andmeturg (SMART 2013/0063) lõpparuandes viidatakse aruande lisale 2, mis peaks sisaldama ka ISCO koodide loetelu, millega antud mõõdikut sisustakse EDM tööriistas. Käesoleva aruande autoritel õnnestunud seda lisa 2 avalikest materjalidest vahearuande valimiseks ajaks leida.

Indikaator 2: Andmettevõtted (*Data companies*) on organisatsioonid, mis on otseselt seotud andmete tootmise, tarnimise ja/või kasutamisega digitaalsete toodete, teenuste ja tehnoloogiate kujul. Need võivad olla nii andmetarnijate kui ka kasutajate organisatsioonid:

- Andmetarnijate põhitegevuseks on digitaalsete andmetega seotud toodete, teenuste ja tehnoloogiate tootmine ja tarnimine. Nad esindavad andmeturu pakkumise poolt.
- Andmete kasutajad on organisatsioonid, mis genereerivad, kasutavad, koguvad ja analüüsivad intensiivselt digitaalseid andmeid ning kasutavad õpitut oma äritegevuse parandamiseks. Nad esindavad andmeturu nõudluse poolt.

Indikaator 3: Andmettevõtete tulud (*Data companies' revenues*) vastavad kõigi Euroopas asuvate andmetarnijate toodetud andmetega seotud toodete ja teenuste koondväärtusele, sealhulgas eksport väljapoole EL-i. Mõõdikus kasutatakse alajaotusi: 3.1. Andmettevõtete (-sideettevõtete) kogutulud EL-is – indikaator 2 järgi arvatud Andmetarnijate kogutulud; 3.2. Andmettevõtete tulude osakaal – Andmetarnijate tulude ja ettevõtete kogutulude suhe sektorites J ja M. Sisuliselt mõõdetakse ettevõtete müügitulu.

Indikaator 4: Andmeturg (*Data Market*) on turg, kus digitaalseid andmeid vahetatakse „toodete“ või „teenustena“ toorandmete väljatöötamise tulemusena. Andmeturu mõõdik on meetodika detailsema kirjelduse järgi olulises osas seotud indikaatoriga 3: andmettevõtete tulud, millele lisatakse andmete hinnanguline väärtus ettevõtete sisesest investeringust, tarbimisest ning väärtusloomest.

Andmeturg on defineeritud kui turg, kus digitaalseid andmeid vahetatakse „toodete“ või „teenustena“ toorandmete väljatöötamise tulemusena. EDM tööriistas määratletakse selle väärtus kui digitaalsete andmete nõudluse koondväärtust, mõõtmata andmete otsest, kaudset ja indutseeritud mõju majandusele tervikuna. Andmeturu väärtus sisaldab importi (andmetooteid ja -teenuseid, mida ostetakse globaalsel digitaalsel turul tarnijatelt, kes ei asu Euroopas) ning ei sisalda Euroopa andmettevõtete eksporti. Mudel põhineb andmete väljavõtmisel IDC andmebaasidest riistvara-, tarkvara- ja teenuste kulukomponentide kohta, mis kuuluvad andmeturu definitsiooni. IDC andmed on väidetavalt segmenteeritud riikide ja tegevusalade lõikes. Andmeturu tuletamiseks kasutatud tarkvara-, riistvara- ja teenusteturu vastavad osad on saadud IDC uuringutest, mis hõlmavad suurandmeid, IT-kulumustreid ja kavatsusi Euroopa turul, ning andmetarnijate ja -kasutajate uuringust peamistes liikmesriikides.

Indikaator 5: Andmemajandus mõõdab andmeturu üldist mõju majandusele tervikuna. See hõlmab digitaaltehnoloogia abil võimaldatud andmete genereerimist, kogumist, salvestamist, töötlemist, levitamist, analüüside koostamist, edastamist ja kasutamist. Andmemajandus hõlmab andmeturu otsesest, kaudset ja indutseeritud mõju majandusele.

- Otsesed mõjud: on andmetarnijate tekitatud esialgne ja vahetu mõju; need esindavad tegevust, mille võivad esile kutsuda kõik andmetootmisega tegelevad ettevõtted. Kvantitatiivseid otseseid mõjusid mõõdetakse seejärel müüdü andmetoodetest ja -teenustest saadava tuluna, st andmeturu väärtusena.
- Kaudsed mõjud: on andmetarnijate poolt ettevõtte tarneahelas genereeritud majandustegevus. Kaudseid mõjusid on kahte erinevat tüüpi: tagasisuunalised kaudsed mõjud ja edasised kaudsed mõjud.
- Indutseeritud mõjud: hõlmavad kogu majanduses tekkivat majandustegevust teisejärgulise mõjuna.

Mõõdikuna kasutatakse SKP osatähtsust ja väärtust (mõju SKT-le aastatel 2018–2019 (€, miljonit; %)). Aruandes tuuakse esile, et mõjude koosseis ajas muutub (perioodil 2019–2025) indutseeritud mõjude kasuks, paljastades seeläbi andmetele juurdepääsu, andmetoodete ja -teenuste vahetamise ning andmete väärtuse jaotuse mõjud majanduses. Andmemajanduse küpsedes muutuvad mõjud üldisele majandusele (indutseeritud) sama oluliseks kui Euroopa tööstusele (kaudsed mõjud). Andmetööstus jääb selle majanduse peamiseks mootoriks, kuid selle otsene mõju kui osakaal kogumõjudest väheneb.

Indikaator 6: Andmeprofessionaalide (spetsialistide) puudus (*Data Professionals Skills Gap*) Andmeprofessionaalide puuduse näitaja kajastab potentsiaalset lõhet andmeoskusega töötajate nõudluse ja pakkumise vahel, kuna oskuste puudumine võib saada takistuseks andmetööstuse arengule ja andmepõhise innovatsiooni kiirele kasutuselevõtule. See põhineb mudelil, mis tasakaalustab peamised andmeoskuste allikad (haridussüsteemist ja ümberõppest ning muudest kandjatest) hinnangulise nõudlusega (kõikide andmeteevõtete poolt).

Andmeprofessionaalide puudus (lõhe) meetoodika on sama, mida rakendas IDC empiirilises uuringus, et hinnata EL ettevõtluse peadirektoraadi (DG GROW) nimel IKT-oskuste (e-oskused) pakkumise ja nõudluse tasakaalu ELis. Mudel töötati esmakordselt välja 2009. aastal ning sellest ajast alates on seda mitu korda edukalt valideeritud ja uuendatud. Tulemusi on Euroopa Komisjon kasutanud e-oskuste poliitika toetamiseks. Andmeoskused ei ole aga IKT-oskuste alamhulk, mistõttu pakkumise ulatus ja nõudluse dünaamika erinevad IDC väljatöötatud e-oskuste mudelist. (vt Hüsing et al 2015).

Indikaatorite kokkuvõttena järeldame, et **andmeökosüsteemi osaks olevaid raamtingimusi arvestatakse kaudselt**, kuna otseselt mõõdikud puuduvad. Lisaks hinnatakse rahvusvahelist mõju USA, Brasiilia ja Jaapani näitel kasutades indikaatoreid 1, 2, 3 ja 4. EDM tööriista koostajad märgivad, et andmeturu, andmeteevõtete, andmeteevõtete tulude ja andmemajanduse näitajate osas olid peamised allikad:

- Eurostati ettevõtete demograafia statistika Euroopa Liidus, mis käsitleb selliseid aspekte nagu ärimajanduses tegutsevate ettevõtete koguarv, nende sündimus, suremus ja ellujäämise määr (viimati uuendatud: detsember 2019);
- Eurostati iga-aastane struktuurne ettevõtlusstatistika suurusklasside kaupa, mis on VKEde analüüsi peamiseks andmeallikaks (viimane uuendus: detsember 2019);

- IDC üksikasjalikud turuprognosid IT riistvara, tarkvara ja IT-teenuste kohta aastatel 2017, 2018 ja 2019;
- IDC Worldwide Black Book (Standard Edition), kvartaalsed uuendused aastatel 2018–2019. Nn Must raamat esindab IDC kvartalialalüüsi ülemaailmse IKT-tööstuse olukorra ja prognoositava kasvu kohta 54 riigis;
- IDC Euroopa vertikaalsete turgude uuring, 2018 ja 2019;
- IMF World Economic Outlook (WEO) andmebaas, oktoober 2019;
- Rahvusvahelise Tööorganisatsiooni statistika ja andmebaasid (ILOSTAT), jaanuar 2019.

Kokkuvõttes järelname, et EDM tööriista metoodilistes alustes on mitmed hinnangute alused läbipaistmatud, kuna kasutatakse IDC eelnevates uuringutes ja juhendites välja töötatud aluseid või määratud osatähtsuseid, mida otseselt EDM tööriista metoodilistes dokumentides välja ei tooda (eelkõige IDC Worldwide Black Book, IDC üksikasjalikud turuprognosid IT riistvara, tarkvara ja IT-teenuste kohta aastatel 2017, 2018 ja 2019). **Seega on indikaatori 4 ja 5 otsene ülevõtmine Eesti andmemajanduse väärtuse hindamiseks piirangutega ja metoodikas on ilmselt vajalik luua riiklikult vastavad alused.**

Eesti andmebaasidest võimalike andmete allikate kaardistamine näitas, et valdavalt on võimalik EDM tööriista raamistikus kasutatud andmed koguda nii Statistikaameti, aga ka äriregistri, töötamise registri (TÖR) ja OSKA uuringutest (vt Lisa 2). Paralleelselt viisime läbi Soome andmemajanduse mudeli kaardistamise¹². Tulemustest selgus, et Soomes kasutatakse nii siseriiklike andmekogusid, aga ka Euroopa andmeturu uuringu ja digimajanduse ja -ühiskonna indeksis kasutatavaid indikaatoreid. Siinkohale võib olla oluline ja vajadus paralleelselt jälgida digimajanduse ja -ühiskonna indeksit (DESI, *The Digital Economy and Society Index*) andmestikku, mida kasutab Euroopa Komisjon liikmesriikide edusammude jälgimiseks digitaalvaldkonnas alates 2014. aastast. Alates 2023. aastast on DESI integreeritud digitaalse aastakümne poliitika programmiga „2030 Kümnnendiaruanne“ ja seda kasutatakse digitaalsete eesmärkide saavutamise edenemise jälgimiseks¹³.

2.3.2. Andmete väärtuse kulupõhine hindamine

Järgnevalt on toodud ülevaade Corrado *et al* (2022) kitsamast metoodikast, kuidas andmete väärtust hinnata lähtudes andmete kogumisel ja nende väärtustamisel tehtud kulutustele. See ülevaade tugineb nende artikli lisas olevale kirjeldusele. Corrado *et al* (2022) on niinimetatud kulupõhine lähenemine, kus eeldatakse, et investeringud andmetesse peavad olema vähemalt nii suure väärtusega, kui on nende soetamiseks tehtud kulud, mis kajastuvad peamiselt läbi tööjõukulude. Tegemist on täiendatud versiooniga võrreldes Kanada Statistikaameti (Statistics Canada 2019a, 2019b) ja Goodridge, Haskel ja Edquist (2021) lähenemistega, mis puudutab ametialasid, mida kaasati kulude hindamiseks.

Metoodika eelduseks on see, et andmete soetusväärtus sõltub tööjõukulust, vahetarbimisest, kapitalikulude (sh kasum) ja netomaksudest.

¹² Sitra - <https://www.sitra.fi/en/articles/making-it-easier-to-monitor-the-development-of-the-data-economy-weve-put-together-a-tool-to-track-indicators/>

¹³ <https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts>

$$\text{Soetusväärtus} = \text{Tööjõukulu} + \text{Vahetarbimine} + \text{Kapitalikulu} + \text{Netomaksud} \quad (1)$$

Esimese sammuna leitakse tööjõukulu kui andmetega tegelevate töötajate täisajale taandatud arv korda nende tööjõutasu.

$$\text{Tööjõukulu} = \text{Töötajate arv} \times \text{Palk} \times \text{Andmetega töötamise osakaal} \quad (2)$$

Praktikas eeldatakse lihtsustatult, et andmete soetusväärtus on proportsionaalne tööjõukuluga. Seega skaleeritakse tööjõukulud üles teguriga, mis kajastab teisi komponente (nn *blow-up factors*).

$$\text{Soetusväärtus} = \text{Tööjõukulu} \times \text{Skaleerimistegur} \quad (3)$$

Taoline lähenemine eeldab detailsete andmete olemasolu töötajate arvu ja nende töötasu kohta ametialade ja tegevusalade lõikes. Samuti on vaja teha eeldusi (või püüda leida seda küsitlusuuringutest), kui suure osa tööajast kulub töötamisele andmetega selliselt, et tekivad uued andmebaasid või uued andmepõhised otsustusmudelid. Viimaks on vaja anda hinnang teiste kulude kohta, mis tavaliselt võetakse ettevõtlusstatistikast.

Corrado *et al* (2022) jagavad andmetega töötavate inimeste aja andmetesse tehtavate investeeringute vaatest kaheks: 1) andmete kogumine ja nende põhjal andmeladude (*data stores*) loomine ja 2) andmeteabe loomine (*data intelligence*), st andmete integreerimine analüütiliste tööriistadega (nt masinõppe algoritmide loomine jmt). Corrado *et al* (2022) kasutavad oma arvutustes järgmisi ametikohti (vt tabel 2). Samas tabelis on toodud ka erinevused võrreldes sellega, mida kasutavad Kanada statistikaamet (Statistics Canada 2019a, 2019b) ja Goodridge, Haskel ja Edquist (2021). Lühend KS-GHE näitab, kas antud ametialad olid kaasatud ka nendes analüüsid.

Tabelis 2 viimases veerus on toodud ka ISCO ametite loetelu, mida OECD oma raportis "*Measuring the Digital Transformation: A Roadmap for the Future*" (OECD 2019) käsitleb info ja kommunikatsioonivaldkonna spetsialistidena. Nagu näha, siis andmevara loovad ametid on tunduvalt laiemad kui kitsalt IKT spetsialistid. Samas on mõned IKT valdkonna inimesed (nagu juhid ja seadmete hooldajad), kes ei panusta otseselt andmete väärtuse loomisse.

Corrado *et al* (2022) eeldavad, et loodus- ja tehnikateaduste tippspetsialistid (ISCO tase 2) kulutavad oma ajast 10% andmeladude loomisele ja 25% andmeteabele (välja arvatud ametigrupp 216 – arhitektid, planeerijad jt, kus mõlemad on 10%). Kontoritöötajad (ISCO tase 4) panustavad andmete loomisse, aga mitte andmeteadausse ning keskastme spetsialistide (ISCO tase 3) panus sõltub konkreetsest ametist.

Tabel 2. Ametialad, mis panustavad andmebaaside ja andmeteabe loomisse (Allikas: Corrado et al, 2022; OECD, 2019)

ISCO ametigrupp	Aeg andme- ladudele <i>Data Stores</i> (%)	KS- GHE	Aeg andme- teabele <i>Data</i> <i>intelligence</i> (%)	KS- GHE	OECD (2019) IKT spetsialist id
133 – IKT juhid	-	-	-	-	Jah
21 - Loodus- ja tehnikateaduste tippspetsialistid					
211 - Füüsika jm loodusteaduste tippspetsialistid	10	Ei	25	Ei	
212 - Matemaatikud, aktuaarid ja statistikud	10	Jah	25	Jah	
213 - Bioteaduste tippspetsialistid	10	Ei	25	Ei	
214 - Tehnikateaduste tippspetsialistid	10	Ei	25	Ei	
215 - Elektrotehnikainsenerid	10	Ei	25	Ei	Jah
216 - Arhitektid, planeerijad, maamõõtjad ja disainerid	10	Jah	10	Jah	
24 - Äri ja halduse tippspetsialistid					
241 - Finantsvaldkonna tippspetsialistid	10	Jah	25	Jah	
242 - Halduse tippspetsialistid	0	Ei	10	Ei	
243 - Müügi, turunduse ja avalike suhete tippspetsialistid	10	Jah	10	Ei	
25 - IKT tippspetsialistid					
251 - Tarkvara ja rakenduste arendajad ning analüütikud	10	Jah	0	Ei	Jah
252 - Andmebaasi ja arvutivõrgu tippspetsialistid	10	Jah	0	Ei	Jah
3 – Keskastme spetsialistid					
331 - Finants- ja matemaatilise arvestuse keskastme spetsialistid	10	Jah	25	Jah	Jah
351 - IKT protsesside, teenuste ja kasutajatoe tehnikud	10	Jah	0	Ei	Jah
352 - Telekommunikatsiooni- ja ringhäälingutehnikud	-	-	-	-	Jah
4 – Kontoritöötajad ja klienditeenindajad					
4132 - Andmesisestajad	5	Jah	0	Ei	
4227 - Uuringute, sh turu-uuringute käsitlejad	5	Jah	0	Ei	
4312 - Statistilise arvestuse, rahanduse ja kindlustuse kontoritöötajad	5	Jah	0	Ei	
432 - Materjaliarvestuse ja transpordivaldkonna kontoritöötajad	5	Jah	0	Ei	
7 – 742 Elektroonika- ja telekom. seadmete paigaldajad ja hooldajad	-	-	-	-	Jah

Huvitaval kombel ei kajasta Corrado *et al* 2022 näiteks andmete loojatena tervishoiu tippspetsialiste (ISCO ametigrupp 22) või sotsiaalvaldkonna tippspetsialiste, kelle tegevuse tulemusena ometigi tekib suur osa tervishoiu- ja sotsiaalvaldkonna andmeid. Vaikimisi eeldatakse, et andmesisestuse töö teevad ära ametirühma 4 (kontoritöötajad ja klienditeenindajad) kuuluvad töötajad.

Ülalmainitud osakaalud on ainult üks võimalus. Näiteks eeldused, mida oma ettekandes näitasid Bugge *et al* (2023) erinevad oluliselt Corrado *et al* 2022 eeldustest. Kahjuks ei ole Bugge *et al* (2023) puhul kasutatud ISCO ametigruppe, mistõttu nende otse kasutamine on keerulisem.

Table 3 Occupations and time spent, Data and Databases, current method (version 3)

Occupations: Data and databases	Time spent	Occupations: Data and databases, cont.	Time spent
Actuaries	54%	Interviewers, Except Eligibility and Loan	21%
Bookkeeping, Accounting, and Auditing Clerks	17%	Management Analysts	63%
Brokerage Clerks	15%	Market Research Analysts and Marketing Specialists	19%
Budget Analysts	46%	New Accounts Clerks	31%
Credit Analysts	19%	Office Clerks, General	11%
Data Entry Keyers	90%	Operations Research Analysts	53%
Database Administrators	36%	Order Clerks	10%
Designers, All Other	33%	Procurement Clerks	17%
Economists	47%	Social Science Research Assistants	81%
File Clerks	21%	Social Scientists and Related Workers, All Other	58%
Financial Analysts	67%	Sociologists	31%
Financial Managers	21%	Statistical Assistants	30%
Financial Specialists, All Other	20%	Statisticians	69%
Information Security Analysts	20%	Stock Clerks and Order Fillers	4%
Inspectors, Testers, Sorters, Samplers, and Weighers	50%	Tax Examiners and Collectors, and Revenue Agents	11%
Insurance Claims and Policy Processing Clerks	20%	Tax Preparers	11%

Joonis 4. Eeldused andmemajandusega tegelemise osakaalu kohta erinevatel töökohtadel (Allikas: Bugge jt, 2023)

Corrado *et al* 2022 kombineerivad ISCO ametigruppide lõikes vajalikud töötajate arvud ametialade lõikes Euroopa tööjõu-uuringute (*Labour Force Survey*) andmetest ja töötasud töötasu-uuringust (*Structure of Earnings Survey*).

Arvutus koosneks seega järgmistest sammudest:

- 1) leida igal ülal väljatoodud ametialal töötavate inimeste arv (võimalusel majandusharude lõikes);
- 2) leida iga ametiala keskmine töine sissetulek (võimalusel majandusharude lõikes);
- 3) eeldada osakaal, kui suure osa oma ajast töötavad inimesed andmebaaside ja andmeteadusega;
- 4) korrutades sammudes 1, 2 ja 3 leitud suurused, leida teise sissetuleku osa, mis kuulub andmebaasidele ning andmeteadusele;
- 5) skaleerimisfaktori rakendamine.

Kui punktides 1 ja 2 kasutatakse andmeid küsitlusuuringutest või selliseid töötasunäitajaid, mis ei ole kooskõlas sisemajanduse koguprodukti leidmisel kasutatavate tööjõukulunäitajatega, siis võib täiendavalt skaleerida ülal toodud näitajad. (Et summaarset näitajat oleks võimalik võrrelda sisemajanduse koguproduktiga).

Andmaks hinnangut andmete koguväärtusele, siis skaleeritakse ülaltoodud tööjõukulud üles erinevate faktoritega, mis võtavad arvesse seda, et lisaks tööjõule on vaja andmete soetamiseks ja analüüsimiseks vaja teha ka kulutusi tarkvarale ja riistvarale, osa kuludest võetakse välja omaniku poolt kasumina ja osa võib võtta täiendavalt avalik sektor maksudena.

Skaleerimisfaktorid leitakse kui tööjõukulude osakaal kogutoodangus erinevates majandusharudes. Corrado *et al* 2022 kasutavad selleks tegevusvaldkondade J62 "Programmeerimine, konsultatsioonid jms tegevused" ja J63 "Infoalane tegevus" ettevõtete majandustegevuse näitajaid. Täiendavalt eeldavad nad, et tegevusalasine vahetarbimine (üks IT ettevõtte ostab teiselt IT ettevõttelt) on nendes sektorites lausa 50% ja seetõttu topeltarvestuse vältimiseks on see osa müügist välja jäetud.

Ülal leitud vahetulemused ja abiarvutused – töötajate arv andmetega seotud ametialadel eri sektoritest, nende keskmine töötasu, IKT ettevõtete majandusnäitajad – on kasutusel ka eraldiseisvate indikaatoritena, vt ptk 2.3.1, mistõttu tuleb need näitajad leida nagunii mõlema meetodi korral. Küll võib erineda konkreetsetelt ametite ja tegevusalade loetelu, mida eri riigid või meetoodikad kasutavad andmetega töötavate spetsialistide defineerimisel või IKT valdkonna kitsendamisel.

Eesti puhul on võimalik töötajate arv leida kas tööjõu-uuringu andmetest või töötamise registri andmetest (TÖR). Töötasu andmed saab samuti leida analoogselt kas küsitlusuurinutest või eelistatult maksu- ja tolliameti TSD andmetest, kus kasutatakse keskmisi töötasusid ametialade lõikes (nt sidudes töötamise registrist (TÖR) ametiala koodi ja TSD andmetest töötasu). Koostöös Statistikaametiga otsustati kasutada TÖRi ja TSD ühendatud andmeid töötajate arvu ja keskmise töötasu (ehk kokku tööjõukulu) leidmiseks.

Saamaks võrreldavaid tulemusi Eesti kohta teiste riikidega, siis tasub esmalt kasutada samu eeldusi ajakulu kohta, mis on toodud Corrado et al 2022. Seejärel peab saadud osakaale valideerima Eesti ekspertidega ning vastavalt sellele tegema täiendavad arvutused Eesti kohta.

Corrado et al (2022) lähenemine on väga sarnane Kanada Statistikaameti (Statistics Canada 2019a, 2019b) ja Goodridge, Haskel ja Edquist (2021) lähenemistega. Erinevus seisneb selles, milliseid ameteid on kaasatud analüüsi. Corrado et al 2022 väidavad, et Kanada Statistikaameti meetoodika ei kaasa piisavas ulatuses andmeteabe (*data intelligence*) loomisega tegelevaid ametikohti ja seetõttu pigem alahindavad andmete väärtust majanduses. Mõlemad, nii Corrado et al (2022) kui Kanada Statistikaameti lähenemine püüab aga anda täpsema hinnangu andmetele ja andmeteabele majanduses, sest ettevõtete endi poolt raporteeritud andmeladude ja andmemudelite kui immateriaalse vara väärtus on väidetavalt alahinnatud rahvamajanduse arvepidamise statistikas.

3. TEISTE RIIKIDE KOGEMUS

Järgnevalt esitame ülevaate sellest, kuidas andmete majanduslikku väärtust (*Economic Value of Data*) on valitud riikides mõõdetud ja mõju hinnatud.

3.1. Ühendkuningriigid

Ühendkuningriikides on varasemalt mõõdetud suurandmete, asjade interneti (IoT) ning nendega kaasnevate analüütiliste lahenduste kasutamisest tulenevat väärtust majanduses (Hogan et al., 2016). Hindamismudel analüüsib kolme mõjuallikat: mõju tõhususele, innovatsioonile ja ettevõtete loomisele. Modelleerimisel järgiti kolme-etapilist lähenemist:

- 1) suurandmete ja asjade interneti kasutamise eeliste välja selgitamine, kasutades küsitluse vastuseid¹⁴ ning statistikaameti (ONS, *Office for National Statistics*) ja ettevõtlus-, innovatsiooni- ja kutseoskuste ministeeriumi (BIS) andmeid;
- 2) praeguse ja tulevase tehnoloogia vastuvõtu määra tuletamine küsitluse põhjal;
- 3) suurandmetest ja asjade internetist saadava kasu arvutamine kogu sektorile ja majandusele.

Tõhususe suurenemine, mida uuriti ettevõtete küsimustikuga, põhineb käibe suurenemisel ja kulude vähenemisel, mille ettevõtted on saavutanud tänu investeringutele suurandmete ja asjade interneti valdkonnas. Tõhususe kasv peegeldab ettevõtete võimet saavutada vähemaga rohkem, mis parandab nende tulemuslikkust ja suurendab nende võimet luua majandusele lisaväärtust.

Ettevõtete innovatsioon tuleneb innovatsiooni ja teadus- ja arendustegevusse tehtavate investeringute tulu suurenemisest ning selliste investeringute tootlikkuse mõjust. Selleks on modelleeritud teadus- ja arendustegevuse kulutuste suurenemist ja sellest tulenevat mõju tulevase pikaajalise müügi suhtes igas tööstusharus. Nii suurandmed kui ka asjade internet võivad suurendada tegevuse tõhusust, mis omakorda suurendab kasumlikkust. Suurenenud kasumlikkuse tulemusena saavad ettevõtted reinvesteerida kasumit tootearendusse. Selleks määrati iga tööstusharu puhul kindlaks, kui suur osa tegevusbilansi brutoülejäagist (kasumile ligikaudne näitaja) kulub teadus- ja arendustegevusele¹⁵, kasutades ONSi tootmise ja tarbimise tabelleid¹⁶. Seejuures kasutati kirjanduse analüüsi selleks, et määrata kindlaks tööstuse toodangu elastsus ehk reageerimisvõime teadus- ja arendustegevusse tehtavate investeringute suhtes. Eeldati, et need tööstusharud, mis näitavad suuremat elastsust, saavad suuremat tulu teadus- ja arendustegevuse investeringutest ning seega on tõenäolisem, et nad saavad kasu suurandmete ja asjade interneti kasutamisest.

Ettevõtete **loomise** mõju tuleneb kasumi suurenemisest, mis tuleneb ettevõtete ökonoomsuse suurenemisest ja signaalist, mida see annab potentsiaalsetele uutele turule tulijatele. Need uued

¹⁴ Ettevõtetest küsiti mitmeid küsimusi nende praeguse ja tulevase eeldatava suurandmete ja asjade interneti kasutamise kohta.

¹⁵ Sealjuures on uurimis- ja arendustegevuse kulud määratletud kui kulutused uurimis- ja arendustegevusele ning reklaamile ja turu-uuringutele.

¹⁶ <https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables>

ettevõtted loovad töökohti ja lisavad majandusele lisaväärtust. Ettevõtete loomisest saadav kasu põhineb suurandmete ja asjade interneti mõju kvantifitseerimisel ettevõtete asutamise arvule. Selle väärtuse arvutamisel hinnatakse VKE-de tootlikkust ning võetakse arvesse suurandmete analüüsist ja asjade internetist tuleneva tõhususe suurenemisest tulenevate turule tuleku tõkete vähenemise mõju.

Bennet Institute'i (2020) poolt avaldatud raportis viitab „väärtus“ sotsiaalse heaolu majanduslikule kontseptsioonile ehk kogu ühiskonna heaolule. Väärtus tuleneb andmetest, kui ettevõtted loovad töökohti või muutuvad tootlikumaks, kui valitsused osutavad tõhusamaid avalikke teenuseid, kui keskkond on puhas ja mitmekesine ning kui inimesed elavad õnnelikumalt ja tervislikumalt. Andmete väärtuse hindamiseks on raportis pakutud mitmeid viise.

Turupõhise hindamise puhul kasutatakse mõõdupuuna aktsiaturgu. Andmepõhised ettevõtted kalduvad näitama kõrgemat börsiväärtust, mis näitab konkurentsieelist. Lisaks sellele on arenenud andmeanalüüsi võimekusega varustatud ettevõtted oma tööstusharu teistest ettevõtetest edukamad. **Tulupõhise hindamise** puhul püütakse hinnata nii praegust kui ka tulevast andmetest tulenevat tulu. Siiski on nendel hinnangutel piirangud, sest sageli ei suudeta kajastada laiemat kasu, mida andmed toovad organisatsioonile või süsteemile, mis ulatub kaugemale finantstehingutest. **Kulupõhised hindamised** seevastu keskenduvad andmete loomise ja säilitamisega seotud kuludele. See hõlmab rahvamajanduse arvepidamises üksikasjalikult kirjeldatud kulusid, eelkõige andmete koostamisega seotud tööjõukulusid.

Mitteturupõhised hindamised uurivad avatud andmete majanduslikku väärtust, kasutades andmekogumite finantsmõju. Need hinnangud koostatakse sageli, kasutades tingimuslikke hindamismeetodeid, mille puhul küsitletakse organisatsioone nende valmisoleku kohta maksta andmetele juurdepääsu eest. Erinevatest vaatenurkadest lähtudes saavad andmeid haldavad organisatsioonid sisemiselt kasu andmetele juurdepääsust saadava tulu või andmete jagamisest tulenevate kulude vähenemise kaudu. Vahendajad, kes kasutavad andmeid toodete ja teenuste loomiseks, aitavad kaasa majanduskasvule, luues töökohti ja uuendusi.

Üks viis, kuidas andmete jagamise ja kasutamise väärtust saab mõõta, on vaadata millist kasu andmekogujad ja -vahendajad ning kasutajad võivad sellest saada. Andmete avalikustamine, jagamine, kombineerimine teiste andmekogumitega loob tavaliselt palju suuremat väärtust laiemale majandusele kui üksikutele andmeomanikele, kellel on õigus andmeid kättesaadavaks teha (Stiglich, 2021).

3.2. Holland

Hollandi statistikaamet (Bondt et al., 2021; Bondt & Mushkudiani, 2021; Statistics Netherlands, 2021) on andmete väärtuse ja mõju hindamisele lähenenud rahvamajanduse arvepidamise seisukohast, mis võimaldab andmeid käsitleda investeringutena. Andmete väärtuse hindamisel oli esimene osa dokumendianalüüs, mis kaardistas teadmiste hetkeseisu valdkonnas. Teine osa sisaldas tutvumist Euroopa IKT uuringu andmetega¹⁷, et mõista ettevõtete andmekasutust (nt suurandmete rakendamine jne). Kolmandas osas viidi läbi intervjuud ettevõtjatega, et saada ülevaade, kuidas

¹⁷ https://ec.europa.eu/eurostat/cache/metadata/en/isoc_e_esms.htm

andmeid ettevõtetes kasutatakse ja kuidas nende andmete väärtus ettevõttes kajastub (nt kasumiaruandes). Uuringu neljandas osas anti esialgne hinnang andmete väärtuse kohta Hollandi ettevõtetes. Arvutuste tegemise aluseks oli Kanada statistikaameti väljatöötatud lahendus (Statistics Canada 2019a; 2019b). Kasutatava meetodi eesmärk on hinnata kulusid, mis on vajalikud nende andmete hankimiseks tehtud kulutuste põhjal. Arvutusi tehakse kolme kategooria (andmed, andmebaasid ja andmeteadus) kulude kohta. Need kulud jagunevad samuti kolmeks osaks: tööjõukulud, muud kulud ja kapitali soetus. Mudelis on olulised muutujad näiteks ametiala, töötasu ja majandusharu. See hõlmab erinevate ametikohtade puhul kindlaksmääramist, kui suur osa palgakuludest kulub andmetele. Andmete väärtuse hinnangu jõudmiseks lisatakse järgmise sammuna muud kulud ja kasumiprotsent.

Digital Realty koostatud raportis (Digital Realty, 2018a, 2018b) on andmemajandust määratletud kui finants- ja majanduslikku väärtust, mis luuakse suurandmete salvestamise, otsimise ja analüüsiga ning keeruka tarkvara ja muude vahendite abil. Andmemajanduse ulatuse hindamiseks töötati välja sektoripõhine mudel, kus iga riigi jaoks välja töötatud mudel jaotati 19 sektori kaupa, mis põhines standardsetel tööstusklassifikatsioonidel¹⁸. Madalmaade puhul jaotati mudel ka piirkondliku geograafilise jaotuse järgi. Kaasati andmeid, nagu ettevõtete käive ja lisandväärtus (*gross value added*), tööhõive sektorite ja ametite kaupa, tootlikkus, tööturuandmed (nt hinnangud tehniliste oskuste puudujääkide ja oskuste puuduse kohta võtmesektorites), ettevõtete demograafilised andmed (andmemajanduse teenuste osutamisega seotud äriüksuste ja asutuste arv). Andmeallikateks olid iga riigi keskse statistikaameti, Eurostati, OECD aga ka ettevõtlusuuringutest saadud andmed. Mudel näitab ajaloolist trendi peamiste majandusnäitajate kohta, mille abil on võimalik tuvastada seoseid tarbimiskulutuste, äriinvesteeringute, tööturu näitajate ja rahvusvahelise kaubanduse vahel.

3.3. Soome

Sitra (Soome Innovatsioonifond) on välja töötanud andmemajanduse seirevahendi¹⁹ (Rastas, 2023). Seirevahend võimaldab näiteks hinnata, kas ettevõtted kasutavad tehisintellekti; kuidas ettevõtted kasutavad andmeid protsesside parandamiseks või teenuste ja töö automatiseerimiseks; kuidas mõjutavad andmed ettevõtte tulemuslikkust; kui palju investeeritakse Soomes teadus-, arendus- ja innovatsioonitegevusse (TAI), ning kui palju kasutatakse ELi rahalisi vahendeid.

Lisaks on Soomes välja pakutud lahendus riiklike avaandmete²⁰ väärtuse ja mõju hindamiseks (Koski & Elinkeinoelämän tutkimuslaitos, 2015), mille järgi võib andmete avamise mõjud jagada majanduslikuks ja sotsiaalseks. Majanduslikku mõju saab hinnata ettevõtete, elanike, avaliku sektori ja majanduse kui terviku tasandil. Kõige märkimisväärsem võimalik majanduslik mõju ettevõtetele on kasv ja tootlikkuse suurenemine tõhususe paranemise ning uute toodete ja teenuste arendamise kaudu, mida võimaldavad avatud andmed. Andmete avamine võib kaasa tuua ka uute ettevõtete

¹⁸ Lisaks Hollandile tehti mudelid ka Ühendkuningriikide, Saksamaa ja Iirimaa kohta.

¹⁹ Seirevahend Power BI aruandena on saadaval siin:

<https://app.powerbi.com/view?r=eyJrjoiNmFmMjZyZctODFjMC00MWUyLWlWl3NmItMmQ2YzYyZGY5M2IzliwidCl6ImlwMTlyMGYxLWFiOTQtNDkzNi04NmQ3LWYyYjQwZjM4ZjRiNyIsImMiOj9>

²⁰ <https://www.avoindata.fi/>

asutamise. Elanike jaoks tuleneb kõige olulisem majanduslik kasu tasuta juurdepääsust andmetele, selle asemel et kasutada tasulisi ressursse. Avaliku sektori organisatsioonide jaoks pakub andmete avamine kulude kokkuhoidu ja võimalust parandada teenuste osutamise tõhusust. Andmete avamise majanduslikku mõju võib täheldada ka majanduse kui terviku tasandil, kui avatud andmeid saab kasutada piisavalt laialdaselt erinevates majandussektorites. Oodatavad peamised mõjud ja näitajad avatud andmete mõju süstemaatiliseks jälgimiseks ja hindamiseks on esitatud esitatud tabelis 3.

Tabel 3. Avaandmete majandusliku mõju hindamise raamistik

Sektor	Oodatav mõju	Indikaator	Potentsiaalne andmeallikas
Ettevõtted	uued tooted ja teenused; uued ettevõtted	avaandmetel loodud uued tooted ja teenused (arv); avaandmetel loodud müügitulu osatähtsus kogu müügitulust	küsitlus
	kasv (müügitulu); tootlikkus	müügitulu muutus; töötajate arv muutus; lisandväärtuse muutus töötaja kohta	äriregister
Kodanikud	aja ja raha kokkuhoid (sääst)	avaandmete hulk x varasem kulu; eeldatav ajakokkuhoid	küsitlus
Avalik sektor	kulude kokkuhoid; tootlikkus	töötundide muutus x töötunni hind; teenuse väljund seoses tehtud ressursside kuluga	automaatne monitooring
Riiklik majandus	majanduskasv, tootlikkus	muutus lisandväärtuses, sh SKP-s	statistikaamet

Lisaks on Soomes hinnatud transpordiandmete mõju (Fintraffic, 2023b, 2023a). Transpordis võib andmetest saadav kasu olla nii otsene (tulu andmepõhistest lahendustest) kui ka kaudne (nt vähem õnnetusi, parem täpsus). Transpordiandmete hindamiseks ja kasutusest ülevaate saamiseks saab viia läbi (iga-aastaseid) uuringuid või intervjuusid ettevõtete andmete ja andmete kasutamise kohta (andmeid kasutavate ettevõtete arv, suurus, tootlikkus, tooted ja teenused, andmete mõju); hinnata andmeprojektide väärtust; mõõta väliste andmeallikate arvu või API päringute arvu; jälgida teenuse kasutajate arvu ja ekstrapoleerida turuosa põhjal üldturgu.

Soome kohta on võimalik täiendavalt leida infot järgmistest allikatest:

- *Digital Economy* mõõtmine (*Quantitative assessment of the size of the digital economy*): <https://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-106.pdf>
- Soomes *Data Economy* programm: <https://www.businessfinland.fi/en/for-finnish-customers/services/programs/data-economy#:~:text=About%20us-The%20goal%20of%20the%20Data%20Economy%20program%20is%20to%20encourage,up%20to%20135%20million%20euros.>
- Finland's Economic Opportunities from Data Centre Investments: <https://copenhageneconomics.com/publication/finlands-economic-opportunities-from-data-centre-investments/#:~:text=Finland%20reaps%20large%20economic%20opportunities,EUR%202.3%20billion%20by%202025>
- Metodoloogia: <https://copenhageneconomics.com/wp-content/uploads/2022/05/annex-finlands-economic-opportunities-from-data-centre-investments.pdf>

3.4. Rootsi

Rootsis on samuti hinnatud andmete avalikustamise ehk avaandmete sotsiaalmajanduslikku mõju ja väärtust. Kombineerides võrreldavate riikide kogemustel põhineva metaanalüüsi ja juhtumiuuringutel põhineva mõjuanalüüsi, on kirjeldatud avaandmete mõju konkreetsetele sektoritele (Ekström & Johannesson, 2020a, 2020b). Metaanalüüsiga keskenduti neljale valdkonnale: georuumilised andmed, aadressiandmed, ettevõtte andmed ja geoloogilised andmed. Uuringu hetkel olid need andmekogumid Rootsis suurel määral piiratud tasudega ning eeldati, et andmete avamisel võiks olla märkimisväärne sotsiaalmajanduslik mõju. Samuti on mitmed Rootsi naaberriigid juba pakkunud avatud juurdepääsu sarnastele andmekogudele. Paljudes nendes riikides on läbi viidud sotsiaalmajanduslikke uuringuid nii enne kui ka pärast andmetele avatud juurdepääsu võimaldamist, mille tulemused viitavad mitmetele eelistele, mis kaasnevad andmete avalikustamisega.

Metaanalüüsi hinnanguliste väärtuste täienduseks viidi läbi juhtumipõhine mõjuanalüüs, kus keskenduti sellele, kuidas siiani tasulised andmekogumid võiksid lisada väärtust viies erinevas sektoris (põllumajandus, IKT, finants ja kindlustus, ehitus, avalik sektor). Iga juhtumiuuringu raames on arvatud andmete makromajanduslik väärtus. Arvutused põhinevad näiteks aja kokkuhoiul, tulude suurenemisel ja turutõrgete (*market frictions*) vähenemisel. Seejärel võrreldi tulemusi varasemate uuringutega ja anti üldine hinnang lisandväärtusele igas sektoris.

Avaandmete väärtust on leitud ka läbi andmetega seotud kulude (Apanasevic, 2021), nagu ülevõtmiskulud, infrastruktuurikulud, hooldus- ja tegevuskulud ning saamata jäänud tulu (vt tabel 4).

Tabel 4. Avaandmete mõju hindamise raamistik

Kulutused	Väärtus/ kasu kategooria	Riskid ja barjäärid
Kohanemiskulud	demograafilise kontrolli suurenemine	juriidilised barjäärid
Infrastruktuurikulud	poliitikas osalemine	riskid
Hooldamis ja opereerimiskulud	sotsiaalne väärtus	tehnilised barjäärid
Müügitulu vähenemine	opereerimise tõhusus ja efektiivsus	organisatsioonilised barjäärid
	majanduskasv	majanduslikud barjäärid

Avatud andmete sotsiaal-majanduslikust mõjust ülevaate saamiseks kasutati tasuvusanalüüsi aspekte koos avaandmetest tuleneva kasu ja riskide kategooriatega. Intervjuude andmete analüüsimisel liigitati vastajate vastused tabelis 4 esitatud andmete mõju kategooriate alusel. Kõige sobivamaks peeti kvantifitseerida tegevuse tõhusust, mida määratleti kui aja kokkuhoiuga seotud tõhususe kasvu.

4. JUHTIMISLAUD JA ANDMEMAJANDUSE VÄÄRTUS

4.1. Andmemajanduse väärtuse hindamise arvutuskäik kulupõhise meetodiga

Andmemajanduse väärtuste hindamiseks kulupõhise meetodiga kasutame Corrado et al (2022), Bugge, Amegble, Sinclair (2023), Statistics Canada (2019a, 2019b) ja Goodridge, Haskel ja Edquist (2021) lähenemist, mida on pikemalt kirjeldatud peatükis 2.3.2.²¹

Andmemajanduse väärtuse hindamine koosnes järgmistest sammudest:

- 1) andmetega potentsiaalselt tegelevate töötajate arvu ja nende töötasu leidmine ametiala ja tegevusalade lõikes;
- 2) eelduste tegemine osakaalu kohta iga ametiala lõikes, mis kulub andmetega töötamiseks;
- 3) andmetega tegelemise tööjõukulu leidmine;
- 4) tööjõukuludele muude kulude lisamine;
- 5) tulemuste agregeerimine.

4.1.1. Töötajate arvu ja töötasu leidmine

Andmetega töötavate töötajate arvu, töötasu ja tööjõukulude leidmiseks kasutatakse andmeid, mis on saadud Statistikaametist päringu tulemusena märtsis 2024. Andmepäring on toodud Lisas 3. Andmed võttis Tartu Ülikooli palvel välja Statistikaameti analüütik Kadri Rootalu. Andmete allikaks on palga andmebaas, mis ühendab töötamise registri (TÖR) ning Tulu- ja sotsiaalmaksu, kohustusliku kogumispensioni ja töötuskindlustusmaks deklaratsiooni (TSD) andmestiku ja mis on kasutusel ka Statistikaameti poolt keskmise brutokuupalga statistika arvutamisel. (vt ka <https://www.stat.ee/et/avasta-statistikat/valdkonnad/tooelu/palk-ja-toojoukulu>).

Käesolevas raportis kasutatav andmestik on sees töölepingud (töösuhte koodid: 1 – tööleping, 507 – viisa alusel tööleping, 800, 801 ja 802 – laevapere töölepingud) ja avaliku teenistuse töötajad ja kõrgemad riigiametnikud. Välja on jäetud võlaõigusseaduse alusel sõlmitud lepingud, sest nende puhul on keeruline hinnata nii töötamise koormust kui ei ole ka teada enamasti ametinimetust.

²¹ Eesti Statistikaamet kasutab ettevõtete poolt oma tarbeks toodetud tarkvara väärtuse hindamiseks samuti palgaandmeid, täpsemalt ettevõtete IT-töötajate palgaandmeid, kuid see meetodika ei olnud käesoleva raporti kirjutamise ajaks veel avalik. (Allikas: Robert Mürsepp, Statistikaameti rahvamajanduse arvepidamise tiimi juht)

Sellest tulenevalt on andmetega töötavate inimeste töö kogus alahinnatud.²² Töötasude leidmisel on välja jäetud ettevõtted, kellel pole Eesti registrikoodi (ja kes ei oma seetõttu tegevusalala).

Potentsiaalselt andmete loomisega tegelevate ametite puhul on kasutatud järgmiseid ametite ISCO koode: 1211, 1346, 2111, 2113, 2114, 2120, 2131, 2132, 2133, 2141, 2142, 2143, 2144, 2145, 2146, 2149, 2151, 2152, 2153, 2161, 2162, 2163, 2164, 2165, 2166, 2411, 2412, 2413, 2421, 2422, 2423, 2424, 2431, 2432, 2433, 2434, 2511, 2512, 2513, 2514, 2519, 2521, 2522, 2523, 2529, 2631, 2632, 3311, 3312, 3313, 3314, 3315, 3511, 3512, 3513, 3514, 4132, 4227, 4311, 4312, 4321, 4322, 4323, 7543

Tulemused on agregeeritud 2- või 4-kohalise koodiga ametite tasemele tegevusalade lõikes EMTAK ühekohalise või kahekohalise tasemel. Tulemuste esitamisel on välja jäetud need read, kus tulemus oli konfidentsiaalne, st kus see jäi alla 3 töötaja või tööandja.

Tulemuste esitamisel on töötajate arv taandatud täistööajale kasutades koormust. Koormus osas on kasutusel keskmine koormus kvartali jooksul (ei arvesta töötatud päevi, vaid ainult TÖRi märgitud töötaja määr ja selle muutusi). Kui töötaja määr pole täidetud, on selleks pandud 1 (ehk täiskohaga töötamine). Seega näitab koormuste summa ligikaudu täisajaga töötavate inimeste arvu. Statistikaametist saadud andmetes on tulemused eraldi kvartalite lõikes, sest statistikaamet kasutab ise kvartaalset perioodi palgaandmete esitamisel. Antud analüüsi käigus esitati tulemused kvartalite keskmisena.

Tulemuste tõlgendamisel peab arvestama, et Sise- ja Kaitseministeeriumi ning nende haldusala töötajate andmeid ei ole võimalik kasutada. Seega alahindab käesolev lähenemine avaliku sektori poolt loodud andmete väärtust tööjõukulu meetodil.

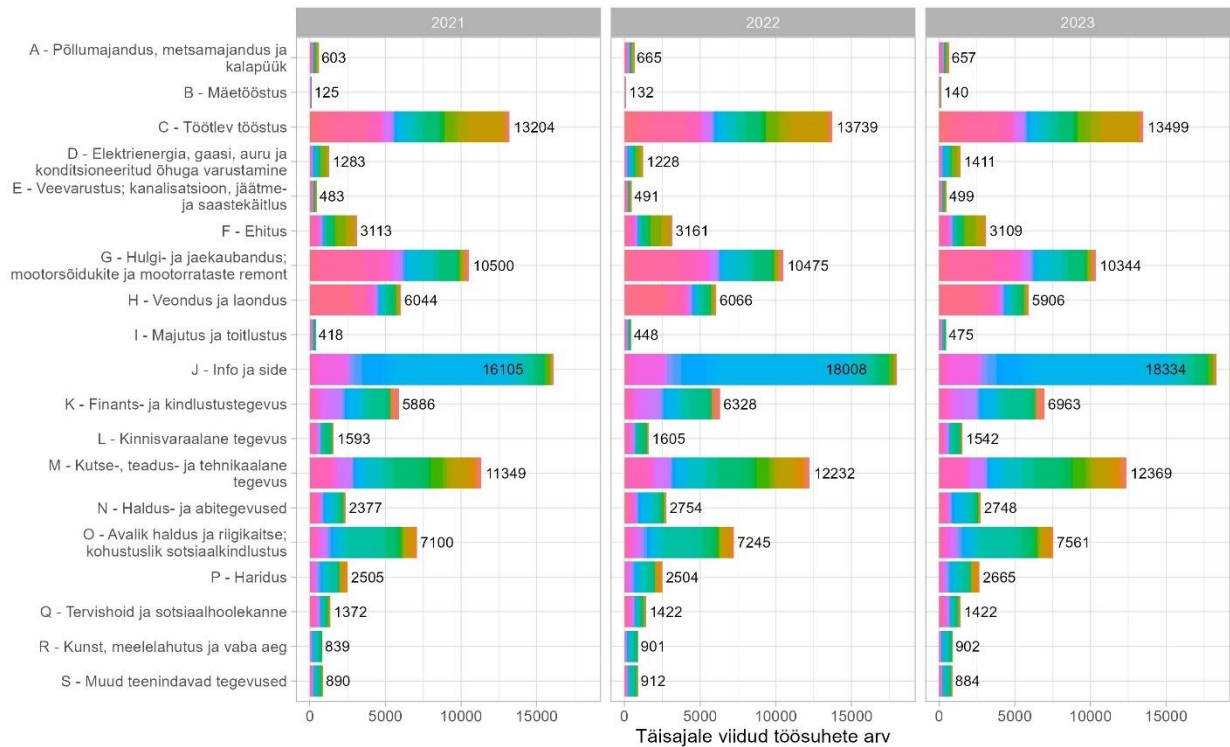
Alljärgnevalt on esitatud andmemajanduse väärtuse leidmisel kasutatud sisendandmed 2021-2023. aastate kohta nii ISCO teise kui neljanda taseme ametialade lõikes.

Oodatult on kõige enam töökohti, mis võiksid olla seotud andmemajandusega info ja side valdkonnas. Sellele järgnevad töötlev tööstus ning kutse-, teadus- ja tehnikaalane tegevus. Oluline on tähele panna, et selleks, et leida summaarne tööaeg, mis nad tegelevad just andmetega, siis see korrutatakse läbi andmetega tegelemise osakaaluga.

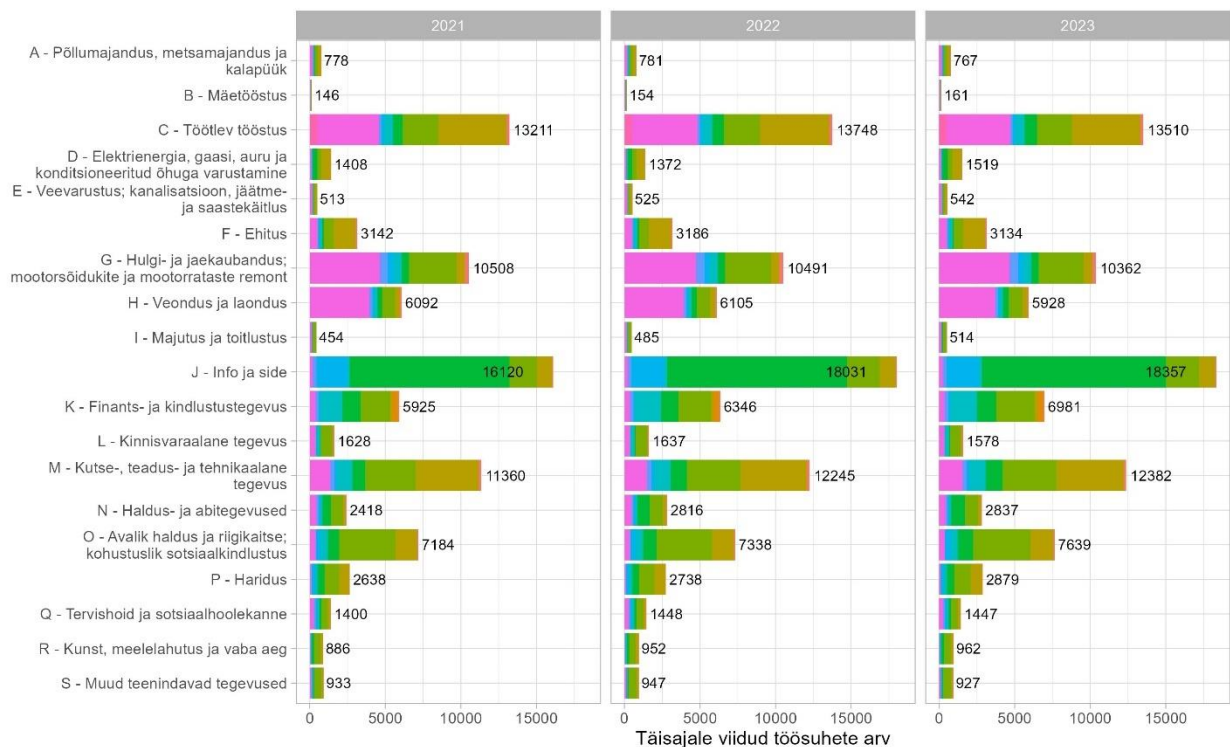
Kui andmeid pärida ISCO neljanda taseme ametialade lõikes, siis tulenevalt andmekaitsest läheb osa töökohtade kaotsi võrreldes teise taseme ametialadega. Kokku on kaotus keskmiselt 1.1% töökoormusest. Suhteliselt suurim kaotus on väikestes tegevusalades. Näiteks põllumajanduse metsanduse ja kalapüügi valdkonnas on see lausa 14%, mäetööstuses 13%, majutuses ja toitlustuses 8% jne. Kui andmeid pärida ISCO teise taseme ametialade lõikes kuid EMTAK 2 tasemel, sooviga täpsemalt hinnata, millistel tegevusaladel on enam andmemajandusega tegelevaid töötajaid, siis andmekaitsest põhjustel on töökoormus 2023. aastal keskmiselt 1,6% väiksem. Seega kui soovitakse kasutada TÖRi ja TSD andmeid tegevusalade ja ametialade lõikes, on soovitatav koondarvutused lasta valmis teha statistikaametis, et vähendada andmekaitsest piirangutest tulenevat ebatäpsust.

²² EMTA andmete järgi moodustas 2023. aastal VÕS lepingute summaarne brutotasu 2,4% kogu töötasust. Samas tegevusvaldkonnas J63 Infoalane tegevus ulatus see 5,7%ni töötasust, mistõttu võib arvata, et andmetega tegelevate inimeste töökoormus on alahinnatud.

Eesti andmemajanduse turuväärtuse hindamise meetodika



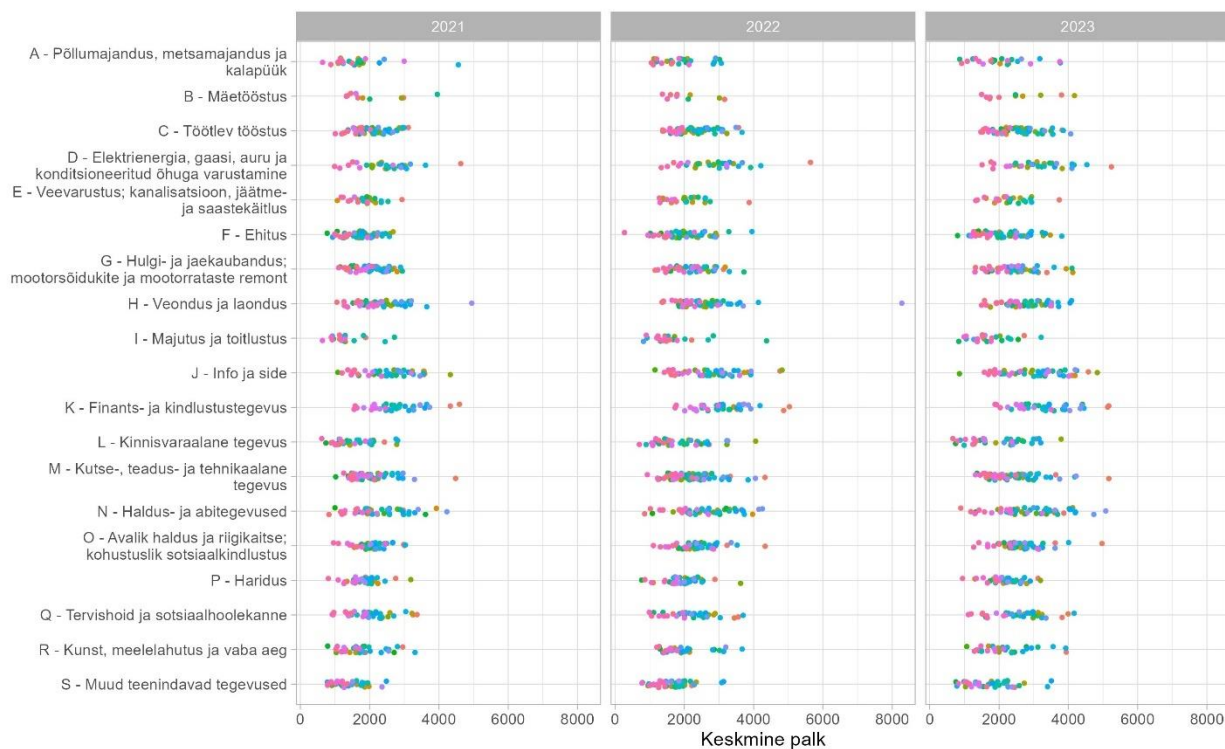
Joonis 5. Täisajale viidud töösuhete arv (4-kohaline ISCO) tegevusalade lõikes üle aastate, mis potentsiaalselt tegelevad andmemajandusega



Joonis 6. Täisajale viidud töösuhete arv (2-kohaline ISCO) tegevusalade lõikes üle aastate, mis potentsiaalselt tegelevad andmemajandusega

Märkus: erinevad värvid näitavad erinevaid ISCO ameteid

Täistööajale viidud töösuhete arv korrutatakse keskmise töötasuga, et saada hinnang kogu tööjõukulule. Keskmise töötasu on leitud kui aastane brutotöötasu summa jagatuna töökoormuste summaga ametialade ja tegevusvaldkondade lõikes. Mõlemad näitajad on saadud eelmainitud andmepäringu tulemusena statistikaametilt.



Joonis 7. Keskmine täisajale taandatud brutopalk ametialade (4-kohaline ISCO) ja tegevusalade lõikes üle aastate

4.1.2. Andmetega töötamise osakaal

Järgmine oluline samm andmete väärtuse leidmisel on hinnangu andmine osakaalule, mis ülal leitud töökoormusest kulub andmetega tegelemisele. Seda osakaalu võib tõlgendada ka kui töötajate osakaal antud ametialal, kes tegelevad täies mahus andmetega (ja teised töötajad ei tegele siis andmetega üldse). Käesolevas raportis võetakse osakaalud varasematest rahvusvahelistest uuringutest, sest meile teadaolevalt Eesti kohta taolisi uuringuid läbi viidud ei ole.

Eeldused, millised ametid panustavad andmete ja andmebaaside loomisse on toodud järgmises tabelis, mis on ka Exceli failis, mis on sisendiks töölauale.

Veerg Bugge et al (töölaua failis on tunnuse nimetus timecan2023) püüab kajastada osakaale, mida kasutatakse Kanada Statistikaametis (vt ettekannet Bugge, Amegble, Sinclair 2023). Et tolles ettekandes ei ole toodud ISCO ametikoode, siis see vastavus tugineb nimetustele. Veerg Corrado et al 2022 (töölaua failis timecorrado2022) kasutab OECD raportis toodud kasutatud osakaale (vt

Corrado, Haskel, Iommi, Jona-Lasinio 2022). Vaikimisi on Eesti arvutustes võetud maksimaalne ülal toodud eeldustest. Seda eeldust saab kergesti töölaua sisendandmeid muutes.

Alljärgnevas tabelis on esitatud ametid, kus vähemalt üks valitud kaaludest on positiivne. Veerus Kaal (töölaua failis kaal) on arvutustes kasutatav osakaal, mida saab kasutaja ise Exceli algfailis muuta. Töölauale on vajalik laadida sisendiks fail, kus peab olema kaks tunnust: iscode (ISCO neljanda taseme kood) ja kaal.

Tabel 5. Andmetega töötamise osakaal ametialade lõikes

ISCO	Nimetus	Bugge et al 2023	Corrado et al 2022	Kaal
1211	Finantsjuhid	21	21	21
1346	Finantsvahendus- ja kindlustusteenuste juhid	21	21	21
2111	Füüsikud ja astronoomid	0	10	10
2113	Keemikud	0	10	10
2114	Geoloogid ja geofüüsikud	0	10	10
2120	Matemaatikud, aktuaarid ja statistikud	54	10	54
2131	Bioloogid, botaanikud, zooloogid jms tippspetsialistid	0	10	10
2132	Põllumajanduse, metsanduse ja kalanduse nõuandjad	0	10	10
2133	Keskkonnakaitse tippspetsialistid	0	10	10
2141	Tööstus- ja tootmisinsenerid	0	10	10
2142	Ehitusinsenerid	0	10	10
2143	Keskkonnatehnika tippspetsialistid	0	10	10
2144	Mehaanikainsenerid	0	10	10
2145	Keemiainsenerid	0	10	10
2146	Mäetööstuse, metallurgia jms valdkondade tippspetsialistid	0	10	10
2149	Tehnikateaduste tippspetsialistid, mujal liigitamata	0	10	10
2151	Elektriinsenerid	0	10	10
2152	Elektroonikainsenerid	0	10	10
2153	Telekommunikatsiooniinsenerid	0	10	10
2161	Ehitusarhitektid	0	10	10
2162	Maastikuarhitektid	0	10	10
2163	Toote- ja rõivadisainerid	0	10	10
2164	Linna- ja liiklusplaneerijad	0	10	10
2165	Kartograafid ja maamõõtjad	0	10	10
2166	Kujundajad ja multimeediadisainerid	0	10	10
2411	Raamatupidamise tippspetsialistid	17	10	17
2412	Finants- ja investeerimisnõustajad	20	10	20
2413	Finantsanalüütikud	67	67	67
2421	Juhtimis- ja organisatsioonianalüütikud	63	63	63
2422	Strateegiate väljatöötajad	10	10	10
2423	Personali ja karjääri tippspetsialistid	10	10	10
2424	Töötajate koolituse ja täiendõppe tippspetsialistid	10	10	10
2431	Reklaami ja turunduse tippspetsialistid	19	10	19
2432	Avalike suhete tippspetsialistid	10	10	10
2433	Tehnika- (v.a IKT) ja meditsiinitoodete müügi tippspetsialistid	10	10	10
2434	IKT-lahenduste müügi tippspetsialistid	10	10	10

2511	Süsteemianalüütikud	10	10	10
2512	Tarkvaraarendajad	10	10	10
2513	Veebi- ja multimeediaarendajad	10	10	10
2514	Rakenduste programmeerijad	10	10	10
2519	Tarkvara ja rakenduste arendajad ning analüütikud, mujal liigitamata	10	10	10
2521	Andmebaaside projekteerijad ja administraatorid	10	10	10
2522	Süsteemadministraatorid	10	10	10
2523	Arvutivõrkude tippspetsialistid	10	10	10
2529	Andmebaasi ja arvutivõrgu tippspetsialistid, mujal liigitamata	10	10	10
2631	Majandusteadlased	47	0	47
2632	Sotsioloogid, antropoloogid jms tippspetsialistid	31	0	31
3311	Väärtpaperi- ja valuutamaaklerid ning -vahendajad	10	10	10
3312	Krediidi- ja laenuhaldurid	19	10	19
3313	Raamatupidamise keskastme spetsialistid	10	10	10
3314	Statistilise ja matemaatilise arvestuse jms keskastme spetsialistid	10	10	10
3315	Hindajad ja kahjuhindajad	10	10	10
3511	IKT operaatorid	0	10	10
3512	IKT kasutajatoe tehnikud	0	10	10
3513	Arvutivõrkude ja -süsteemide tehnikud	0	10	10
3514	Veebitehnikud	0	10	10
4132	Andmesisetajad	21	5	21
4227	Uuringute, sh turu-uuringute käsitlejad	21	5	21
4311	Raamatupidamise kontoritöötajad	17	17	17
4312	Statistilise arvestuse, rahanduse ja kindlustuse kontoritöötajad	15	15	15
4321	Laoarvestuse kontoritöötajad	5	5	5
4322	Tootmisarvestuse kontoritöötajad	5	5	5
4323	Transpordivaldkonna kontoritöötajad	5	5	5
7543	Toodete (v.a toidud ja joogid) testijad	50	0	50

Andmetega töötamisele kuuluva täisajale viidud töötajate arvu leidmiseks korrutame täisajale viidud töösuhete arvu läbi ülaltoodud eeldusega, kui palju ajast inimesed tegeleva andmete või andmeteabe loomisega.

$$\text{Andmetega töötajate tingarv}_j = \sum_k \text{täistööajale viidud töötajate arv ametis}_{j,k} \times \text{kaal}_k \quad (5)$$

4.1.3. Andmetega töötavate inimeste näitajad

Kombineerides punktides 4.1.1 toodud hõivatute arvud ametialade lõikes ja punktis 4.1.2 toodud eeldused andmetega töötamise kohta Nii töötas ülaltoodud eelduste kohaselt 2023. aastal andmetega 91 429 töötajat ja kui kaaluda see läbi andmetega töötamise tööajaga, siis 13 970 inimest.

Saadud tulemus esitatakse töölaual nii tegevusalade lõikes kui kokku. Töölauale võib lisada ka kaalumata näitajad või erinevad suhtarvud (vt tabel 6).

Tabel 6. Andmetega töötavate täisajale viidud töötajate arv ja ajaga kaalutud

Aasta	Rea number	2021	2022	2023
Andmetega tegelevate töötajate arv	(1)	85787	90315	91429
Ajaga kaalutud töötajate arv	(2)	12813	13564	13970
Kasv võrreldes eelmise aastaga	(3)		5.9%	3.0%
Kas võrreldes aastaga 2021	(4)		5.9%	9.0%
<i>Taustaandmed</i>				
IKT sektori täistööajale taandatud töötajate arv *	(5)	30578	34100	
Kokku täistööajale taandatud töötajate arv *	(6)	442934	463720	
<i>Suhtarvud</i>				
Andmetega tegelevate töötajate arvu suhe IKT sektori töötajate arvu	(7) = (1)/(5)	280.6%	264.9%	
Andmetega tegelevate ajaga kaalutud töötajate arvu suhe IKT sektori töötajate arvu	(8) = (2)/(5)	41.9%	39.8%	
Suhe täistööajale taandatud töötajate arvu	(9) = (1)/(6)	19.4%	19.5%	
Suhe täistööajale taandatud töötajate arvu	(10) = (2)/(6)	2.9%	2.9%	

Allikas: * Statistikaamet, tabel EM001: Ettevõtete majandusnäitajad tegevusala ja tööga hõivatud isikute arvu järgi.

Märkus: hallil taustal on kuvatud näitajad, mis on esitatud töölaual prototüübil.

Töölaual lehel kuvatakse lehel "Tööhõive" väljal "Andmetega tegelev tööhõive, 2023" näitajat 13970, mis on ümardatud 14.0 tuhandeni, samuti viimast kasvu (3%) ja kumulatiivset kasvu alates 2021. aastast (9.0%). Samuti kuvatakse võrdlus täisajale taandatud töötajate arvuga kokku ja IKT sektoris. Töölaual võib kuvada ka ajaga skaleerimata töötajate arvu (rida (1)) ja suhtarve IKT sektori töötajate arvu suhtes (rida (7)) või kogu töötajate arvu suhtes (rida (9)).

Samuti kuvatakse töölaual tegevusalade lõikes ajaga kaalutud töötajate arv sektorite lõikes (tabel 7, veerud (4), (5), (6)). Samuti leitakse tegevusalad, kus aastased kasvud on kõige kiiremad ja kõige aeglasemad (veerg (7)).

Tabel 7. Andmetega töötavate täisajale viidud töötajate arv ja ajaga kaalutud töötajate arv tegevusalade lõikes, 2021-2023

Tegevusala	Andmetega tegelevate töötajate arv			Ajaga kaalutud töötajate arv			Kasv	
	Aasta	2021	2022	2023	2021	2022	2023	2023/2022
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
A - Põllumajandus, metsamajandus ja kalapüük		603	665	657	96	102	101	-0.9%
B - Mäetööstus		125	132	140	16	17	18	5.1%
C - Töötlev tööstus		13204	13739	13499	1832	1896	1875	-1.1%
D - Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine		1283	1228	1411	212	218	244	12.2%
E - Veevarustus; kanalisatsioon, jäätme- ja saastekäitlus		483	491	499	69	70	73	4.5%
F - Ehitus		3113	3161	3109	433	438	427	-2.4%
G - Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont		10500	10475	10344	1348	1339	1315	-1.9%
H - Veondus ja laondus		6044	6066	5906	624	638	645	1.0%
I - Majutus ja toitlustus		418	448	475	72	80	84	5.0%
J - Info ja side		16105	18008	18334	1970	2209	2265	2.5%
K - Finants- ja kindlustustegevus		5886	6328	6963	1293	1438	1691	17.6%
L - Kinnisvaraala tegevus		1593	1605	1542	296	302	291	-3.6%
M - Kutse-, teadus- ja tehnikalaane tegevus		11349	12232	12369	1830	1989	2024	1.8%
N - Haldus- ja abitegevused		2377	2754	2748	367	410	414	0.9%
O - Avalik haldus ja riigikaitse; kohustuslik sotsiaalkindlustus		7100	7245	7561	1280	1328	1386	4.3%
P - Haridus		2505	2504	2665	501	487	527	8.3%
Q - Tervishoid ja sotsiaalhoolekanne		1372	1422	1422	239	255	252	-1.3%
R - Kunst, meelelahutus ja vaba aeg		839	901	902	143	147	146	-0.4%
S - Muud teenindavad tegevused		890	912	884	192	202	192	-4.8%
Kokku		85787	90315	91429	12813	13564	13970	3.0%

Märkus: hallil taustal on kuvatud näitajad, mis on esitatud töölaua prototüübil.

4.1.4. Andmetega tegelemisele kuluv tööjõukulu

Eelnevates punktides 4.1.1. ja 4.1.2 on toodud vajalikud sisendid andmetega tegelemisele kuluva tööjõukulu leidmiseks. Tööjõukulu kokku võib leida, korrutades kokku punktides 4.1.1 leitud täiskoormusega töötavate inimeste arvu ja nende keskmise töötasu koos sotsiaalmaksuga. Statistikaametilt küsiti tegelikult ka aastane brutopalgat ja tööjõumaksude summa kokku, seega puudub vajadus komponente eraldi välja arvutada. Neid on kasulik eraldi vaadata vaid selleks, et aru saada

Alternatiivne võimalus, mis vähendab andmekaitselisi probleeme, on leida aastane brutotööjõukulu koos sotsiaalmaksuga kokku ISCO 4-kohaliste ametite ja EMTAK esimese tasandi tegevusalade lõikes. Seega kasutatakse antud töös mitte eelpool toodud valemit (2) vaid juba tööjõukulu koondtulemust

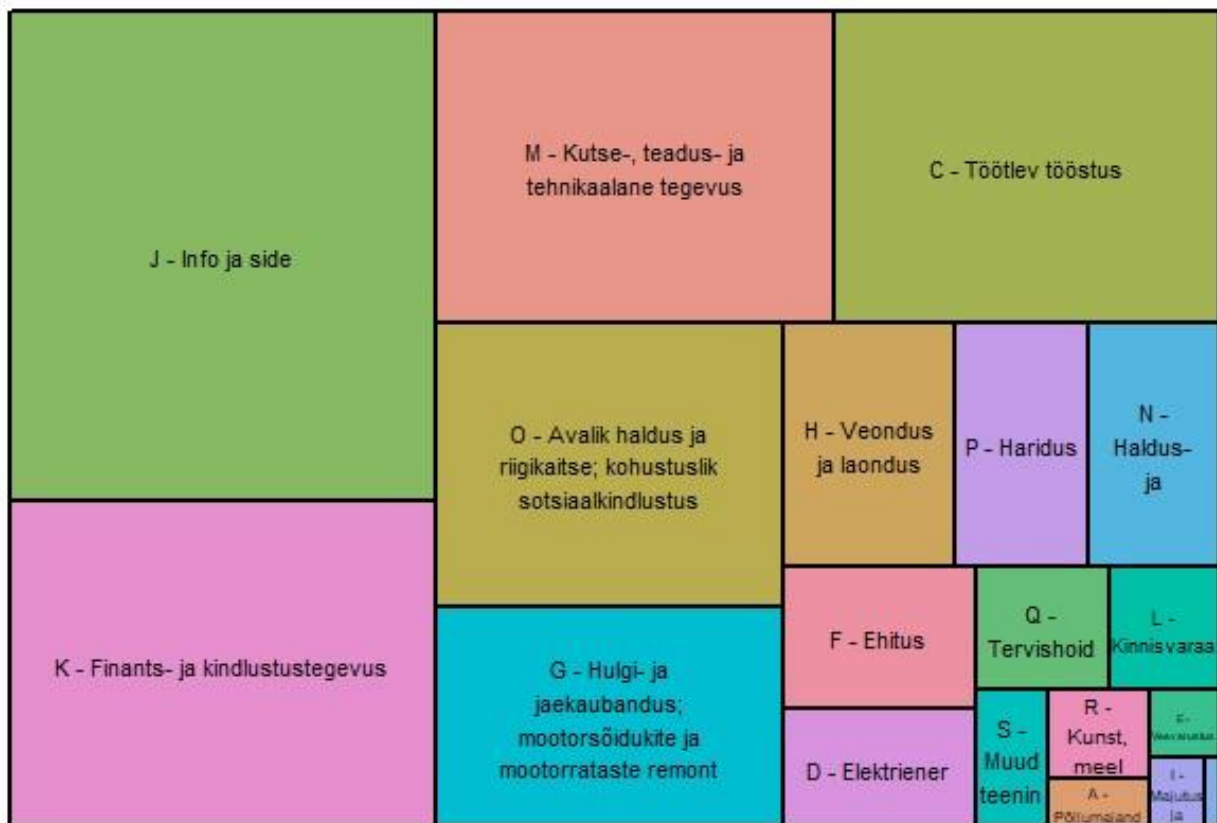
$$Tööjõukulu = Tööjõukulu kokku valitud ametitel \times Andmetega töötamise osakaal \quad (6)$$

Tehniliselt leitakse tööjõukulu tegevusalal j liites kokku brutopalgaga ja sotsiaalmaksu (mis sisaldab ka tööandja töötuskindlustusmaksed) ametis k tegevusalal j , korrutades läbi andmetega töötamise osakaaluga ametis k .

Töölaual kasutatakse statistikaametist saadud päringu tulemust "tulemused_SA.xls" lehelt osa2_4k.

$$Tööjõukulu tegevusalal_j = \sum_k (brutopalg_summa_{k,j} + sm_summa_{k,j}) \times kaal_k \quad (7)$$

Tööjõukulu andmetega tegelemisel on samas vaid vahetulemus, mida võib esitada ka eraldi soovi korral. Alljärgneval joonisel on toodud andmetega tegelemise tööjõukulu tegevusalade lõikes 2023. aastal.



Joonis 8. Andmetega tegelemise summaarne tööjõukulu tegevusalade lõikes 2023. aastal

Töölaua uuendamisel peab arvestama, et kui lisanduvad uued potentsiaalsed ametid, mis võiksid tegeleda andmetega, siis tuleb uuesti pärida ka brutopalgaga ja sotsiaalmaksu aastased kogusummad nende ametite kohta.

4.1.5. Muude kulude lisamine

Muud andmetega tegelemise kulud leitakse Eesti puhul skaleerimistegur abil. Selleks kasutatakse ettevõtete majandusnäitajate andmestikku (Statistikaameti andmebaasi tabel EM001 "Ettevõtete majandusnäitajad") aastate 2021. ja 2022. kohta tegevusalade J62 "Programmeerimine,

konsultatsioonid jms tegevused" ja J63 "Infoalane tegevus" jaoks. Lisas 4 on toodud detailsem päring andmete kohta.

Muudest ostetud teenustest ja müügi eesmärgil ostetud teenustest eeldame nii nagu Corrado et al 2022, et 50% on sektorisesed IT-teenused ja lähevad kohe edasi müügiks ilma täiendava tööjõupanuseta. Ülejäänud müügitulu peab siis katma tööjõukulud, põhivara kulumi ja kasumi. Praegune meetodika eeldab, et pool sisseostetud kuludest kategooriates "Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused" ja "Muud ostetud teenused" on tehtud omakorda IKT sektorist.

Selle tulemusena moodustavad tööjõukulud 2021-2022. aastal keskmisena 43% omatoodangust, mis teeb skaleerimisteguriks 2.32. Töölaud kasutab vaikimisi aastate 2020-2022 keskmisi näitajaid, mis annavad skaleerimisteguriks 2.27. Viimase arvutuskäik on illustreeritud tabelis 8.

Tabel 8. Skaleerimisteguri arvutus

Rida	Näitaja	Tegevusala	2020	2021	2022	Kokku
(1)	Müügitulu, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	1805928	2504284	3832206	8142418
(2)	Müügitulu, tuhat eurot	Infoalane tegevus	300105	388203	491305	1179613
(3)	Muud ostetud teenused, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	736677	1375698	2513577	4625951
(4)	Muud ostetud teenused, tuhat eurot	Infoalane tegevus	125206	166159	224745	516109
(5)	Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	28223	43337	27230	98790
(6)	Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused, tuhat eurot	Infoalane tegevus	2856	1613	2420	6889
(7)	Tööjõukulud, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	679831	823521	1096322	2599674
(8)	Tööjõukulud, tuhat eurot	Infoalane tegevus	97518	103830	144769	346117
	Skaleerimisfaktori arvutus ridade põhjal	$\{(1) + (2) - 0.5*[(3) + (4) + (5) + (6)]\} / \{(7) + (8)\}$				2.27

Allikas: Statistikaameti andmebaasi tabel EM001

Saadud tulemus on sarnane sellele, mida Corrado et al 2022 kasutasid oma töös: Suurbritannia 2.1, Prantsusmaa 2.2, Saksamaa 2.3, Hispaania 2.4, Rootsi 2.8 ja Itaalia 3.4.

4.1.6. Koondtulemus

Skaleerimistegurit rakendatakse kõikidele töajookuludele, mis on seotud andmetega, sõltumata ametist või tegevusalast ühte moodi.

$$\text{Andmete väärtus tegevusalal}_j = \text{Töajookulu tegevusalal}_j \times \text{skaleerimistegur} \quad (7)$$

Sellisel leitud andmete väärtuse koondtulemused tegevusalade lõikes on järgmises tabelis. 2023. aasta andmete väärtus on kokku 1310 mln ehk ca 3.5% SKPst.

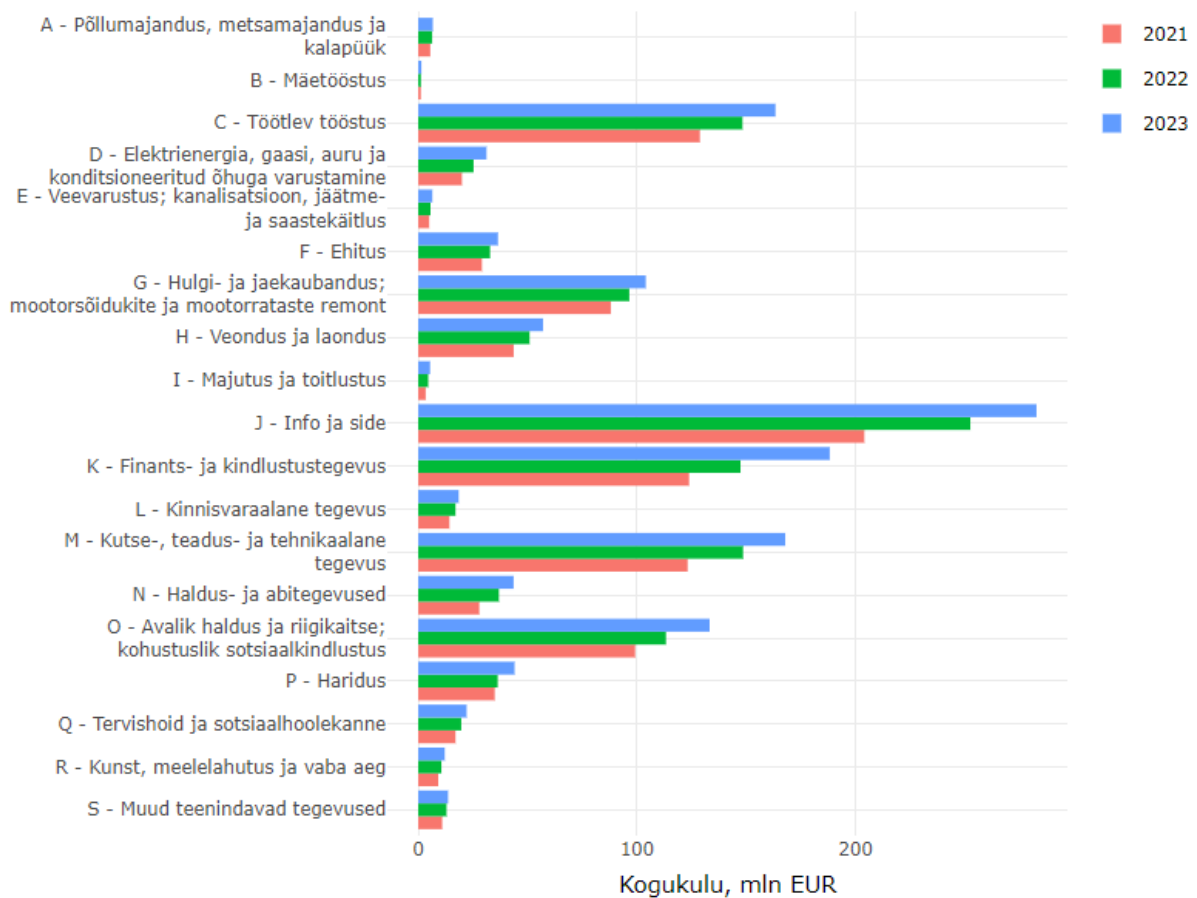
Tabel 9. Andmete väärtus ja tööalal esitatud näitajad

Aasta	2021	2022	2023
Töajookulu kokku (mln eurot)	426	502	576
Skaleerimistegur	2.27	2.27	2.27
Andmete väärtus kokku (mln eurot)	969	1142	1310
Aastane kasv		17.9%	14.7%
Kasv alates 2021		17.9%	35.2%
Keskmine kasv aastas			16.3%
<i>Taustatunnused</i>			
Sisemajanduse koguprodukt (SKP)* (mln eurot)	31169	36011	37682
IKT sektori käive müügitulu * (mln eurot)	5608	7417	
<i>Suhtarvud</i>			
Andmete väärtus, % SKP	3.11%	3.17%	3.48%
Andmete väärtus, % IKT sektori müügitulusse	17.3%	15.4%	

Allikas: * Statistikaamet, tabelid RAA0012 ja EM001

Märkus: hallil taustal on kuvatud näitajad, mis on esitatud tööalua prototüübil.

Täiendavalt saab leida andmete väärtuse eraldi ka tegevusalade lõikes. Alljärgnev joonis esitab tulemuse tegevusalade ja aastate lõikes. Sama tulemus on esitatud ka tööalual, eraldi 2023. aasta kohta tulpdiaagrammina ja üle aastate joondiagrammina.



Joonis 9. Andmemajandusele omistatud kogukulud

Saadud tulemusi saab võrrelda kas sisemajanduse koguproduktiga, kulutustega teadus- ja arendustegevusele või IKT valdkonna kulutustega.

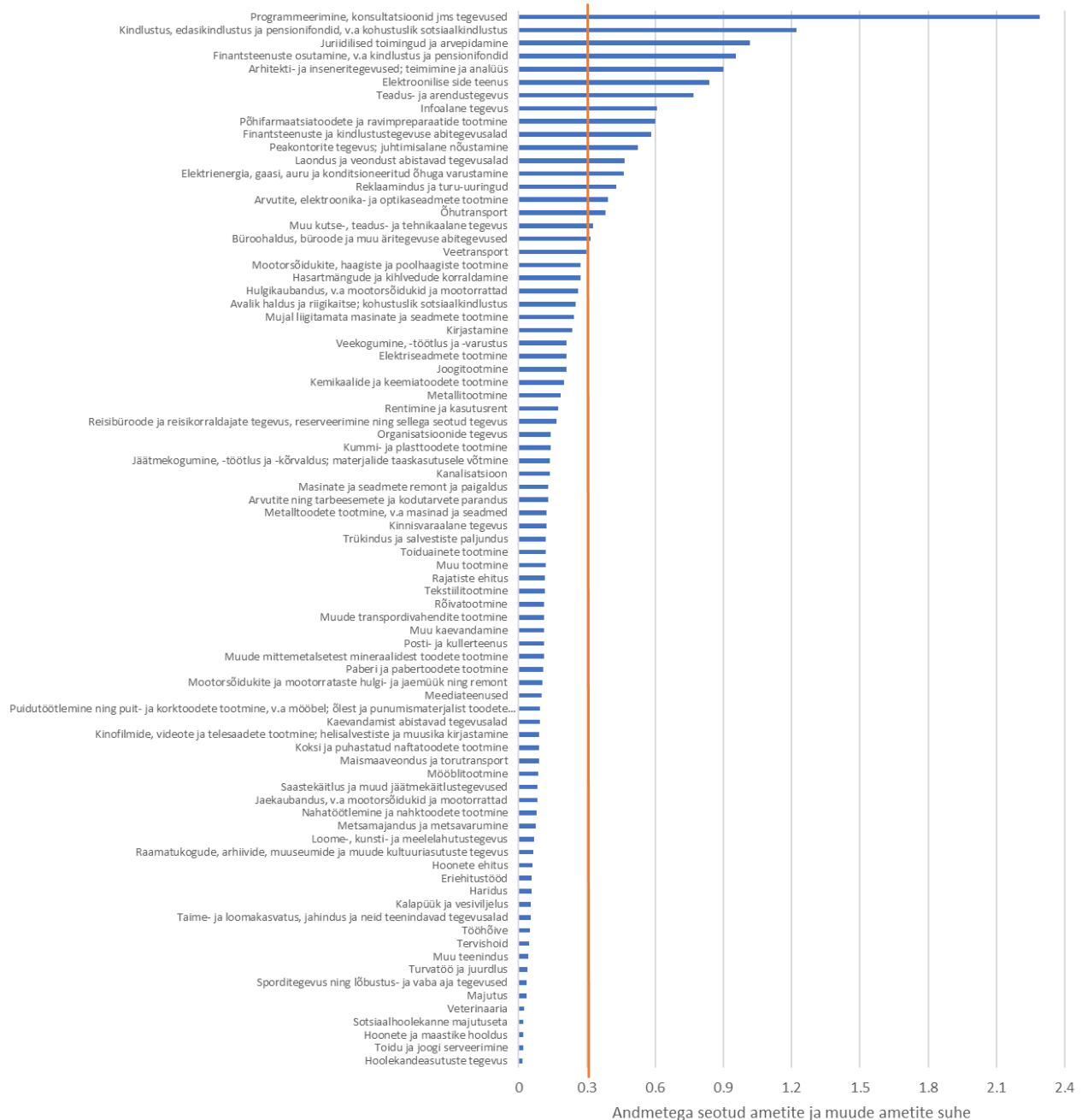
Võrreldes Corrado et al (2022) analüüsis toodud tulemustega teiste riikide kohta, on Eesti tulemus vaikumisi valitud eelduste kohaselt pigem madal, vt nende tabel 3. Peamine erinevus tuleb sellest, et ülal tehtud eeldused aja kohta, mis inimesed tegelevad andmetega, on väiksemad. Samuti kasutas Corrado et al ametialade ISCO kolmanda tasandi koode, mis tähendas, et andmetega töötamist laiendati suuremale hulgale inimestest.

Meie tulemused on ka väiksemad kui väljaandes IDC (2022) European Data Market Update toodud väärtused Eesti kohta 2022. aastal (10.5% SKPst). Et viimase meetodika ei ole aga reprodutseeritav Eesti kohta, siis me ei tea, millest on tingitud nii suur hinnangute erinevus.

4.2. Ettevõtete indikaatorid juhtimislaua

Lisaks kulupõhisele andmete väärtuse leidmisele ning selle komponentidele (andmetega töötavate inimeste täistööajale taandatud arv) sisaldab töölaud veel teisi indikaatoreid andmemajanduse kohta.

Et andmekaitsealsetel põhjustel ei saanud leida, kas igas üksikus ettevõttes on tööl andmetega tegelevad inimesed, siis töölaua prototüübi jaoks valiti andmemajandusega ettevõtted välja EMTAKi 2-kohalise tegevusala järgi. Andmetega tegelevate ettevõtetenäe käsitletakse kõiki ettevõtteid, mis kuuluvad EMTAK 2-kohalise tegevusala kategooriasse, kus andmetega tegelevad töötajad moodustasid vähemalt 30% ülejäänud töötajatest. Andmetega tegelevad töötajad defineeriti kui töötajad, kus ISCO neljanda taseme ametiala kood kuulus punktis 4.1.1. (või tabelis 5) toodud loetellu. Täpsem andmepäring on toodud lisa 3.



Joonis 10. Andmemajandusega seotud tegevusalade valimine

Nii defineeriti 2022. aastal selliselt valitud kriteeriumi järgi andmetega tegelevateks järgmised tegevusalad:

- Arhitekti- ja inseneritegevused; teimimine ja analüüs
- Büroohaldus, büroode ja muu äritegevuse abitegevused
- Arvutite, elektroonika- ja optikaseadmete tootmine
- Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine
- Elektroonilise side teenus
- Finantsteenuste ja kindlustustegevuse abitegevusalad
- Finantsteenuste osutamine, v.a kindlustus ja pensionifondid
- Infoalane tegevus
- Juriidilised toimingud ja arvepidamine
- Kindlustus, edasikindlustus ja pensionifondid, v.a kohustuslik sotsiaalkindlustus
- Laondus ja veondust abistavad tegevusalad
- Muu kutse-, teadus- ja tehnikaalane tegevus
- Õhustransport
- Peakontorite tegevus; juhtimisalane nõustamine
- Põhifarmaatsiatoodete ja ravimpreparaatide tootmine
- Programmeerimine, konsultatsioonid jms tegevused
- Reklaamindus ja turu-uuringud
- Teadus- ja arendustegevus

Neist omakorda kolm tegevusala: elektroonilise side teenus (tegevusala 61), programmeerimine, konsultatsioonid jms tegevused (tegevusala 62) ja infoalane tegevus (tegevusala 63) kuuluvad IKT sektori alla.

(Tulevikus soovitame määrata statistikaameti abil, kas ettevõtte kuulub andmemajanduse alla lähtuvalt sellest, kas ettevõtte on tööl inimesed, kes tegelevad andmeteabe või andmebaaside loomisega ning teha seejärel ära ettevõtete majandusnäitajate agregeerimine.)

Valitud tegevusaladel töötavate ettevõtete kohta valiti välja järgmised näitajad:

- arv, müügitulu, lisandväärtus, investeeringud immateriaalsesse põhivarasse 2022. aastal (kui viimane teadaolev aasta)
- Eraldi samad näitajad tehnoloogia ja tarkvara (IKT) arendusettevõtete kohta (tegevusalalt 61, 62 ja 63).

Kõik näitajad on pärit statistikaameti andmebaasi tabelist EM001.

Lisaks absoluutnäitajatele on esitatud ka andmemajandusega tegelevate ettevõtete majandusnäitajate suhe kogu Eesti ettevõtete koondnäitajatesse ja IKT ettevõtete näitajatesse

4.3. Muud indikaatorid

Muude näitajadena on töölaual kuvatud taustana indikaatorid teisest andmeallikast:

- Teadus- ja arendustegevuse kulud (Statistikaameti tabel TD050)
- Ettevõtted, kus oma töötajad tegelevad andmeanalüütikaga, % (Statistikaameti tabelist IT151)
- IKT-spetsialistide olemasolu ettevõtetes (Statistikaameti tabel IT138)
- IKT osakaal kõrgharidusse vastuvõetud üliõpilaste seas (Statistikaameti tabeli HT303 põhjal)

4.4. Töölaua tehniline rakendus ja sisendandmed

Uuringu raames töötati välja juhtimislaud, mille eeskujuks olid Eesti Statistikaameti²³ ja Eesti Töötukassa²⁴ tööpoliitika näitajate töölaud, Rahandusministeeriumi makromajanduse olukorra ilmakaart²⁵ ning perepoliitika ja vanemahüvitise näitajate töölauda prototüüp²⁶. Juhtimislaud loodi veebirakendusena kasutades Pythoni tarkvara.

Andmelaua rakendusega saab tutvuda aadressil <https://andmemajandus.stacc.cloud/>.

Andmemajanduse prototüüp sisaldab järgmisi lehti (Joonis 10):

- Koguväärtus
- Tööhõive
- Ettevõtted
- Välislingid
- Taustainformatsioon

4.4.1. Andmemajanduse koguväärtus

Andmemajanduse töölauda lehel Koguväärtus on esitatud tulemused andmete väärtuse kulupõhise leidmise kohta, reprodutseerides dünaamiliselt alapunktis 4.1.6 toodu. Tulemused on esitatud nii kokku absoluutsummana, protsendina SKPst, kasvuna viimase ja esimese perioodi vahel ja ka keskmise kasvuna ja võrdlusena IKT sektori müügitulusse.

Andmemajandus

See andmelaud kuvab statistikat ja hinnanguid Eesti andmemajanduse kohta.

- Vahelehel "Koguväärtus" näeb hinnanguid andmemajanduse suuruse kohta.
- Vahelehel "Tööhõive" näeb hinnanguid andmemajandusega tegeleva tööhõive kohta.
- Vahelehel "Ettevõtted" näeb statistikat andmemajandusega tegelevate ettevõtete kohta.
- Vahelehel "Välislingid" on loetletud mõned huvipakkuvad näitajad, mida saab näha teistel lehekülgedel.
- Vahelehel "Taustainformatsioon" näeb informatsiooni EMTAK kategooriate ning kasutatud meetoodika ja andmete kohta ning saab muuta arvutustes kasutatud eeldusi ametite andmetega tegelemise ajakulu kohta.

Vali vaheleht, mida vaadata:

[Koguväärtus](#) [Tööhõive](#) [Ettevõtted](#) [Välislingid](#) [Taustainformatsioon](#)

Joonis 11. Andmemajanduse töölauda kuvatõmmis, päis

²³ <https://tooturg.stat.ee/>

²⁴ <https://www.tootukassa.ee/et/statistics-and-research/peamised-statistilised-naitajad>

²⁵ <https://www.fin.ee/riigi-rahandus-ja-maksud/fiskaalpoliitika-ja-majandus/majanduse-ilmakaart>

²⁶ <https://vorkanalytics.shinyapps.io/VHToolaud/>

Koguväärtuse leidmiseks kasutatavad sisendandmed võetakse kolmest kohast:

- Exceli failist nimega "tulemused_SA.xlsx", lehelt "osa2_4k", mis sisaldab ISCO neljanda taseme tööjõukulu EMTAK1 tegevusalade lõikes. Fail on saadud päringu tulemusena statistikaametilt (vt Lisa 3 päringu struktuuri kohta). Oluline on säilitada sama andmestiku struktuur tulemuste uuendamisel.
- Exceli failist nimega "eeldusedajakohta.xlsx", mis sisaldab eeldusi ISCO neljanda taseme ametite lõikes tööaja osakaalu kohta, mille tulemusena tekib kas andmekogu või andmeteave. Saadud tabeli saab ka ise üles laadida uute eeldustega. Oluline on, et andmestik sisaldaks tunnuseid *iscocode* (ISCO neljanda taseme kood) ja *kaal*.
- Statistikaameti andmebaasist API lehelt <https://andmed.stat.ee/api/v1/et/stat/EM001> vajalikud sisendandmed skaleerimisfaktori jaoks. Vaikimisi on töölaua koodis kasutatud 2020-2022 keskmisi näitajaid faktori leidmiseks.

Töölaual kuvatakse ka suhtarvud SKP ja IKT lisandväärtuse suhtes. Need näitajad on praegu töölaua koodis käsitsi sisestatud failis metrics.py (tunnused `_LAST_YEAR_GDP_VALUE_MILLIONS = 37_682.4` ja `_2022_ICT_VALUE_MILLIONS = 7_417.0234`)

Andmemajanduse koguväärtus

Kuna andmemajandus ei ole otse mõõdetav, kasutatakse siin andmelaua hinnangut. Tööhõive andmete ja ametite andmetega tegelemise hinnangute abil on arvatud andmetega tegeleva töö tööjõukulu hinnang. See arv on korrutatud konstandiga, mis hindab tööjõukulude ja kogutulu suhet, et saada hinnang kogutulule. Rohkemate detailide jaoks vt vahelehte "Taustainformatsioon". Kõik rahalised väärtused on jooksevhindades, inflatsiooni ei ole arvestatud

Andmemajanduse koguväärtus, 2023 ⓘ

1310 mln €

Andmemajanduse osa SKTst, 2023 ⓘ

3.48 %

Andmemajanduse osa IKTst, 2022 ⓘ

15.4 %

Andmemajanduse aastane kasv, 2023

14.7 %

Andmemajanduse kogukasv, 2021-2023

35.3 %

Andmemajanduse keskmine aastane kasv, 2021-2023 ⓘ

16.3 %

Suurima kasvuga sektorid:

1. Finants- ja kindlustustegevus (K): 27.7%
2. Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine (D): 23.3%
3. Haridus (P): 21.2% .

Vähima kasvuga sektorid:

1. Põllumajandus, metsamajandus ja kalapüük (A): 4.8%
2. Muud teenindavad tegevused (S): 5.5%
3. Hulgi- ja jaekaubandus; mootorsõidukite ja mootorrataste remont (G): 7.8% .

Aastate lõikes

Nendelt joonistelt näeb andmemajanduse koguväärtuse muutust aja jooksul ja saab võrrelda muutust terves andmemajanduses ja huvipakkuvas sektoris.

Terve andmemajandus:



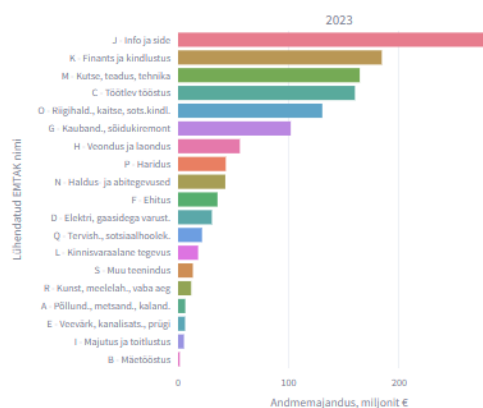
Selle EMTAK sektori andmemajandus:



Viimane aasta sektorite lõikes

Sellel joonisel saab võrrelda andmemajanduse koguväärtust eri sektorites.

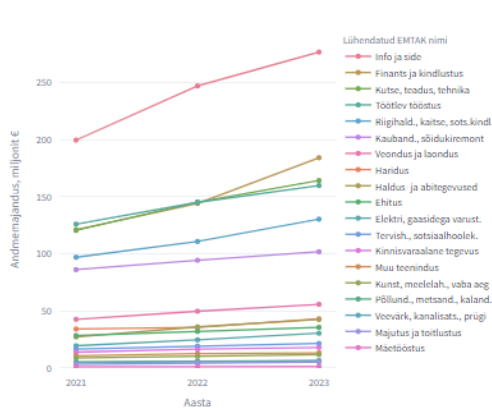
Joonise ülesehituse seletus: (kliki, et avada ja sulgeda)



Aastate ja sektorite lõikes

Sellel joonisel saab võrrelda andmemajanduse koguväärtuse muutust aja jooksul eri sektorites.

Joonise ülesehituse seletus: (kliki, et avada ja sulgeda)



Joonis 12. Andmemajanduse töölaua kuvatõmmis, koguväärtuse vaheleht

4.4.2. Tööhõive

Lehel "Tööhõive" on toodud andmetega töötavate inimeste täistööajale taandatud inimeste arv, mis on skaleeritud täiendavalt osakaaluga, mida eeldame, et inimesed tegelevad andmetega. Valitud eelduste korral on 2023. aastal tulemuseks 13970 inimest, mis on töölaual ümardatud 14.0 tuhande inimese peale. Kaalumata väärtus (vt ka punkti 4.1.2, tabel 6) oleks 91429. Et andmetega tegelemise osakaalud ei muutu üle aastate ja tegevusalade, siis sisuliselt näitab suhteline kasv üle aastate ametite arvu kasvu, kus andmetega töötamise osakaal oli positiivne.

Andmetega tegelev tööhõive

Andmetega tegeleva tööhõivena on siin käsitatud nende ametite tööhõivet, mis tegelevad hinnangu põhjal vähemalt 10% ajast andmetega, kaalutud andmetele kuluva aja osakaalu hinnanguga. Tööhõivet mõõdetakse täistöökohta ekvivalentides. Rohkemate detailide jaoks vt vahelehte "Taustainformatsioon".

Andmetega tegelev tööhõive, 2023	Andmetega tegeleva tööhõive osa tervest tööhõivest, 2022	Andmetega tegeleva tööhõive osa IKT tööhõivest, 2022
14.0 tuhat	2.92 %	39.8 %
Andmetega tegeleva tööhõive aastane kasv, 2023	Andmetega tegeleva tööhõive kogukasv, 2021-2023	Andmetega teg. tööh. keskmine aastane kasv, 2021-2023
3.0 %	9.03 %	4.43 %

Suurima kasvuga sektorid:

1. Finants- ja kindlustustegevus (K): 17.6%
2. Elektrenergia, gaasi, auru ja konditsioneeritud õhuga varustamine (D): 12.2%
3. Haridus (P): 8.3% .

Vähima kasvuga sektorid:

1. Muud teenindavad tegevused (S): -4.8%
2. Kinnisvaraala tegevus (L): -3.6%
3. Ehitus (F): -2.4% .

Aastate lõikes

Nendelt joonistelt näeb andmetega tegeleva tööhõive muutust aja jooksul ja saab võrrelda muutust kogu majanduses ja huvipakkuvus sektoris.

Terve andmetega tegelev tööhõive:

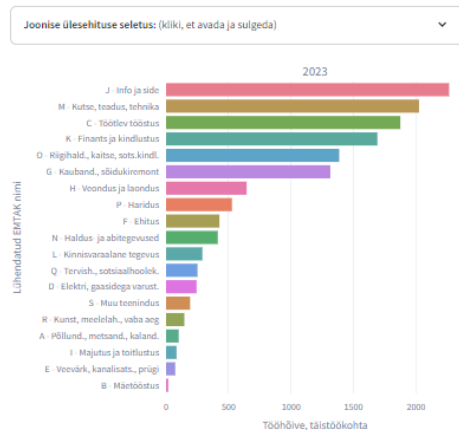


Selle EMTAK sektori andmetega tegelev tööhõive:



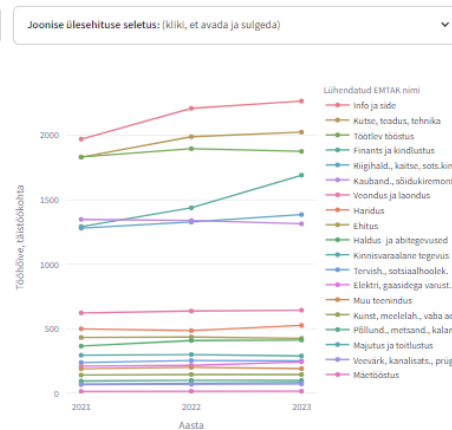
Viimane aasta sektorite lõikes

Sellel joonisel saab võrrelda andmetega tegeleva tööhõivet eri sektorites.



Aastate ja sektorite lõikes

Sellel joonisel saab võrrelda andmetega tegeleva tööhõive muutust aja jooksul eri sektorites.



Joonis 13. Andmemajanduse töölaua kuvatõmmis, tööhõive vaheleht

Hõive väärtuse leidmiseks kasutatavad sisendandmed võetakse järgmistest kohtadest:

- Exceli failist nimega "tulemused_SA.xlsx", lehelt "osa1_4k", mis sisaldab summaarset töökoormust ISCO neljanda taseme ja EMTAK1 tegevusalade lõikes. Fail on saadud päringu tulemusena statistikaametilt (vt Lisa 3 päringu struktuuri kohta). Oluline on säilitada sama andmestiku struktuur tulemuste uuendamisel.
- Exceli failist nimega "eeldusedajakohta.xlsx", mis sisaldab eeldusi ISCO neljanda taseme ametite lõikes tööaja osakaalu kohta. (Vt punkti 4.3.1)

Lehe päises on toodud võrdlus kõigi tegevusalade täistööajale taandatud töötajate arvuga ja kitsamalt info- ja sidetehnoloogia tegevusalal. Need näitajad on praegu töölaua koodis käsitsi sisestatud failis metrics.py (tunnused _2022_TOTAL_LABOR_THOUSANDS = 463.720 ja _LAST_YEAR_ICT_LABOR_THOUSANDS = 34.100). Nende arvude allikaks on Statistikaameti tabel EM001.

4.4.3. Ettevõtted

Lehel Ettevõtted on antud hinnang andmetega tegelevate ettevõtete arvule. Andmetega tegelevate ettevõtetenä on siin käsitletud kõiki ettevõtteid, mis kuuluvad EMTAK 2-kohalise tegevusala kategooriasse, kus andmetega tegelevate töötajate suhe kõikidesse teistesse töötajatesse on 30%. Valitud piir 30% on samas subjektiivne. IKT ettevõtetenä on käsitletud kõiki ettevõtteid, kes on tegevusvaldkonnas 61, 62 või 63. Prototüübis toodud näitajad ei ole automaatselt uuendatavad, vaid on käsitsi eelnevalt välja arvatud ja seejärel esitatud failis app.py.

Andmetega tegelevad ettevõtted

Andmetega tegelevate ettevõtetenä on siin käsitletud kõiki ettevõtteid, mis kuuluvad EMTAK 2-kohalise tegevusala kategooriasse, kus vähemalt 30% tööhõivest tegeleb andmetega. Rohkemate detailide jaoks vt vahelehti "Tööhõive" ja "Taustainformatsioon".

Andmetega tegelevad ettevõtted, 2022 [Ⓢ]	Andmetega teg. ettevõtete osa kõigist ettevõtetest, 2022 [Ⓢ]	Andmetega teg. ettevõtete osa IKT ettevõtetest, 2022 [Ⓢ]
39.9 tuhat	27.6 %	385 %
Andmetega tegelevate ettevõtete müügitulu, 2022 [Ⓢ]	Andmetega teg. ettev. müügitulu osa kõigist ettev., 2022 [Ⓢ]	Andmetega teg. ettev. müügitulu osa IKT ettevõtetest, 2022 [Ⓢ]
19.5 mlrd €	19.8 %	379 %
Andmetarnijate müügitulu suhe sektori J (Info ja side) müügituluga võrreldes, 2022		
25.4 %		

Loodud lisandväärtus (LV)

Lisandväärtus on käibe ja toodangu sisendkulude vahe.

Andmetega tegelevate ettevõtete LV, 2022 [Ⓢ]	Andmetega teg. ettevõtete LV osa kõigist ettevõtetest, 2022 [Ⓢ]	Andmetega teg. ettevõtete LV osa IKT ettevõtetest, 2022 [Ⓢ]
5.67 mlrd €	25.7 %	326 %

Suurima lisandväärtusega hõivatud kohta andmetega tegelevad EMTAK 2-kohalsed tegevusalad:

1. Elektrienergia, gaasi, auru ja konditsioneeritud õhuga varustamine (35): 240 tuhat €
2. Elektroonilise side teenus (61): 94.8 tuhat €
3. Peakontorite tegevus; juhtimisalane nõustamine (70): 85.4 tuhat €

Vähima lisandväärtusega hõivatud kohta andmetega tegelevad EMTAK 2-kohalsed tegevusalad:

1. Muu kutse-, teadus- ja tehnikaalane tegevus (74): 34.1 tuhat €
2. Juridilised toimingud ja arvepidamine (69): 34.5 tuhat €
3. Põhifarmaatsiatoodete ja ravimpreparaatide tootmine (21): 36.1 tuhat €

Investeeringud immateriaalsetesse põhivaradesse (IIP)

Immateriaalne põhivara hõlmab füüsilise vormita vara nagu näiteks tarkvara, intellektuaalomandeid või lubasid.

Andmetega tegelevate ettevõtete IIP, 2022 [Ⓢ]	Andmetega teg. ettevõtete IIP osa kõigist ettevõtetest, 2022 [Ⓢ]	Andmetega teg. ettevõtete IIP osa IKT ettevõtetest, 2022 [Ⓢ]
213 mln €	71.0 %	120 %

Joonis 14. Andmemajanduse töölaua kuvatõmmis, ettevõtete vaheleht

Andmetega tegelevaid ettevõtteid oli 39857 (töölaua ümardatud kui 39,9 tuhat). Nende müügitulu on 19.5 miljardit eurot, lisandväärtus 5.67 miljardit eurot ja investeeringud immateriaalsesse põhivarasse 213 miljonit eurot. Siinkohal tuleb muidugi rõhutada, et nende näitajate puhul ei ole tegemist üksnes andmetest tulenevate rahavoogudega.

Töölaua on toodud nende näitajate suhe kõikide ettevõtete majandusnäitajatesse: vastavalt 27.6% kõikidest ettevõtetest, 19.8% müügitulust, 25.7% lisandväärtusest ja 71% investeeringutest immateriaalsesse varasse.

IKT sektori vastavad osakaalud oleksid 7% ettevõtetest, 5% müügitulust, 8% lisandväärtusest ja lausa 59% investeeringutest immateriaalsesse varasse.

Töölaua on esitatud ka andmetega tegelevatest tegevusaladest kolm suurima ja kolm vähima lisandväärtusega töötaja kohta tegevusala.

Töölaua on esitatud ka andmetarnijate müügitulu osakaal kogu IKT sektori müügitulus, mis oli 2022. aastal 25.4%. Andmetarnijatena on defineeritud tegevusaladel "Elektroonilise side teenus" ja "Infoalane tegevus" tegutsevad ettevõtted.

Viimaks on töölaua prototüübis toodud ka kõikide andmetega tegelevate ettevõtete suhe IKT sektori ettevõtete näitajatesse.

4.4.4. Välislingid

Leht Välislingid on reserveeritud välistele andmetele, mis võetakse kas automaatselt või käsitsi teistest andmeallikatest. Samuti saab siia lehele koondada rahvusvahelistest uuringutest leitud tulemused Eesti andmemajanduse kohta. Tegemist on näitajate komplektiga, mis pakub andmemajandusele erinevaid erinevast vaatenurgast.

Praegusel hetkel on toodud lehel tulemused järgmistest tabelitest:

- IT151: Suurandmeid analüüsinud ettevõtte tegevusala ja tööga hõivatud isikute arvu järgi
- IT138: IKT-spetsialistide olemasolu ettevõtte tegevusala ja tööga hõivatud isikute arvu järgi
- HT303: Vastuvõetud üliõpilased õppekavagrupi, haridusastme ja emakeele järgi
- TD050: Kulutused teadus- ja arendustegevusele institutsionaalse sektori ja kulu liigi järgi

Hetkel on prototüübis esitatud arvud sisestatud käsitsi faili "app.py". Samas failis asuvad ka lingid välistele andmetele. Neist kõige olulisem on Statistikaameti Info- ja kommunikatsioonitehnoloogia juhtimislaud.

Välislingid

Siin on loetletud mõned huvipakkuvad näitajad, mida saab täpsemalt näha teistel lehekülgedel.

Ettevõtted, kus oma töötajad tegelevad andmeanalüütikaga

Ettevõtted, kus oma töötajad tegelevad andmeanalüütikaga, 2023

18,8 %

Vaata täpsemalt [sellest Statistikaameti tabelist](#).

IKT spetsialistid

IKT-spetsialistidega ettevõtted, 2022

17.0 %

↑ 2.12 %

Vaata täpsemalt [sellelt Statistikaameti leheküljelt](#) või [sellest Statistikaameti tabelist](#).

IKT osakaal kõrghariduse vastuvõetute seas

IKT osakaal kõrghariduse vastuvõetute seas, 2023

10.3 %

↓ - 8.81 %

Vaata täpsemalt [sellest Statistikaameti tabelist](#).

Teadus- ja arendustegevus

Kulutused teadus- ja arendustegevustele, 2022

642 mln €

↑ 16.5 %

Vaata täpsemalt [sellelt Statistikaameti leheküljelt](#) või [sellest Statistikaameti tabelist](#).

Vaata veel

[Statistikaameti IKT valdkonna lehekülg](#)

[Statistikaameti IKT juhtimislaud](#)

Joonis 15. Andmemajanduse töölaua kuvatõmmis, välislinkide vaheleht

4.4.5. Taustainformatsioon

Leht Taustainformatsioon on reserveeritud meetoodilisteks juhisteks. Leht sisaldab hetkel lühikokkuvõtet meetoodikast, mis käsitleb kulupõhist andmete väärtuse leidmist. Samale lehele saab panna taustainformatsiooni ka teiste alamlehtede kohta.

Kui soovid kasutada teistsuguseid ajakasutuse hinnanguid, siis lae üles ajakulu eelduste Excel fail (pole kohustuslik):

Drag and drop file here
Limit 200MB per file • XLSX

Browse files

testeeldusedajakohta.xlsx 13.5KB

Laadi alla ajakasutuse hinnangute fail, et seda muuta

Fail edukalt üles laetud!

Laadi alla ajakasutuse hinnangute fail, et seda muuta

Joonis 16. Andmemajanduse töölaua kuvatõmmis, ajaelduste üleslaadimine

Lehele on võimalik üles laadida teistsugused eeldused andmetega tegelemise aja osakaalu kohta erinevates ametites. Antud eeldus on kriitiline osa andmetele väärtuse leidmise meetoodikas, aga samas ei toetu see otseselt ühelegi Eesti uuringule. Seetõttu on otstarbekas erinevaid stsenaariume läbi mängida muutes neid eeldusi.

4.5. Töölaua valideerimine

Poliitikakujundamiseks sobiva ja praktilise andmemajandust iseloomustava töölaua lahenduse valideerimiseks viidi läbi intervjuud erinevate valdkondade esindajatega. Järgnevalt on esitatud ülevaade **intervjuude käigus antud hinnangutest** vajalike andmemajanduse indikaatorite ja juhtimislaua funktsionaalsuse kohta ning laiemalt andmemajanduse määratlemisest.

Nii avaliku kui erasektori esindajad määratlesid **andmemajandust** kui andmete väärindamise ja kasutamise protsessi majandusliku ja ühiskondliku väärtuse loomiseks. Avaliku sektori esindajate hinnangul on andmemajandus mitmetahuline, keskendudes väärtuse loomisele, ühiskondliku kasu saavutamisele ning efektiivsete otsuste tegemisele. Peamised aspektid hõlmavad väärtuse loomist, andmevahetust ja lisandväärtuse mõõtmist, andmete väärindamist (andmete müük teenuste loomiseks ja tulu teenimiseks), ühiskondlikku kasu ning automatiseerimist ja otsuste tegemist.

Erasektori esindajad näevad andmemajandust kui protsessi, mis hõlmab andmete kogumist, töötlemist ja kasutamist majandusliku ja ühiskondliku kasu saavutamiseks, otsuste toetamiseks ning protsesside ja süsteemide tõhususe parandamiseks. Nende hinnangul on andmemajandus laiem kui lihtsalt andmete ostmine ja müümine, sisaldades ka andmepõhiseid tööviise ja meetodikaid ning inimeste ja organisatsioonide rolli. Peamisteks aspektideks on andmete kogumine ja analüüs, andmete kasutamine majandusliku kasu saavutamiseks, andmesüsteemid ja -infrastruktuur, andmete vahetamine ja koostöö ning otsustusprotsessid.

Andmemajanduse hindamise ja kirjeldamise **peamised mõõdikud** sõltuvad eesmärkidest, kuid üldiselt hõlmavad need väärtuse loomist, andmete kvaliteeti, tööjõu ja kulude analüüsi, tehnilisi näitajaid ning turuinfo ja tarbijaelistuste jälgimist. Avaliku sektori esindajad rõhutasid mõõdikute vajadust ja eesmärgi (mõõdikud sõltuvad sellest, mida soovitakse saavutada), pikemat ajavaadet, andmete kvaliteeti ja kättesaadavust ning tööjõu ja kulude-tulude analüüsi olulisust. Tähtsaks peetakse ka tehnilisi mõõdikuid, protsesside efektiivsus ning trende ja tulevikuväärtusi.

Erasektori esindajad peavad oluliseks turu mahtu ja tellimuste hulka (IT-ettevõtete andmesüsteemide tööde tellimuste hulk), otsustusprotsesside kvaliteeti, andmete kasutamist ja vahetamist majanduslikku kasu saavutamiseks ning turuinfot, sh turu müüke, tarbijaelistusi ja konkurentide majandusaasta andmed. Samuti rõhutati vajalike andmespetsialistide olemasolu ja ettevõtete huvi andmete kasutamise vastu.

Andmemajanduse edukust või tõhusust saab avaliku sektori esindajate hinnangul mõõta läbi eesmärgipõhiste mõõdikute, keskendudes andmete mahule, hoiustamisele ja avaliku ning erasektori koostööle. Koostöö edukus võib olla oluline andmemajanduse edukuse näitaja. Samuti saab andmemajanduse edukust mõõta läbi väärtuse loomise ja tasuvusanalüüsi (kas andmealgatused toovad kasu võrreldes nende kuluga), keskkonnamõju, sh energiatarbimist ja digitaalprügi haldamist (pikemaajalise andmesalvestuse ja ebavajalike andmete hindamine aitab vähendada keskkonnamõju). Oluline on hinnata andmetehnoloogiate praktilist rakendamist ja nende tõhusust avalike teenuste parendamisel. Samuti on tähtis juurdepääs ja andmete kvaliteet, kuna avaliku sektori projektides tuleb tagada tasakaal avatud andmete kasulikkuse ja konfidentsiaalsuse vahel.

Erasektori esindajate hinnangul on andmemajanduse edukuse peamiseks näitajaks andmete kättesaadavus ja sagedus. Näiteks kinnisvara ja IT sektorites on andmete kättesaadavus ja sagedus

(kvartaalsed või aastased andmed) kriitilise tähtsusega. Oluline on hinnata, kuidas ettevõtted kasutavad nii sisemisi kui väliseid andmeid, millist kasu andmevahetustest saadakse ja milline on andmevahetuse maht ja mõju. Rõhutati ka andmete võimet pakkuda väärtuslikku konteksti äritegevuse jaoks (müügikasv, turu mõistmine), toetades strateegilisi otsuseid ja operatiivset tõhusust. Samuti on võtmetähtsusega andmesüsteemide ja -protsesside integratsiooni toimivuse hindamine ning otsuste kvaliteet tõhususe suurendamiseks ja äriksvu edendamiseks.

Kokkuvõttes keskendub avalik sektor sotsiaalsele kasule, keskkonnamõjule ja koostöö efektiivsusele, samas kui erasektor rõhutab ärijõudlust, operatiivset tõhusust ja andmete konteksti väärtust. Mõlemad sektorid peavad oluliseks andmepõhist otsuste tegemist ja andmeressursside tõhusat kasutamist.

Töölaua kuvamise selguse ja visuaali hindamisel rõhutasid avaliku sektori esindajad lihtsust ja arusaadavust. Abiks oleks selgituste lisamine samale lehele vähendamaks liikumist lehtede vahel. Oluline on selgelt välja tuua andmemajanduse ja kogusektori võrdlus.

Erasektori esindajate hinnangul on oluline, mida nende andmetega edasi tehakse. Toodi välja, et infograafikud peaksid olema lihtsad ja informatiivsed, koos selgitusega, kuidas neid lugeda. Värvide kasutamine peaks olema mõõdukas ja funktsionaalne, et vältida ülekoormatust. Välja toodi, et võrdlused sektori sees ja riigi keskmisega on olulisemad kui erinevate sektorite võrdlus.

Töölaua peamiste omaduste ja funktsioonide osas rõhutasid avaliku sektori esindajad põhimõtete selgituste vajalikkust mõistmaks, kuidas mõõdikud on koostatud.

Erasektori esindajate hinnangul peaksid visuaalid sisaldama projektsioone ja trendianalüüse, pakkudes laiemat konteksti. Lisatekstit ja infokastid peaksid aitama kasutajatel mõista jooniste tähendust ja konteksti. Mõlema sektori esindajad pidasid oluliseks võimalust valida ja võrrelda erinevaid sektoreid ja ajaperioode (kuni viis aastat).

Töölaua näitajana, mis võiks toetada organisatsiooni tööd, tõid avaliku sektori esindajad välja andmete kvaliteedi ning tööjõu ja kulude analüüsi. Samuti toetab organisatsiooni tööd avaliku ja erasektori koostöö tõhusus, spetsialistide olemasolu ja nende haridustase, mis mõjutab andmemajanduse edukust.

Erasektori esindajate hinnangul on organisatsiooni tööd toetavateks näitajateks samuti andmete kättesaadavus ja sagedus ning andmete kasutamise tõhusus ja andmevahetuse maht. Samuti sektori sisene võrdlus, näiteks ettevõtete edetabel, mis näitab ettevõtete paiknemist sektoris ning turuinfo.

Kokkuvõttes rõhutavad nii avaliku kui erasektori esindajad andmemajanduse selgust ja praktilist kasutust, kusjuures olulised on võrreldavad andmed ja nende kontekstuaalne väärtus. Andmete kvaliteet, kättesaadavus, koostöö ja spetsialistide olemasolu on võtmetegurid andmemajanduse edukuse ja tõhususe mõõtmisel.

Erinevate **IKT ja haridusevaldkonna näitajate vajalikkuse** kokkuvõtte on toodud alljärgnevalt.

- **IKT-spetsialistidega ettevõtted andmemajanduseettevõtetes.** Avaliku sektori esindajate hinnangul on näitajat oluline jälgida, kuid ametikohad on erinevad ning nende üldine jälgimine ei oma väärtust. Oluline on näiteks täiendkoolitamise vajaduse jälgimine. Erasektori esindajate hinnangul on samuti tegemist olulise näitajaga, kuid eelistatakse täpset jaotust kutse- ja kõrghariduse vahel.

- **Informaatika ja infotehnoloogia õppekava grupis kutsehariduse omandamine (võivad huvi pakkuda).** Avaliku sektori esindajate hinnangul on seda näitajat oluline jälgida, kuid oluline on ka kõrgharidus ja täiendõpe.
- **Riiklike avaandmete hulk (registreeritud arv) ja avaandmete pakujate nimekiri (kasutusjuhud ja kasulikkus oluline).** Nii avaliku kui erasektori esindajate hinnangul on olulised näitajad, kuid sama oluliseks peeti andmete kvaliteeti ja ajakohasust. Avaliku sektori esindajad märkisid, et lisanduda võiks ka avaliku ja erasektori koostöö näitaja.
- **Tehisintellekti tehnoloogiaid kasutanud ettevõtted tegevusala ja tööga hõivatud isikute arv (jaotusega andmemajandusettevõtete hulgas, IKT sektori ettevõtete hulgas).** Avaliku sektori esindajate hinnangul on vajalik mõista, kuidas AI-tehnoloogiaid kasutatakse ja kuidas need aitavad kaasa valdkonna arengule. Erasektori hinnangul on antud näitaja kasulik teada, kuid otseselt mittevajalik.
- **E-kaubandust kasutanud ettevõtted tegevusala ja tööga hõivatud isikute arvu järgi (jaotusega andmemajandusettevõtete hulgas, IKT sektori ettevõtete hulgas).** Erasektori hinnangul on antud näitaja vähem oluline.
- **Internetiühenduse kiirus, pilveteenuse kasutamine, veebilehe olemasolu, sotsiaalmeedia kasutamine.** Avaliku sektori esindajate hinnangul on tegemist vajaliku näitajaga, et andmemajandus toimiks, tagades andmete kättesaadavuse ja hoiustamise. Samas veebilehe olemasolu ja sotsiaalmeedia kasutamist hinnati vähem oluliseks, kuna lisandväärtus võib olla laialivalguv. Erasektori esindajate hinnangul on samuti tegemist olulise näitajaga, kuid täpsustamist vajab, kas pilveteenus hõlmab ainult lihtsaid teenuseid või ka suuri infosüsteeme.

Täiendavate soovitusena toodi avaliku sektori esindajate poolt välja selgituste ja konteksti lisamist, et paremini tagada näitajate mõistetavus. Lisaks peaksid infograafikud olema lihtsad ja informatiivsed. Erasektori esindajate poolt toodi välja andmete kasutatavust, et andmeid oleks võimalik kasutada masinloetavalt. Mõlema sektori esindajate poolt toodi välja, et tähelepanu tuleks pöörata andmete täpsusele ja ajakohasusele.

KOKKUVÕTE JA JÄRELDUSED

Käesoleva uuringu eesmärgiks oli koostada meetodika ülevaade Eesti andmemajanduse mahu ning turuväärtuse hindamiseks teiste riikide ja mudelite näitel. Selleks tehti uuringu esimeses etapis ülevaade rahvusvaheliselt kasutatavatest meetoditest ja andmetest andmemajanduse hindamiseks. Uuringu teises etapis viidi läbi andmete väärtuse kulupõhine hindamine Eesti näitel ning tehti valmis töölaua prototüüp tulemuste kuvamiseks. Töölaud sisaldab täiendavalt näitajaid, mis tuletati käesolevas raportis andmetega tegelevate töötajate ja ettevõtete arvu kohta ning viited valdkondlikele näitajatele.

Uuringu **I etapis** läbiviidud kaardistuse tulemusel saame välja tuua järgmised olulised punktid:

- EDM tööriista mudel on andmemajanduse vaates mitmekülgne ja arvesse on võetud andmeökosüsteemi vaade. Mudel klassifitseerub nii andmeturu, aga ka andmemajanduse suuruse kombineeritud lahenduseks. Valdavas osas on mudeli parameetrid kasutatavad Eestis ja andmed kogutavad riiklikust statistikast ning riiklikest andmebaasidest.
- EDM tööriista meetodilises lahenduses on kasutatud siiski mitmeid erinevaid uuringuid, millest ülevõetud hinnangute alused on ebaselged. Kasutatakse IDC eelnevates uuringutes ja juhendites välja töötatud aluseid või määratud osatähtsuseid, mida otseselt EDM tööriista meetodilistes alustes välja ei tooda (eelkõige IDC Worldwide Black Book, IDC üksikasjalikud turuprognosid IT riistvara, tarkvara ja IT-teenuste kohta aastatel 2017, 2018 ja 2019).
- EDM tööriista mudeli kasutamise tugevuseks on mitmekülgse vaate andmine. Puuduseks on suur hinnangute hulk, mille kohta puuduvad avalikustatud alused. Mudeli ülevõtmine Eesti andmemajanduse väärtuse kaardistamiseks on võimalik, kuid eeldatavalt tuleb kasutada hinnangute andmisel ja osatähtsuste määramisel täiendavaid siseriiklike arvamusi.
- Andmete väärtuse kulupõhise hindamise raamistiku näide, mida rakendab Kanada Statistikaamet ja Corrado *et al* (2022), võimaldab hinnata andmete väärtust nende soetusväärtuse kaudu.
- Kulupõhise mudeli eeliseks on kvantitatiivsete andmete kasutamine, aga puuduseks on eeldused või hinnanguid, kui suure osa tööajast kulub töötamisele andmetega selliselt, et tekivad uued andmebaasid või uued andmepõhised otsustusmudelid. Mudeli eeliseks võib nimetada lõpptulemust, mida on võimalik edasi anda arvuliselt kokku agregeeritud väärtusena ehk andmemajanduse kulud kokku on võrdne andmemajanduse väärtusega.
- Järeldame, et Eestis on kulupõhine mudel rakendatav, kuna riiklikud registrid võimaldavad kvantitatiivseid andmeid töötamise kohta. Siiski tuleb arvestada, et üksnes kulupõhine mudel on oluliselt kitsam kui EDM tööriistas rakendatud indikaatorid ja ülevaade.
- Lähiriikide andmemajanduse väärtuse kaardistamise võrdluses ilmnes, et kõik riigid on suhteliselt edumeelselt andmemajanduse hindamisega tegelenud. Lisaks on tegeletud avaandmete väärtuse kaardistamisega, mis on oluline ka Eestis. EDM tööriist ja kulupõhine mudel otseselt avaandmete sotsiaalmajandusliku väärtuse hinnanguid ei kajasta. Kulupõhise mudeliga on võimalik siiski avaandmetele väärtus leida.
- Lisaks juhime tähelepanu, et andmemajanduse mõõtmiseks puudub riikide vahel või ka rahvusvahelistel institutsioonidel (Eurostat, OECD, Euroopa Komisjon) ühtne meetodika.

Uuringu II etapis viidi läbi andmemajanduse kulupõhine hindamine Eesti jaoks, teiste indikaatorite väljavalimine andmemajanduse mõõtmiseks, kvalitatiivne uuring soovide kohta töölauale ning töölaua prototüübi loomine.

Tuginedes varasemates uuringutes toodud eeldustele andmetega töötavate inimeste ametialade kohta, näitasid 2023. aasta kohta tehtud arvutused, et täisajale viidud töötajaid, kes loovad andmeteavet või andmebaase on Eestis ca 90 tuhat. Rahvusvahelise meetodika põhjal tegelevad nad vaid andmetega keskmiselt kuuendiku oma tööajast, mis teeb ligi 14 tuhat täisajaga vaid andmeteavet ja andmebaase loovat töötajat. Nende töötasu põhjal tehtud arvutused koos muude kulude lisamisega andsid 2023. aasta kohta andmete turuväärtuseks ca 3.5% SKPst. Arvestades töös kasutatud konservatiivseid eeldusi ja TÖR ning TSD registriandmete andmekaitsest tulenevaid piiranguid on see pigem andmete turuväärtuse hinnangu alumine piir. Suurim osa tuleb andmete väärtusest on info ja side valdkonnas (21%). Järgnevad finants- ja kindlustustegevus (14%), kutse-, teadus- ja tehnikaalane tegevus (13%), töötlev tööstus (12%) ja avalik haldus, riigikaitse ja sotsiaalkindlustus (10%).

Teisest küljest arvatati kõikide ettevõtete, kes potentsiaalselt võiksid tegeleda andmemajandusega, tulenevalt sellest, et nende tegevusalas töötab palju andmete tegelevaid inimesi, müügitulu ja lisandväärtus, mis moodustas vastavalt 20% ja 26% kõikide ettevõtete müügitulust ja lisandväärtustest. Selle näol tegemist pigem ülehinnanguga andmemajanduse suurusele, sest kogu tegevusala majandusnäitajad omistati andmemajandusele.

Töös ei suudetud lõpuni läbi töötada IDC meetodikat European Data Market näitajate kohta ja reprodutseerida selle üllatuslikku suurt hinnangut Eesti kohta (10% SKPst), sest selle meetodikamaterjalid ei olnud teadlaskonnale avalikult kättesaadavad. Meie soovitus on kasutada võimalusel tellija kontakte Euroopa Komisjoniga ja paluda tole uuringu detailset meetodikamaterjale.

Uuringu **II etapi** kohta saame välja tuua järgmised olulised punktid:

- Andmetele kulupõhise väärtuse leidmine on arvutuslikult võimalik ja arvutuskäik on läbipaistev ning reprodutseeritav. Samas Eesti kohta puuduvad seni uuringud, mis võimaldaks hinnata, kui suure osa erinevatel ametitel inimesed tegelevad andmeladude ja andmeteabe loomisega. Et tegemist on kõige enam määramatust sisaldava informatsiooniga, mis samas mõjutab kõige enam andmete kulupõhist väärtust, siis tuleb suhtuda tulemustesse kriitiliselt. Töölaua prototüübil jäeti need andmed muudetavaks sisendfaili abil, mis võimaldab teha tundlikkusanalüüsi eelduste osas.
- Kui soovida jätkata ametite põhise lähenemisega andmete väärtuste leidmisel, siis soovitame läbi viia täiendava uuringu Eestis hindamiseks, kui suur osakaal erinevatest ametitest tegeleb andmete väärindamisega või kui suure osa tööajast see võtab.
- Andmetele kulupõhise väärtuse leidmisel ja ka andmetega tegelevate inimeste töökoormuse hindamiseks TÖRi ja TSD põhjal jäid välja võlaõiguslikud lepingud, sest neis enamasti puudub ametiala ja koormus. 2023. aastal moodustas VÕS lepingute summaarne brutotasu tegevusvaldkonnas J63 (Infoalane tegevus) ligi 6%. Kui edaspidi kasutada registriandmeid andmemajanduse hindamiseks, siis tuleb koos statistikaametiga töötada välja meetod, kuidas arvesse võtta võlaõiguslikke lepinguid. Arvestades paindlike töösuhete kasvu, sealjuures informaatika ja andmeanalüütika valdkonnas, on oht tulevikus aina enam alahinnata andmemajandusega seotud inimeste arvu ja nende tööpanust.

- Sarnast kulupõhist lähenemist investeringute rahalise väärtuse hindamisel kasutab meile teadaolevalt ka Eesti Statistikaamet. Nimelt arvutatakse ettevõtete poolt oma tarbeks toodetud tarkvara väärtuse hindamiseks ettevõtete IT-töötajate palgaandmeid, kuid see meetoodika ei olnud käesoleva raporti kirjutamise ajaks veel avalik meile. Et tegemist on väga sarnaste lähenemistega, siis soovime tulevikus kooskõlastada need kaks lähenemist.
- Andmetele kulupõhise väärtuse leidmine eeldab detailseid andmeid ametiala ja tegevusala lõikes, mis saadi TÕRi ja TSD andmete põhjal, kus aga tuli vastu andmekaitse piirang. Veel enam mõjutab andmekaitse seda, kui soovida määrata andmetega töötavate ettevõtete arvu, nende ettevõtete käivet või lisandväärtust tuginedes TÕRi või TSD andmetele. Käesolevas projektis tehti ühekordsed päringud Statistikaametile, mille vastuses jäi osad andmeväljad tühjaks konfidentsiaalsuse tõttu. Selle vältimiseks soovime tulevikus viia andmemajanduse töölaud statistikaameti alla, mis võimaldaks ilma kadudeta leida täpsemaid koondhinnanguid andmetega tegelevate töötajate arvu ja töötasu kohta tegevusalade lõikes, andmetega tegelevate ettevõtete arvu ja nende majandusnäitajate kohta tegevusalade lõikes ning viimaks ka andmemajanduse kui terviku väärtuse kohta.
- Projekti käigus töölaua prototüübi loomine näitas samuti, et otstarbekas on koostöö statistikaametiga, sest lisaks käesoleva projekti raames loodud indikaatoritele, mis tuginesid statistikaameti andmetele, on ka väga suur osa teisestest andmetest pärit statistikaameti andmebaasist. Omakorda suur osa andmetest on pärit statistikaameti infotehnoloogia juhtimislaualt. Seetõttu käesolevas projektis statistikaameti töölaudadel kuvatud andmeid prototüüpi esialgu üle kandma vaid toodi eraldi viide statistikaameti vastavatele tabelitele ja töölaudadele.
- Intervjuudest saadud tagasiside näitas ootuste väga suurt heterogeensust töölaua suhtes. Ühest küljest sooviti koondnäitajat andmemajanduse mõõtmiseks ja teisest küljest oli ootus, et see kataks kõikvõimalikke näitajaid, mis on vähegi seotud andmetega, digimajandusega ja IKT valdkonnaga. Vastavalt projekti esimeses etapi tulemustele ei soovitud käesoleva projekti töölaua prototüüpi informatsiooniga üle küllastada vaid koguti kokku just need indikaatorid, mis töötati välja käesoleva projekti raames. Samas on töölaud piisavalt paindlik, et koondada sinna selleks loodud lehele laiem indikaatorite valik
- Töölaua prototüüp ehitati valmis Pythoni koodis koos selgitustega, mis võimaldab seal kuvatud indikaatoreid ja visuaalne kergesti modifitseerida või uuendada.

Käesolevat uuringut kokku võttes näeme, et andmemajanduse mõõtmiseks nii rahvusvaheliselt kui ka Eesti siseselt on väga palju erinevaid indikaatoreid. Käesolevas projektis keskenduti neile, mis on võimalik arvutada välja statistikaameti valduses olevate töötajate ja ettevõtete registriandmete põhjal ja mida saab seetõttu kergesti uuendada.

Andmetele kulupõhise väärtuse leidmise plussideks on kokkuvõtvalt: arvutuskäigu läbipaistvus, kerge uuendamine, sarnasus muudes valdkondades väärtuse imputeerimisel statistikaametis ja lihtne tundlikkusanalüüsi teostamise võimalus peamiste eelduste kohta. Lähenemise miinusteks on see, et andmetega töötamise osakaalu eeldus tugineb uuringutele, mis on teiste tööturgude kohta (peamiselt USA). Samuti andmekaitse piirangute ja VÕS lepingute väljajätmise tõttu alahindame tööjõukulu ja seega andmete väärtust, eriti väikestel tegevusaladel.

Uuringus kasutatakse andmete väärtuse mõõtmiseks rahvusvahelist rakendust leidnud meetoodikat, mis tugineb andmetega töötavatele töötajate arvule ja nende töötasule. Samas võrreldes

rahvusvaheliselt kasutatud meetodikaga (nt töös viidatud Kanada statistikaameti lähenemine ja OECD uuringud), kasutatakse käesolevas uuringus registriandmete põhjal (TÖR ja TSD) tuletatud töötajate arvu ja tööjõukulu. Varasemates rahvusvahelistes uuringutes on tuginetud seevastu küsitlusandmetele. Registriandmete põhine analüüs, mida saab teha statistikaamet, võimaldab saada hinnanguid kiiremini ja tuginedes üldkogumile ka täpsemalt. Kumbki lähenemine – kas küsitlusandmete põhine või registriandmete põhine – ei vabasta siiski vajadusest teha eeldused, millised ametialad ja millises mahus on seotud andmetega. Käesolevas projektis on kasutatud eeldusi rahvusvahelisest kirjandusest. Need on eeldused, mida saab samas muuta, et testida lõpptulemuste tundlikkust. Samuti, kui tekitavad täiendavad uuringud selle kohta, millised ametialad on Eestis enam seotud andmetele väärtuse loomisega, siis saab vastavaid eeldusi täpsustada.

Uuring soovib järgmise sammuga käesolevas uuringus välja pakutud andmete väärtuse mõõtmise meetodika ühtlustada statistikaametis juba kasutusel oleva lähenemisega infotehnoloogiasse tehtud investeeringute rahalise väärtuse hindamiseks. Samuti soovime kaasata statistikaametit selleks, et kasutada töötajate-ettevõtete omavahel seotud andmestikke, mis võimaldaks leida andmemajandusega tegelevate ettevõtete käekäik üksikult üldisele agregeerides.

Paralleelselt on Eesti Kutsekoda välja töötamas ametialade ja oskuste omavaheliste seoste registrit, kaasates sinna ka MKMi eksperdid. Seetõttu soovime selle registri valmimisel andmetega seotud ametialade loetelu valideerida ka selle registri tulemustega.

KASUTATUD KIRJANDUS

1. Apanasevic, T. (2021). *Socio-economic effects and the value of open data: A case from Sweden*. <https://www.econstor.eu/handle/10419/238004>
2. Bondt, H. de, Kuipers, A., Mushkudiani, N., & Pluijmen, C. (2021). *De waarde van data 2001-2017*. <https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2020/de-waarde-van-data-2001-2017>
3. Bondt, H., & Mushkudiani, N. (2021). *Estimating the Value of Data in the Netherlands*. IARIW-ESCoE Conference "Measuring Intangible Assets and Their Contribution to Growth", London. https://www.researchgate.net/publication/368667379_Estimating_the_Value_of_Data_in_the_Netherlands
4. Brenda Bugge (Statistics Canada), Koami Amegble (Statistics Canada) and Amanda Sinclair (Statistics Canada) "Estimation of Investment in and Stock of Data in the Canadian System of National Accounts", ettekanne konverentsil IARIW-CIGI Conference November 2-3, 2023
5. Coyle, D., Diepeveen, S., Wdowin, J., Tennison, J., & Kay, L. (2020). *The Value of Data—Summary report 2020*. <https://www.bennettinstitute.cam.ac.uk/publications/value-data-summary-report/>
6. Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C. (2022) Measuring data as an asset: Framework, methods and preliminary estimates OECD. Economics Department Working Papers No. 1731 <https://read.oecd.org/10.1787/b840fb01-en?format=pd>
7. Digital Realty. (2018a). *Data Economy Report 2018*. https://go2.digitalrealty.com/rs/087-YZJ-646/images/Report_Digital_Realty_1805_Data_Economy_Report_2018.PDF?mkt_tok=eyJpIjoiWIRJME9XTXpOaIF4TURjNCIsInQiOiI6QmNJdmtTRGUzZ1wvUHBkRFMwdEYxYU02RjRjSnJueUNqZUdUZlYzYkR1QWRyeEtEZ25Gb1pVXF0dGRQMGtIcUZVNxB0aERFd1RZWVJzZ0RcL3lnVWZqbFVlSWxRREZlVIAzNzhDanM5eW5jbGQ5ZE1CT1dKSElxcDhwbjZWSWhQln0%3D
8. Digital Realty. (2018b). *Digital Realty's Data Economy Report*. <https://userfiles.mailswitch.nl/c/6a7bed2fdc49043b50345fd3469b347c/1045-cd1bca252e892ceaf9e278b815f98e8f.pdf>
9. European Commission, Directorate-General for Communications Networks, Content and Technology, Cattaneo, G., Micheletti, G., Glennon, M. et al. (2020), The European data market monitoring tool – Key facts & figures, first policy conclusions, data landscape and quantified stories – D2.9 final study report, Publications Office, <https://data.europa.eu/doi/10.2759/72084>
10. European Commission, IDC: Gabriella Cattaneo, Mike Glennon, Marianne Kolding, Carla La Croce; Irene Magnani; Giorgio Micheletti; Martin Sunblad. Lisbon Council: Cristina Moise, David Osimo. (2021), European Data Market Study 2021–2023 –D1 Inception Report, Publications Office
11. European Commission, Directorate-General for Communications Networks, Content and Technology
12. (2022), Survey of businesses on the data economy 2022 – Final report, Publications Office of the European

13. Union, <https://data.europa.eu/doi/10.2759/392138>
14. European Commission, Directorate-General for Communications Networks, Content and Technology, Cattaneo, G., Micheletti, G., Raczko, N. et al. (2023), European Data Market Study 2021-2023, Publications Office, <https://digital-strategy.ec.europa.eu/en/library/results-new-european-data-market-study-2021-2023>
15. Ekström, D., & Johannesson, J. (2020a). *The value of open data—Economic impact assessment of open public sector information*. <https://www.lantmateriet.se/contentassets/e16a59e08cb744149c878776256560e6/open-data-report-summary-2020-06-04.pdf>
16. Ekström, D., & Johannesson, J. (2020b). *Värdet av öppna data—Samhällsekonomisk nyttoanalys av tillgängliggörande av särskilt värdefulla data*.
17. Farboodi, M., & Veldkamp, L. (2021). A Model of the Data Economy (Working Paper 28427). National Bureau of Economic Research. <https://doi.org/10.3386/w28427>
18. Fintraffic. (2023a). *Impact Potential of the Traffic Data Ecosystem*. https://www.fintraffic.fi/sites/default/files/2023-08/Impact%20Potential%20of%20the%20Traffic%20Data%20Ecosystem%20-%20Report%20_FintrafficIntra.pdf
19. Fintraffic. (2023b). The data economy creates a value of hundreds of millions of euros in the transport data economy in Finland. *Fintraffic in English*. <https://www.fintraffic.fi/en/the-data-economy>
20. Fleckenstein, M., Obaidi, A., & Tryfona, N. (2023). A Review of Data Valuation Approaches and Building and Scoring a Data Valuation Model. *Harvard Data Science Review*, 5(1). <https://doi.org/10.1162/99608f92.c18db966>
21. Goodridge, P., Haskel, J., Edquist, H. (2021). We See Data Everywhere Except in the Productivity Statistics. *Review of Income and Wealth*. <https://doi.org/10.1111/roiw.12542>
22. Hanna, N.K. Assessing the digital economy: aims, frameworks, pilots, results, and lessons. *J Innov Entrep* 9, 16 (2020). <https://doi.org/10.1186/s13731-020-00129-1>
23. Hogan, O., Holdgate, L., & Jayasuriya, R. (2016). *The Value of Big Data and the Internet of Things to the UK Economy*. https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf
24. Hüsing, T., Korte, W., B., Dashja, E. (2015). e-Skills in Europe Trends and Forecasts for the European ICT Professional and Digital Leadership Labour Markets (2015-2020), empirica Working Paper, https://eufordigital.eu/wp-content/uploads/2019/10/e-Skills-in-Europe_Trends-and-Forecasts.pdf
25. ISWGNA (2020), "Recording and Valuation of Data in National Accounts", issue note prepared by the ISWGNA subgroup on digitalization for the 14th Meeting of the Advisory Expert Group on National Accounts, October 2020. https://unstats.un.org/unsd/nationalaccount/aeg/2020/M14_5_3_1_Recording_of_Data_in_SNA.pdf
26. Koski, H. & Elinkeinoelämän tutkimuslaitos. (2015). The Impact of open data –a preliminary study. *Ministry of Finance publications*.

- <https://vm.fi/documents/10623/1107406/The+Impact+of+Open+Data/1c432b3a-a5e8-41ea-a5ea-135280a69ea3?version=1.0>
27. McKinsey (2013) Open data: Unlocking innovation and performance with liquid information, https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx
 28. Mitchell, J., D. Ker and M. Leshner (2021), "Measuring the economic value of data", *OECD Going Digital Toolkit Notes*, No. 20, OECD Publishing, Paris, <https://doi.org/10.1787/f46b3691-en>.
 29. OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264311992-en>.
 30. Rastas, T. (2023, juuni 14). Making it easier to monitor the development of the data economy – we've put together a tool to track indicators. *Sitra*. <https://www.sitra.fi/en/articles/making-it-easier-to-monitor-the-development-of-the-data-economy-weve-put-together-a-tool-to-track-indicators/>
 31. Statistics Canada (2019a). Measuring investment in data, databases and data science: Conceptual framework. <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm>
 32. Statistics Canada (2019b). The value of data in Canada: Experimental estimates. <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.ht>
 33. Statistics Netherlands. (2021, veebruar 5). CBS examines the value of data. *Statistics Netherlands*. <https://www.cbs.nl/en-gb/corporate/2021/02/cbs-examines-the-value-of-data>
 34. Statistikaamet. Tabel EM001: ETTEVÕTETE MAJANDUSNÄITAJAD, viimati vaadatud 31.03.2024
 35. Statistikaamet. Andmepäringu tulemused. 14.03.2024
 36. Stiglich, L. (2021, oktoober 28). Policy to unlock the economic value of data. *Open Data Institute*. <https://theodi.org/news-and-events/blog/policy-to-unlock-the-economic-value-of-data/?saved>

Lisa 1. Dimensionaalse andmete hindamise mudeli raamistik

Mõõde	Kirjeldus
Omandiõigus	Andmekogumi omandiõigus ja tulenevad piirangud (litsentsid)
Maksumus	Hankimise, hoolduse ja asendamise kulud
Kasutamine	Integreerimise võimalused, kasutamise ulatus, kasutussagedus, metaandmed, nõudluse kasv, väärtuse vähenemine ajas
Vanus	Värskendussagedus ja saadaolev ajalugu (aegrida)
Privaatsus	Tundlike andmete olemasolu ja privaatsusstandardid
Andmete kvaliteet	Täielikkus, täpsus, järjepidevus, usaldusväarsus ja ajakohasus
Maht ja mitmekülgsus	Kirjete arv, teabe ulatus

Lisa 2. Andmemajanduse väärtuse mudelite andmete kaardistus

Andmemajanduse väärtuse mudelite andmete kaardistus (Exceli tabel)



Lisa 2_mudelite
kaardist.xlsx

Lisa 3. Andmepäring statistikaametile

Päring

Tartu Ülikooli Rakendusuuringu Keskuse andmepäring Statistikaametile

1. Täiskohale taandatud töölepinguga töötajate arv aastas samaaegselt tegevusalade (EMTAK 1 tase) ja valitud ametialade lõikes (vt jaotust all) 2021, 2022 ja 2023. Kui koormust ei ole lepingus arvuliselt määratud, siis koormus võetakse võrdseks 1-ga.

Näide:

Aasta	EMTAK 1	ISCO	Arv
2021	A	12	1234
2021	A	13	5678
2021	A	1211	120
...			

2. Tegelik brutopalgade väljamaksete summa aastas, tööpanuste summa aastas, sotsiaalmaksu summa aastas ja töötuskindlustusmaksu summa aastas samaaegselt tegevusalade (EMTAK 1 tase) ja valitud ametialade lõikes (vt jaotust all) 2021, 2022 ja 2023.

Näide:

Aasta	EMTAK 1	ISCO	Brutopalgade summa	Tööpanuste summa	Sotsiaalmaksu summa	Töötuskindlustusmaksu summa
2021	A	12	123456	987654
2021	A	13	345678
2021	A	1211	1234

3. Valitud ametialade (vt loetelu all) töölepinguga töösuhete arv kokku aastas tegevusaladel (EMTAK 2 tase), kõikide töölepinguga töösuhete arv tegevusaladel (EMTAK 2 tase) kokku aastatel 2021, 2022, 2023

Näide:

Aasta	EMTAK 2 tase	Valitud töösuhete summa	Kokku töösuhete summa
2021	01	98765	123456
2021	02	43210	54321

4. Unikaalsete isikute arv ja keskmine koormus EMTAK 2 taseme ja ISCO 2 taseme järgi.

Näide:

Aasta	EMTAK 2	ISCO 2	Unikaalsete isikute arv	Keskmine koormus
2021	01	12	1234	1123
2021	01	13	5678	4578
2021	01	14	120	132

5. Andmetega potentsiaalselt tegelevate töötajate töökoormuste summa ja unikaalsete isikute arv detsembri lõpus EMTAK2 lõikes.

Valitud ametialade loetelu:

- a. Esimene tase ISCO kahekohaline tase järgmiste koodidega: 12, 13, 21, 24, 25, 26, 33, 35, 41, 42, 43, 75
- b. Võimalusel, kui konfidentsiaalsus lubab, siis minna detailsemaks ja esitada lisaks ISCO neljakohalised koodid järgmiste ametite kohta:

1211, 1346, 2111, 2113, 2114, 2120, 2131, 2132, 2133, 2141, 2142, 2143, 2144, 2145, 2146, 2149, 2151, 2152, 2153, 2161, 2162, 2163, 2164, 2165, 2166, 2411, 2412, 2413, 2421, 2422, 2423, 2424, 2431, 2432, 2433, 2434, 2511, 2512, 2513, 2514, 2519, 2521, 2522, 2523, 2529, 2631, 2632, 3311, 3312, 3313, 3314, 3315, 3511, 3512, 3513, 3514, 4132, 4227, 4311, 4312, 4321, 4322, 4323, 7543

Tulemus

Päringu tulemuseks on saadud Exceli andmetabel järgmiste lehtede ja veergudega, mis tõsteti kokku faili "tulemused_SA.xlsx".

- Leht: **osa1_2k** – keskmine töökoormus kokku EMTAK1 ja ISCO 2 tasemete lõikes kvartalite kaupa. Tunnused:
 - AASTA – andmete aasta
 - EMTAK_KOOD1 – EMTAK 1 taseme kood
 - AMET_KOOD2 – ISCO 2 taseme kood
 - keskm_koormus_summa_Q1, keskm_koormus_summa_Q2, keskm_koormus_summa_Q3, keskm_koormus_summa_Q4 – keskmised töökoormused kvartalite kaupa
- Leht: **osa1_4k** – keskmine töökoormus kokku EMTAK1 ja ISCO 4 tasemete lõikes kvartalite kaupa. Tunnused samad, mis eelmisel lehel, välja arvatud AMET_KOOD4 – ISCO 4 taseme kood
- Leht: **osa2_2k** - töökoormuste summa EMTAK1 ja ISCO 2 tasemete lõikes aastas. Tunnused:
 - AASTA - andmete aasta
 - EMTAK_KOOD1 – EMTAK 1 taseme kood
 - AMET_KOOD2 – ISCO 2 taseme kood
 - brutopalgasumma – summaarne brutopalk aastas
 - toopanusumma – summaarne tööpanuste summa aastas
 - sm_summa – summaarne sotsiaalmaksu summa aastas
 - tk_summa – summaarne töötuskindlustusmaksu summa aastas
 - tm_summa – summaarne tulumaksu summa aastas
- Leht: **osa2_4k** - töökoormuste summa EMTAK1 ja ISCO 4 tasemete lõikes aastas. Tunnused samad, mis eelmisel lehel, välja arvatud AMET_KOOD4 – ISCO 4 taseme kood
- Leht: **osa3** – töösuhete arv kokku aastas potentsiaalselt andmemajandusega tegelevatel ametitel ja töösuhete arv kokku EMTAK 2 lõikes. Tunnused:
 - AASTA - andmete aasta
 - EMTAK_2K – EMTAK2 taseme kood
 - toosuhted_it_amet – potentsiaalselt andmetega töötavate ametite töösuhete arv kokku aastas
 - toosuhted_kokku – töösuhete arv kokku aastas
- Leht: **osa1_2k_E2K** – unikaalsete isikute arv ja keskmine koormus EMTAK 2 taseme ja ISCO 2 taseme järgi. Tunnused:
 - AASTA
 - EMTAK_KOOD1
 - EMTAK_KOOD2

- AMET_KOOD2
- unik_inimesed_Q1, unik_inimesed_Q2, unik_inimesed_Q3, unik_inimesed_Q4 – unikaalsete inimeste summa kvartalis
- keskm_koormus_summa_Q1, keskm_koormus_summa_Q2, keskm_koormus_summa_Q3, keskm_koormus_summa_Q4 – keskmine summaarne koormus kvartalis
- Leht: **osa3_detsember** - unikaalsete isikute arv ja keskmine koormus detsembri lõpu seisuga potentsiaalselt andmemajandusega tegelevatel ametitel ja kokku. Tunnused:
 - AASTA
 - ENIMETUS2 – tegevusala nimetus EMTAK2 tasemel
 - unik_inimesed_it_amet - potentsiaalselt andmetega töötavatel ametitel töötavate unikaalsete isikute arv kokku
 - unik_inimesed_muu – teistel ametitel töötavate unikaalsete isikute arv kokku
 - koormus_summa_it_amet – töökoormuste summa kokku potentsiaalselt andmetega töötavatel ametitel
 - koormus_summa_muu – teistel ametitel töötavate isikute töökoormuste summa kokku
- info – taustainfo andmepäringule

Lisa 4. Muude kulude skaleerimisfaktori sisendfail

Muude kulude arvutamise aluseks on Statistikaameti andmed tabelist EM001

https://andmed.stat.ee/et/stat/majandus_ettevetete-majandusnaitajad_ettevetete-tulud-kulud-kasum_aastastatistika/EM001

Andmed võib laadida otse Statistikaameti andmebaasist API kaudu või käsitsi ja lugeda sisse.

Töölaud vaikimisi laeb andmed API kaudu tabelist <https://andmed.stat.ee/api/v1/et/stat/EM001> kasutades järgmist päringut.

```
enterprise_statistics_query = {
    'query': [
        {
            'code': 'Näitaja',
            'selection': {'filter': 'item', 'values': ['V12110', 'CUB_C_100_110_120', 'CUB_C_345_350',
'V13310']},
        },
        {'code': 'Tegevusala', 'selection': {'filter': 'item', 'values': ['J62', 'J63']}},
        {'code': 'Tööga hõivatud isikute arv', 'selection': {'filter': 'item', 'values': ['TOTAL']}},
        {'code': 'Vaatusperiood', 'selection': {'filter': 'item', 'values': ['2020', '2021', '2022']}},
    ],
    'response': {'format': 'csv'},
}
enterprise_statistics_response = requests.post(
    enterprise_statistics_url, json=enterprise_statistics_query, timeout=5
)
```

Alternatiivne võimalus on kasutada csv faili, mis peaks vastama järgmisele kujule:

EM001: ETTEVÖTETE MAJANDUSNÄITAJAD Näitaja, Tegevusala, Vaatusperiood ning Tööga hõivatud isikute arv			
Näitaja	Tegevusala	Vaatusperiood	Kokku
Müügitulu, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	2022	3832206
Müügitulu, tuhat eurot	Infoalane tegevus	2022	491304.5
Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	2022	27230
Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused, tuhat eurot	Infoalane tegevus	2022	2420.2

Muud ostetud teenused, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	2022	2513576.7
Muud ostetud teenused, tuhat eurot	Infoalane tegevus	2022	224744.6
Tööjõukulud, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	2022	1096322
Tööjõukulud, tuhat eurot	Infoalane tegevus	2022	144769.3
Müügitulu, tuhat eurot	Programmeerimine, konsultatsioonid jms tegevused	2021	2504283.6
...

Skaleerimistegur = (Müügitulu, tuhat eurot - 0.5 Müügi eesmärgil ostetud kaubad, kinnisvara ja teenused, tuhat eurot - 0.5* Muud ostetud teenused, tuhat eurot) / Tööjõukulud, tuhat eurot*

Lisa 5. Töölaua näitajate uuendamise protsess

Uuendamaks töölaual kuvatud andmemajandust iseloomustavaid indikaatoreid, mis loodi käesoleva projekti raames, on vajalikud järgmised sammud.

- 1) Andmetega potentsiaalselt tegelevate ametialade määramine ametite neljanda taseme klassifikatsiooni põhjal
- 2) Uuendatud andmepäring statistikaametile TÖR ja TSD andmete põhjal järgmiste näitajate leidmiseks
 - a. TÖR ja TSD andmete põhjal aastaste tööjookulude leidmine tegevusalade esimese (EMTAK 1) ja ametialade neljanda taseme (ISCO 4) lõikes
 - b. TÖR ja TSD andmete põhjal täisajale viidud töökoormuse leidmine tegevusalade esimese (EMTAK 1) ja ametialade neljanda taseme (ISCO 4) lõikes
 - c. TÖR ja TSD andmete põhjal aasta lõpus unikaalsete isikute töösuhete arv tegevusalade teise (EMTAK 2) lõikes; eraldi kes on töötavad potentsiaalselt andmetega ja kõik ülejäänud
- 3) Statistikaametist saadud andmete koondamine faili nimega tulemused_SA.xlsx, eelnev punkt (a) lehele osa1_4k, eelnev punkt (b) lehele osa2_4k (või siis muuta vastavalt Pythoni koodi).
- 4) Uuendatud (või siis sama) eelduste komplekt tööaja kohta, mis on seotud andmebaaside loomise või andmeteabega. Eelduste üleslaadimine kodulehele failina, kus on kaks tunnust (iscocode ja kaal), või siis muuta vastavalt Pythoni koodi.
- 5) Ühekordselt leida tegevusalad teise taseme järgi, kus andmetega potentsiaalselt tegelevate töötajate osakaal on suurem kui 30%.
 - a. Valitud tegevusalade kohta leida summaarsed näitajad Statistikaameti tabelist EM001 (müügitulu, lisandväärtus, investeeringud immateriaalsesse põhivarasse)
 - b. Uuendada Pythoni koodi (faili app.py) uute arvudega
- 6) Ühekordselt kontekstiindikaatorite lisamine (SKP väärtus aastas, IKT sektori käive, täiskohale taandatud töötajate arv majanduses jmt) Pythoni koodi. Soovi korral saab nende uuendamise automatiseerida päringutega Statistikaameti andmebaasist.
- 7) Valida Pythoni koodis, mis aastate ettevõtlusnäitajate põhjal arvutatakse välja skaleerimistegurid tööjookuludele andmemajanduse kulupõhiseks hindamiseks. (Hetkel on vaikumisi kasutusel 2020-2022)

Lisa 6. Töölaua prototüübi kood

Töölaua prototüübi kood pakitud kujul failis "dasso-andmemajandus-e2d473bd4cf4.zip", mis on edastatud tellijale.