

Datasäädöksen soveltamisalaan liittyvät markkinat Suomessa

Heli Koski

Datatalousfoorumi

29.4.2024

LVM:n tilaaman selvityksen tavoite

- Arvioida datasäädöksen soveltamisalan mukaisen liiketoiminnan laajuutta Suomessa.
- Käytämme Tilastokeskuksen toimialatason luokittelua (TOL 2008) määrittämään datasäädöksen soveltamisalaan kuuluvat yritykset ja arvioimaan niiden taloudellisen toiminnan laajuutta.
- Liiketoiminnan laajuutta mitataan liikevaihdolla, arvonnisäyksellä ja dataomaisuuden arvolla yhteensä, toimialoittain & yrityskoon mukaan.
- Laskelmissa esitettyihin lukuihin tuovat epävarmuutta erityisesti datasäädökseen liittyvä tulkinnanvaraisuus ja tilastointikäytäntöihin liittyvät seikat.

Datasäädöksen soveltamisala Tilastokeskuksen aineistojen valossa

Datasäädöksen soveltamisala

- Verkkoon liitettävien laitteiden (Esineiden Internet; IoT) ja niihin liittyvien palveluiden tarjoajat (ml. virtuaaliavustajat), pois lukien tuotteet, joiden pääasiallisena toimintona on datan tallennus, käsittely tai siirto muun osapuolen kuin käyttäjän toimesta.
- **Verkkoon liitettävien laitteiden valmistuksen osalta** datasäädöksen soveltamisalaan kuuluvat selvimmin seuraavat toimialat:

Elektronisten komponenttien ja piirilevyjen valmistus; Mittaus-, testaus- ja navigointivälineiden ja -laitteiden valmistus, kellot; Säteilylaitteiden sekä elektronisten lääkintä- ja terapialaitteiden valmistus; Sähkölaitteiden valmistus; Muiden koneiden ja laitteiden valmistus; Moottoriajoneuvojen, perävaunujen ja puoliperävaunujen valmistus; Muiden kulkuneuvojen valmistus.

Datasäädöksen soveltamisala

- **Verkkoon liitettäviin tuotteisiin liittyvät palvelut:**
 - ✓ Digitaaliset palvelut (pois lukien sähköiset viestintäpalvelut, ml. ohjelmistot), joita ilman kytketty tuote ei toimisi tai jotka on lisätty myöhemmin parantamaan verkkoon liitetyn tuotteen toiminnallisuutta.
 - ✓ Virtuaaliavustajat = ”ohjelmistot, jotka voivat käsitellä pyyntöjä, tehtäviä tai kysymyksiä...”. (Datasäädös rajaa tuotteeseen liittymättömät ohjelmistot pääsääntöisesti sen soveltamisalan ulkopuolelle.)
 - ✓ Datankäsittelypalvelut, ml. pilvi- ja reunapalvelut, esim. ohjelmisto palveluna (software-as-a-service, SaaS).
- Palvelujen tarjoajia löytyy keskeisesti televiestintä, tietopalvelutoiminta (mm. www-hakuportaalit, tiedon käsittely ja palvelintilan vuokraus ja muita toimintoja, jotka mahdollistavat tiedon hakemisen) ja ohjelmistot, konsultointi ja siihen liittyvä toiminta –toimialoilta.

Tilastokäytäntöihin liittyvät ongelmat datasäädöksen soveltamisalan mukaisen toiminnan arvioinnissa

- Rekisteriaineistoilla ei päästä liiketoiminnan luonteeseen kiinni yrityksen toiminnan pääasiallista toimialatasoa tarkemmin.
- Datasäädöksen soveltamisalaan kuuluvia ohjelmistoalan yrityksiä ei voida erotella yrityksistä, joilla ei ole asetuksen soveltamisalaan kuuluvaa liiketoimintaa (vain osa ohjelmistoista kuuluu soveltamisalaan; toimintamallien kirjo on ohjelmistoalalla laaja).
- Velvoitteet eivät koske julkisen sektorin hätätilassa tekemiä datapyyntöjä lukuun ottamatta pieniä yrityksiä (alle 50 työll. tai taseen loppusumma <10 milj. €), paitsi jos omistusyhteys/sidosyritys, joka ei ole mikro- tai pienyritys tai jos verkkoon liitetty tuote/palvelu teetetään alihankintana.

Tilastokäytäntöihin liittyvät ongelmat datasäädöksen soveltamisalan mukaisen toiminnan arvioinnissa

- Tilastoissa käytettävä toimialajako ei pysty kovin hyvin erottelemaan verkkoon liitettävien tuotteiden valmistuksen ja niihin liittyvien palveluiden tarjonnan liiketoiminnan laajuutta.
- ✓ Monilla verkkoon liitettävien koneiden ja laitteiden valmistajilla oheispalvelut muodostavat merkittävät tulovirran, mutta tilastointikäytäntöjen vuoksi tällaiset palvelut luetaan kuitenkin teolliseen toimintaan.
- Verkkoon liitettäviä laitteita valmistavilla toimialoilla voi toimia yrityksiä, joiden tuotevalikoimassa ei ole verkkoon liitettäviä tuotteita tai niihin liittyviä palveluita.
- Emme voi tunnistaa tilastoista yrityksiä, jotka valmistavat verkkoon liitettäviä laitteita tai tarjoavat niihin liittyviä palveluita, mutta joiden pääasiallinen toiminta ei kuulu datasäädöksen soveltamisalaan määritellylle toimialalle.
- Lähtökohtaisesti kaikki toimialat ovat datasäädöksen soveltamisalan piirissä.

Dataomaisuuden arvon määrittäminen

Dataomaisuuden arvon määrittäminen

- Laskelmien tekeminen datasäädöksen soveltamisalaan kuuluvan liiketoiminnan laajuudesta liikevaihdolla ja arvonlisäyksellä mitattuna on toimialarajausten tekemisen jälkeen suoraviivaista.
- Datasäädöksen soveltamisalan mukaisen toiminnan laajuuden selvittäminen datavarantojen osalta on haasteellista, koska yritysten tilinpäätöstiedot eivät kerro niiden datavarantojen arvosta.
- Kustannusperusteinen lähestymistapa: yhdenmukainen kansantalouden tilinpidon standardikäytäntöjen kanssa: mikäli arvoa ei voida havainnoida suoraan markkinatransaktioiden kautta suositellaan käytettäväksi ”kustannusten summa” –menetelmää (datan tuotantokustannukset).

Kustannusperusteinen menetelmä

- Dataomaisuuden arvon määrittäminen kustannusperusteisella menetelmällä perustuu datan luomiseen, keräämiseen, varastointiin ja tietokantojen korvaamiseen liittyvien kustannusten tunnistamiseen ja laskemiseen.
- Oletus: yritykset maksavat datan hyödyntämiseen liittyviä kuluja vain, jos he odottavat saavansa toiminnasta vähintäänkin samanarvoista taloudellista hyötyä. Täten laskelmasta saatu tulos on datatalouden arvon alaraja.
- Laskelmien tekeminen on verrattain helppoa, mikäli rekisteriaineistoja on saatavilla.

Dataomaisuus

- Käytämme Kanadan tilastoviranomaisen (Statistics Canada, 2019) kehittämää viitekehystä ja luokittelemme dataomaisuuden kolmeen pääkategoriaan:
 1. **Data** tarkoittaa digitaalisia havaintoja, jotka voidaan tallentaa, siirtää ja käsitellä, ja joiden avulla voidaan lisätä ymmärrystä erilaisista ilmiöistä.
 2. **Tietokannat** ovat jäsenneiltyjä datavarastoja, jotka on organisoitu tiedonhakua ja käsittelyä varten.
 3. **Datatiede** (data science) tarkoittaa erilaisia tekniikoita, menetelmiä, algoritmeja ja järjestelmiä, joiden avulla yksilöt hankkivat tietoa ja lisäävät ymmärrystä dataa hyödyntämällä.

Datan tuotantokustannukset arvonluonnin mittarina

- Hyödynnämme kv-kirjallisuudessa käytettyä menetelmää, joka perustuu datan tuotantokustannusten laskentaan dataperusteiseen arvonluontiin liittyvien työpanosten kautta.
 - ✓ Edellyttää dataperusteiselle arvonluonnille relevanttien ammattiryhmien tunnistamista sekä arviota kunkin ammattiryhmän dataperusteiseen arvonluontiin keskimäärin käyttämästä ajasta suhteessa koko työaikaan.
 - ✓ Sovellamme aiempaa kv-työtä (Statistics Canada, 2019) ammattiryhmien tunnistamisen ja ajankäytön osalta suomalaisiin rekisteriaineistoihin.
- Tilastokeskuksen ammattiryhmäluokitus poikkeaa kv-tutkimuksissa käytetyistä luokituksista → suomalaisesta ammattiluokituksesta on ensin tunnistettava kansainvälisiä luokkia vastaavat relevantit ammattiryhmät.

Statistics Canada (2019) Ammattiluokitus	Tilastokeskus Ammattiluokitus 2010	Datatyöhön käytetty aika (keskim. %)
Data		
Financial and investment analysts	2413 Rahoitusanalyytikot	10-20
Customer and information services supervisors	121 Liiketoiminta- ja hallintojohtajat	40-50
Other customer and information services representatives		
Data entry clerks	4132 Tallentajat	100
Survey interviewers and statistical clerks	4227 Tutkimus- ja markkinatutkimushaastattelijat	90-100
Mathematicians, statisticians and actuaries	2120 Matematiikan ja tilastotieteen erityisasiantuntijat	20-30
Economists and economic policy researchers and analysts	2631 Ekonomistit	20-30
Tietokannat		
Computer and information systems managers	133 Tieto- ja viestintäteknologiajohtajat	30-50
Database analysts and data administrators	252 Tietokantojen, -verkkojen ja järjestelmäohjelmistojen erityisasiantuntijat	90-100
Information systems testing technicians	351 Informaatio- ja tietoliikenneteknologian asiantuntijat	30-50
Datatiede		
Financial and investment analysts	2413 Rahoitusanalyytikot	50-60
Statistical officers and related research support officers	3314 Tilastointi- ja matematiikka-asiantuntijat	90-100
Mathematicians, statisticians and actuaries	2120 Matematiikan ja tilastotieteen erityisasiantuntijat	50-60
Economists and economic policy researchers and analysts	2631 Ekonomistit	50-60

Dataomaisuuden arvon laskeminen

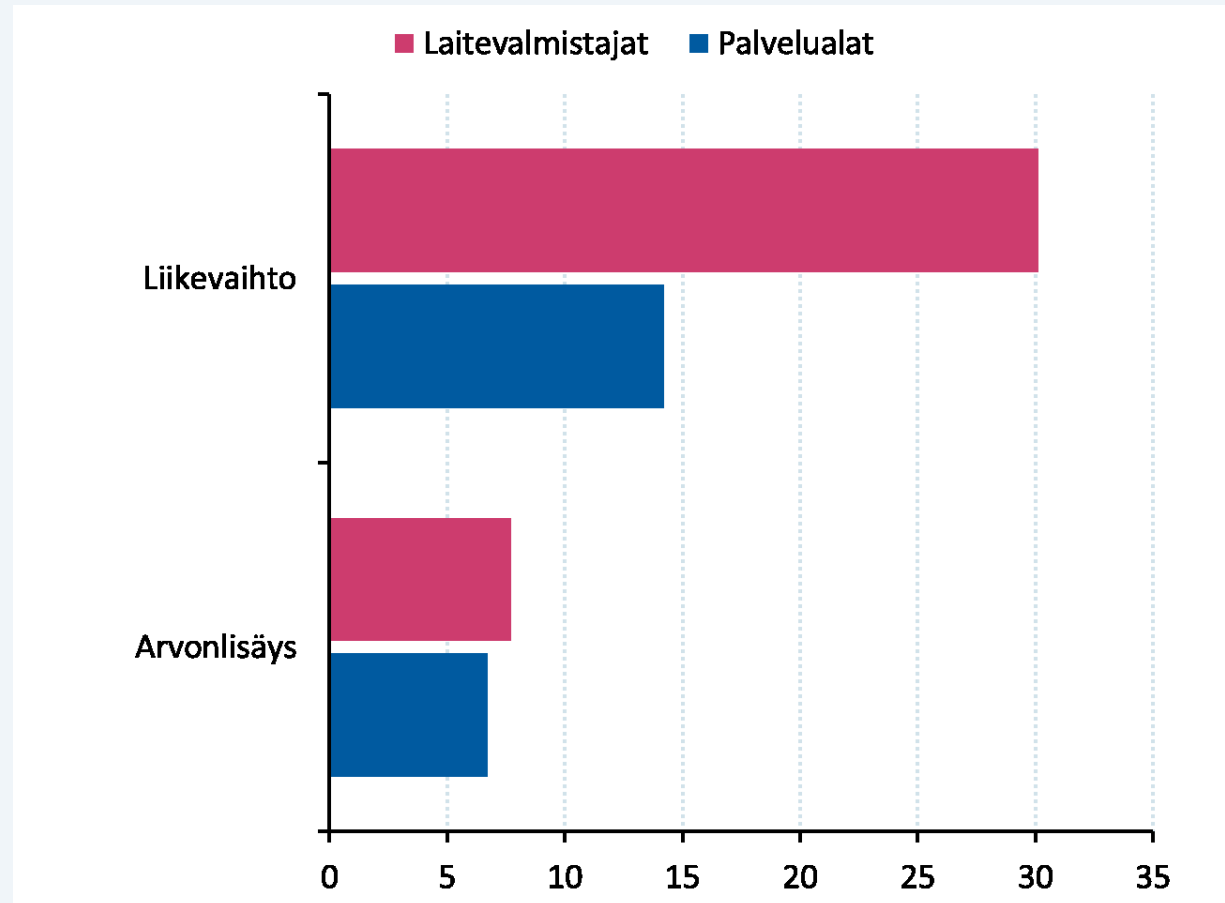
- Dataomaisuuden tuotantokustannukset: palkat + dataomaisuuden luonnin epäsuorat kustannukset (esim. yleiskustannukset).
- Goodridge ym. (2022) soveltavat EU 27-maissa yleisesti käytössä olevaa menetelmää epäsuorien kustannusten huomioimiseksi ja kertovat palkkakustannukset kahdella.
- Dataomaisuuden arvon laskeminen: yleisesti pääomakannan mittaamisessa käytetty investointikertymämenetelmä (Perpetual Inventory Method, PIM).
- ✓ Pääoman kuluminen lasketaan olettaen, että datainvestointien tuotantokapasiteetti pienenee vuosittain kiinteällä osuudella: ohjelmistot, tietokannat & data 33% vuosipoisto; datatiede 20% vuosipoisto (sovelletaan myös t&k-investointeihin).

Laskelmat dataomaisuuden arvosta suuntaa antavia

- Kustannusperusteisella menetelmällä tehtävä arvio datainvestoinneista vaatii oletuksia esimerkiksi eri ammattiryhmien ajankäytöstä ja että käytettyjä datan arvottamismenetelmiä ollaan yhä kehittämässä.
- Esittämämme arviot dataomaisuuden suuruudesta ovat suuntaa antavia.
- Osa yritysten dataomaisuudesta sisältää liikesalaisuuksia. Datasäädöksen puitteissa tuotteen valmistajan ei tarvitse luovuttaa liikesalaisuuksia sisältävää dataa. Käytettävissä olevien tietojen perusteella ei ole kuitenkaan mahdollista arvioida, mikä osa yritysten datavarannoista tullaan lukemaan liikesalaisuuksia koskevaksi.

Datasäädöksen soveltamisalaan kuuluvien yritysten liiketoiminnan laajuus ja dataomaisuuden arvo

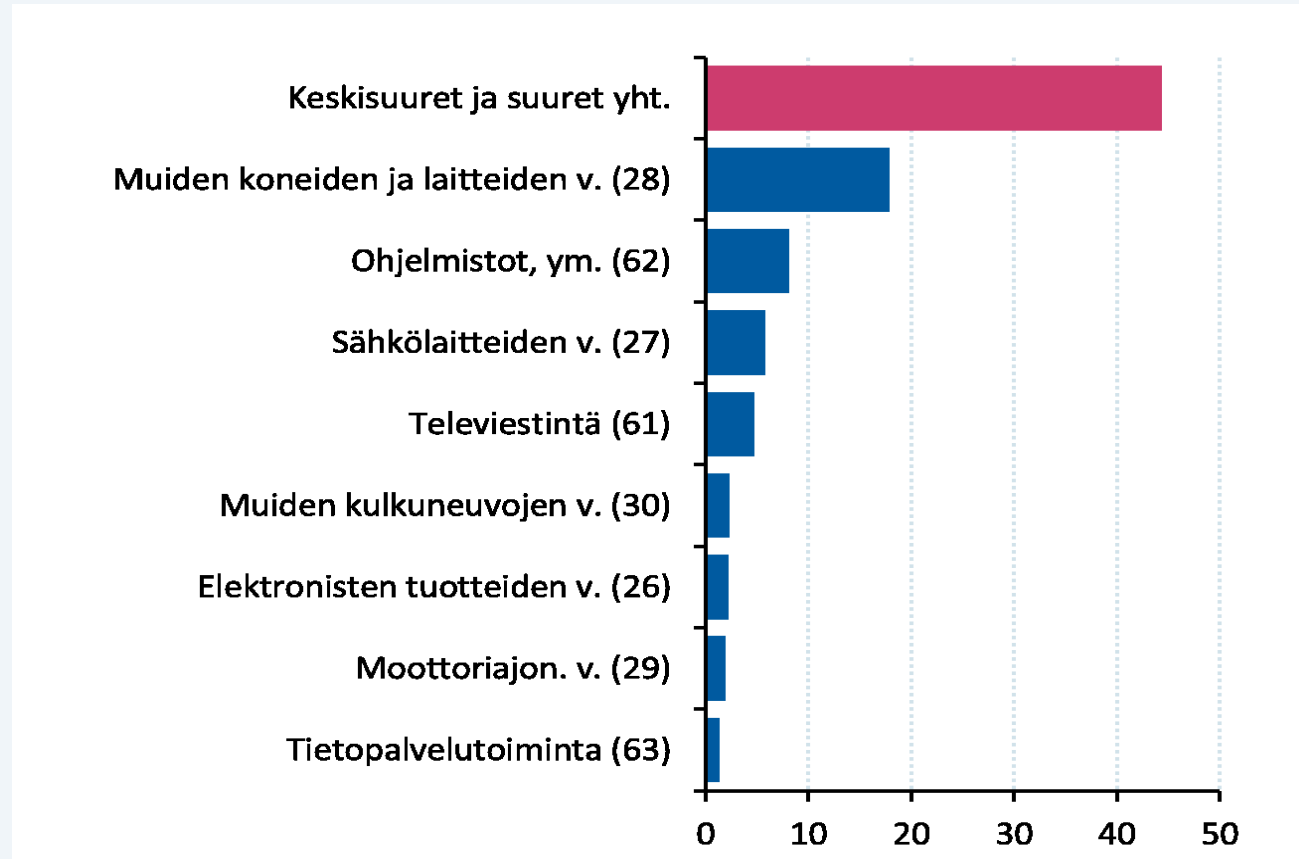
Verkkoon liitettäviä laitteita valmistavien ja niihin liittyviä palveluita tarjoavien toimialojen liikevaihto ja arvonlisäys datasäädöksen soveltamisalaan kuuluvissa keskisuurissa ja suurissa yrityksissä v. 2021, miljardia euroa



Suurten ja keskisuurten yritysten liikevaihto oli v. 2021 yli 44 miljardia euroa ja arvonlisäys 14 miljardia euroa, eli 5,8 % BKT:sta.

Lähde: Koski & Pajarinen (2024)

Liikevaihto datasäädöksen soveltamisalaan kuuluvissa keskisuurissa ja suurissa yrityksissä v. 2021, miljardia euroa



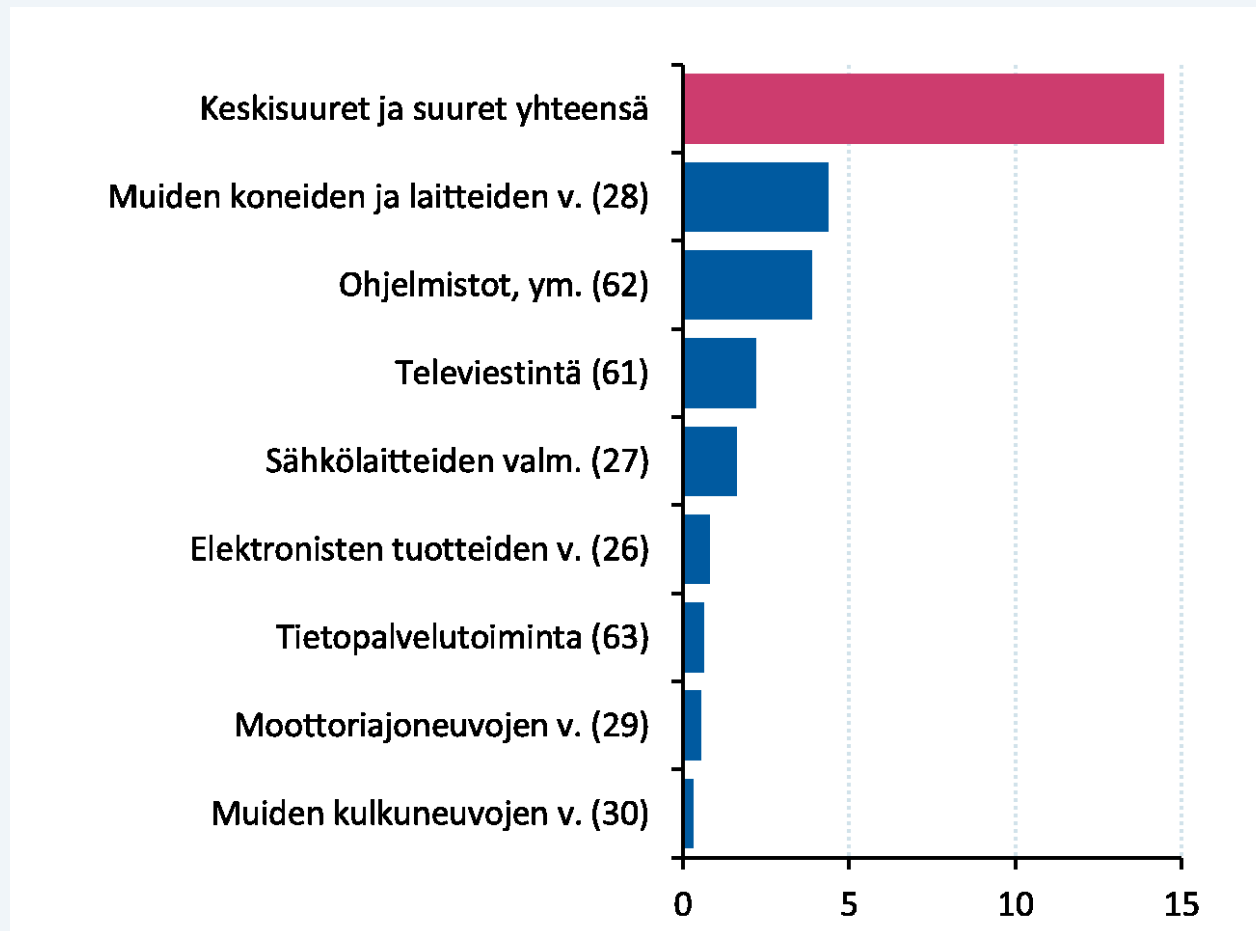
Kaikkien yritysten (ml. pienet) lv:

Verkkoon liitettävien laitteiden valmistus: liki 33 miljardia euroa

Palvelualat: 18 miljardia euroa.

Lähde: Koski & Pajarinen (2024)

Arvonlisäys datasäädöksen soveltamisalaan kuuluvissa keskisuurissa ja suurissa yrityksissä v. 2021, miljardia euroa



Kaikkien yritysten (ml. pienet) arvonlisäys:

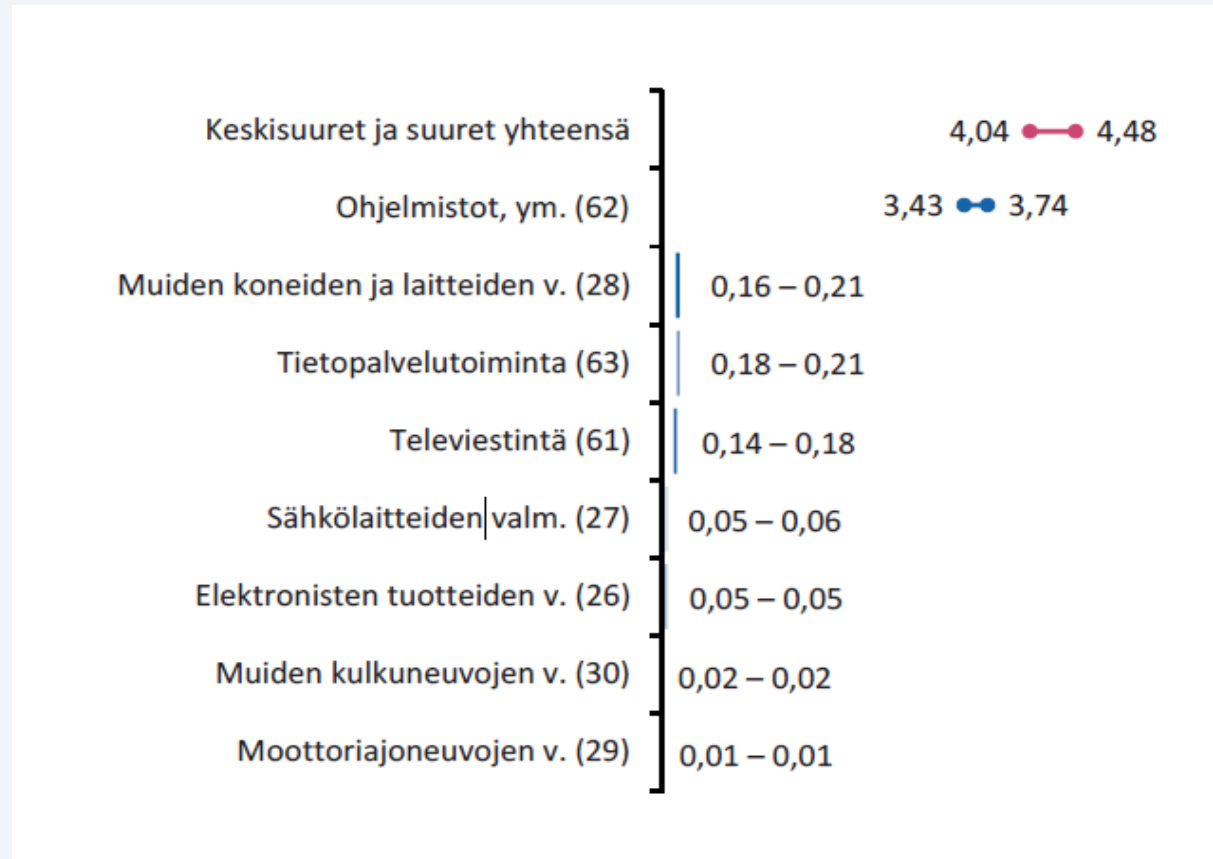
Verkkoon liitettävien laitteiden valmistus: 8,6 miljardia euroa (3,4 % BKT).

Palvelualat: 8,7 miljardia euroa (3,5 % BKT)

Yhteensä yli 17 miljardia euroa eli 6,9 % BKT:sta.

Lähde: Koski & Pajarinen (2024)

Dataomaisuuden arvo datasäädöksen soveltamisalaan kuuluvissa keskisuurissa ja suurissa yrityksissä v. 2021, miljardia euroa



Kaikkien yritysten (ml. pienet) dataomaisuuden arvo:

Verkkoon liitettävien laitteiden valmistus: 314–387 miljoonaa euroa.

Palvelualat: 5,2–5,8 miljardia euroa.

Lähde: Koski & Pajarinen (2024)

Lopuksi

- Esitettyihin laskelmiin liittyy huomattavaa epävarmuutta: i) Datasäädöksen soveltamiseen liittyy tulkinnanvaraisuutta ja tulkinta tulee muotoutumaan vasta oikeuskäytäntöjen perusteella & ii) Rekisteriaineistoilla ei päästä yrityksen liiketoiminnan luonteeseen kiinni pääasiallista toimialatasoa tarkemmin.
- Data on vuodelta 2021 (tilastoviive) ja mm. generatiivisen tekoälyn laajamittainen hyödyntäminen tulee muuttamaan liiketoimintamalleja ja yritysten hallussa olevia datavarantoja ennen syksyä 2025, jolloin datasäädöstä aletaan soveltaa.
- Tarkennuksia datasäädöksen soveltuvuuden osalta ja päivitettyjä laskelmia tietopohjaksi datasäädöksen soveltamisalaan kuuluvan liiketoiminnan laajuuden osalta voidaan tehdä vasta, kun datasäädöstä on alettu soveltamaan.