



Tiedon laatukriteerit ja mittaristo – soveltamisohje

Tiedon laatukehikko, TiHA TP3

Tekijät:

Janika Tarkoma
Antti Jakobsson
Mervi Haakana
Jari Kosonen
Jarmo Ranki

tiedonlaatu@stat.fi

Kansikuva: iStock
Kannen suunnittelu: Riikka Turunen
Taitto: Riikka Turunen

Tietoja lainattaessa lähteenä on mainittava Tilastokeskus.

© 2023 Tilastokeskus

Helsinki 2023

Sisällys

1	Johdanto	5
2	Laatukriteerien mittariston soveltaminen käytännössä	6
2.1	Erilaisia tietoaaineistoja	8
2.2	Arviointitasot	8
2.3	Mittaristopohja	9
2.4	Käytetyistä termeistä	10
3	Laatukriteeriryhmä: Miten tieto kuvaa todellisuutta?	11
3.1	Laatukriteeri: Virheettömyys	11
3.1.1	Menetelmällisesti tuotetut arvot	12
3.1.2	Virheelliset ominaisuusarvot	12
3.1.3	Väärinluokittelu	13
3.2	Laatukriteeri: Tarkkuus	14
3.2.1	Keskihajonta	14
3.2.2	Poikkeavat havainnot	15
3.3	Laatukriteeri: Johdonmukaisuus	16
3.3.1	Tieto loogisuustarkastettu	16
3.4	Laatukriteeri: Ajantasaisuus	17
3.4.1	Viiteajankohta	18
3.4.2	Luontiajankohta	18
3.4.3	Tarkistusajankohta	20
3.4.4	Muutosajankohta	20
3.5	Laatukriteeri: Kattavuus	21
3.5.1	Ajallinen tavoitekattavuus	22
3.5.2	Alueellinen tavoitekattavuus	22
3.5.3	Tavoitekohdeyksiköt	23
3.5.4	Tavoiteominaisuuksissa puutteita	24
3.5.5	Puuttuvat kohdeyksiköt	24
3.5.6	Ylimääräiset kohdeyksiköt	26
3.5.7	Puutteelliset kohdeyksiköt	27
3.5.8	Puutteelliset ominaisuustiedot	28
4	Laatukriteeriryhmä: Miten tieto on kuvattu?	29
4.1	Laatukriteeri: Alkuperäisyys	29
4.1.1	Tietolähde	30
4.1.2	Tiedon elinkaari	31
4.1.3	Muutosten hallinta	32
4.2	Laatukriteeri: Metatietojen ymmärrettävyys	33
4.2.1	Aineistokuvaus	33
4.2.2	Käsitteiden määritelmät	34
4.2.3	Ominaisuuksien tietokuvaukset	35
4.2.4	Ymmärrettävyyden asiakaspalaute	35
4.3	Laatukriteeri: Suositustenmukaisuus	36
4.3.1	Noudatettavat säädökset ja standardit	37

5	Laatukriteeriryhmä: Miten tietoa voi käyttää?	38
5.1	Laatukriteeri: Koneluettavuus	38
5.1.1	Tietoaineiston tietomalli	38
5.1.2	Kohdeyksikön pysyvä tunniste	39
5.1.3	Koneluettavuuden asiakaspalautte	40
5.2	Laatukriteeri: Käyttöoikeudet	41
5.2.1	Käyttöoikeus	41
5.2.2	Käytön rajoitukset	42
5.3	Laatukriteeri: Oikea-aikaisuus	43
5.3.1	Määräpäivien noudattaminen	43
5.3.2	Päivitystiheys	44
5.3.3	Päivityksessä muuttuneet ominaisuustiedot	45

1 Johdanto

Tämä on pilotoinnin pohjaksi tuotettu ensimmäinen versio Tiedon laatukriteerien ja mittariston soveltamisohjeesta. Dokumentti perustuu Tiedon laatukriteerit ja mittaristo -suositusehdotukseen, joka on soveltamisohjeen tämän version kirjottamishetkellä tiedonhallintalautakunnan jatkotyöstössä lautakunnan suositukseksi.

Tiedon laatukriteerien ja mittariston soveltamisohje on laadittu tiedon laatuvausten käyttöönoton helpottamiseksi. Soveltamisohjetta on tarkoitus käyttää tiedonhallintalautakunnan suosituksen rinnalla ja siksi ohje ei sisällä kaikkea suosituksessa olevaa tietoa vaan keskittyy erityisesti mittariston arvojen tuottamiseen. Soveltamisohjeeseen on kerätty aktiivisesti esimerkkejä, jotka helpottavat mittarien ymmärtämistä ja tuovat selkeitä ohjeita mittarin arvon tuottamiseen. Esimerkkejä toivotaan kuitenkin lisää, koska ensimmäisten versioiden esimerkit ovat keskittyneet laatukehikon tuottamisessa aktiivisimpien tahojen tietoaaineistojen malleihin. Voit toimittaa esimerkkejä soveltamisohjeen käyttöön sähköpostilla tiedonlaatu@stat.fi.

Soveltamisohjeen rinnalle on saatavilla myös excel mittaristopohja, joka sisältää mittariston mittarit ja hyvin tiiviit ohjeet mittarin arvon tuottamiseksi. Mittariston mittarit jaoteltiin mittaristopohjaan paitsi laatukriteeriryhmittäin myös mittarin arviointitason mukaan. Tietoaaineistotasoa tarkastelevat mittarit pyrkivät luomaan kuvan tietoaaineiston laadusta yleensä ja ominaisuustason tarkastelulla tuodaan lisätietoa keskeisimpien ominaisuuksien laadusta.

Soveltamisohjeessa esittelemme kaikki tiedon laatukriteerien mittarit uudelleen painottaen nimenomaan käytännön näkökulmiin. Soveltamisohjeen alussa saat yleistietoa tekniikoista, joita mittarien arvojen määrittämisessä kannattaa hyödyntää. Yksikkötasoinen tietoaaineisto on luultavasti selkein esimerkki, koska mittarien arvoja voidaan laskea suoraan tietoaaineistosta. Kaikissa tietoaaineistoissa jokaisen kohdeyksikön tietojen laadun arviointi ei ole mahdollista näin suoraviivaisesti ja siksi laatuarvion tuottaminen vaatii arvioinnin hyödyntämistä. Arviointi voi perustua esimerkiksi laadun määrittelyä varten poimittuihin otostutkimuksiin tai prosessointihistorian antamaan kokemukseen.

Soveltamisohjeen yleisesittelyosan jälkeen esittelemme mittaristoa laatukriteerien ryhmittelyn mukaisessa järjestyksessä. Kunkin mittarin kohdalla esitellään mittarin nimi, sen määrittely, arviointitaso, ehdotettu esitysmuoto ja soveltamisen ohjeita eri tyyppisten tietojen tapauksessa. Lisäksi mahdollisimman monen mittarin kohdalle olemme koonneet esimerkkejä soveltamisesta.

2 Laatumittariston soveltaminen käytännössä

Kullekin laatukriteerille on haettu kyseistä kriteeriä kuvaavia mittareita, joiden arvot tuovat konkreettista ymmärrystä tarkasteltavasta laatu- ja ominaisuuskulmasta. Käsiteltävän tiedon tyyppi vaikuttaa paljon siihen, miten mittarien arvoja on edes mahdollista tuottaa. Yksikkötasolla olevien perusaineistojen mittarien arvojen tuottaminen lienee selkein vaikkakin ehkä työläin tapaus. Mittarien laatuarviot kohdistuvat joko koko tietoaineistokokonaisuuteen tai yksittäisiin ominaisuuksiin. Ominaisuuksien mittarien arvoja toivotaan lopulta tuotettavan jokaiselle aineistossa olevalle ominaisuudelle vaikka aluksi kannattaa keskittyä keskeisimpiin ominaisuuksiin.

Laatumittariston mittarien arvoja tuottaessa ensisijaisesti pyritään varsinaisista ominaisuuden tai tietoaineiston saamista tiedoista tai arvoista laskemalla tuottamaan tarkka mittarin arvo aina kun se on mahdollista. Aina haluttua tietoa ei kuitenkaan ole mahdollista laskea ja tällaisissa tilanteissa voidaan ilmoittaa arviointiin perustuvia arvoja.

Laadun arvioiden olisi tärkeää perustua testaukseen tai muuten selkeästi perusteltavissa oleviin havaintoihin. Testaamista voidaan toteuttaa suoralla testaamisella erilaisin testimenetelmin. Esimerkiksi rekisterissä olevien arvojen oikeellisuutta voidaan tarkastaa pienen otoksen avulla kysymällä tai muuten tarkastamalla otoksen kohdeyksiköiden ominaisuuksien arvot ja verrata niitä rekisterissä oleviin. Epäsuoraksi testaamiseksi kutsutaan tekniikoita, joissa laatua arvioidaan hyödyntämällä aiempaa tietoaineiston prosessointia. Jos tiedon prosessointihistoria tunnetaan, voidaan tehdä hyvin osuvia arvioita esimerkiksi tietoaineiston sisältämien virheiden osuudesta ellei tiedon lähteissä tai keruussa ole tapahtunut suuria muutoksia.

Laadun arviot voivat perustua myös asiakastarpeen perusteella asetettuun laatutavoitteeseen. Tiedon laadussa voidaan puhua täydellisen laadun tavoittelusta, mutta käytännössä täydellistä laatua tuskin saavutetaan eikä se ole välttämättä edes mielekästä resurssoinnin näkökulmasta. Täydellisyyden tavoittelun sijaan onkin parempi tähdätä riittävään laatuun, jonka tason määrittelyssä asiakastarpeet ovat keskeisin tekijä. Tiedon laadun arviot voivat perustua myös tällaiseen asetettuun laatutavoitteeseen, jolloin tarkastelu keskittyy siihen, onko laatutavoite saavutettu vai ei. Esimerkkinä tällaisesta laatutarkastelusta voit tutustua Maanmittauslaitoksen [Maastotietokannan](#) laatutavoitteisiin.

Kuva 1. Mittarin arvon tuottaminen voi perustua useampaan erilaiseen lähestymistapaan



2.1 Erilaisia tietoaineistoja

Tieto tai tietoaineisto voi hyvinkin erilaista organisaatiosta ja tiedon käsittelyn tasosta riippuen. Tiedon laatukriteerien kehittämisessä on rajauduttu rakenteisen datan käsittelyyn, mutta rakenteista dataakin on hyvin monenlaista. Seuraavassa esittelemme muutaman esimerkin siitä, mitä milläkin tietoaineistotyypillä tarkoitamme.

Yksikkötasoisella tietoaineistolla tai yksikköaineistolla tarkoitamme rekisteriaineistotyypistä tietoa, jossa jokaisella rivillä on oma kohdeyksikkönsä ja sarakkeisiin on merkitty kyseisen kohdeyksikön saamat ominaisuuksien arvot.

Koostetulla tietoaineistolla tarkoitetaan jollakin tavalla yksikkötason aineistosta summattua tietoa. Tällaista on esimerkiksi tilastotieto tietystä ilmiöstä. Esimerkiksi tilastossa kuvataan kuinka monta henkilöä Suomessa asuu ja jaotellaan henkilömäärät tarkemmin ominaisuuksien kuten asuinpaikan, iän, sukupuolen tai koulutuksen suhteen. Tilastossa voidaan kuvata toisaalta kuinka paljon liikevaihtoa yritykset, jotka kuuluvat tiettyyn kokoluokkaan, tietyille toimialoille ja sijaitsevat tietyissä maakunnissa, tuottavat vuoden aikana yhteensä eli tietynlaisten yritysten liikevaihdon summaa. Koostettu aineisto voi kertoa myös ominaisuuksien muita tunnuslukuja kuten mediaaneja tai keskiarvoja kutakin tarkasteltavaa kohdeyksikköryhmää kohti.

Koostettua tietoaineistoa tarkastellessa tulee mittarinarvon määrittämisessä pysyä tilaston raportointitasolla. Esimerkiksi puuttuvuutta kuvataan sillä tasolla, millä tietoja on kuvattu tilastossa vaikka puuttuvuudesta voisi antaa kuvauksen myös tarkemmalla luokituksen tasolla. Esimerkiksi, jos tilastossa kerrotaan yrityksen toimiala luokituksen kirjainta käyttäen eli hyvin yleisellä tasolla, tilastotietoon kohdistuu puuttuvuutta vain, jos joku kirjaintason luokka puuttuu kokonaan. Tätä tasoa tarkemman luokituksen laatua kuvataan yksikkötasoisien tietoaineiston laatu-kuvauksessa, joten tiedot ovat tarvittaessa käytettävissä sitä tarkastelemalla.

2.2 Arviointitasot

Mittaristossa on jaettu mittarin kohdistumisen arviointitaso kahteen luokkaan: tietoaineisto ja ominaisuus. Tämä tarkoittaa sitä, että osa

mittareista arvioi koko tietoaaineistoa ja osa kohdistuu erityisesti yksittäisten ominaisuuksien laatuun. Mittaristossa on myös mittareita, joiden arvo tulisi tuottaa sekä tietoaaineistolle että ominaisuuksille. On hyvä huomata, että kaikkia tietoaaineiston sisältämiä ominaisuuksia tarvitse heti tarkastella vaan aluksi on mielekästä kuvata keskeisimpien ominaisuuksien laatua.

Koko tietoaaineistoon kohdistuvilla mittareilla pyrimme tuomaan yleiskuvan tietoaaineistosta ja sen laadusta ylätasolla tarkasteltuna. Monelle tiedon käyttäjälle on kuitenkin tärkeää saada tietoa myös keskeisimpien tietoaaineiston ominaisuuksien laadusta. Vaikka tietoaaineisto olisi tuotettu käyttäjän kannalta olennaisia standardeja käyttäen, voi väärä ominaisuuden luokitus tuottaa haasteita tietoaaineiston hyödyntämiseen ja laatuksivauksen avulla käyttäjä osaa valmistautua tarvittaviin lisätoimenpiteisiin.

2.3 Mittaristopohja

Laadun arvioinnin tueksi tarjottava pilotointipohja on excel-pohja, joka sisältää oman välilehtensä koko tietoaaineistoon kohdistuvien mittarien käsittelyyn ja oman välilehtensä esimerkiksi valituille noin viidelle tyypillisimmälle tai keskeisimmälle ominaisuudelle. Tarkasteltavien ominaisuuksien määrää on pilotointipohjan esimerkkiluonteen takia rajattu. Pohjasta on saatu näin selkeämpi ja helpommin havainnollistava.

Mittaristopohjan tarkoitus ei ole toimia mallina laatutietojen kokoamiselle vaikka sitä voi luonnollisesti myös tähän tarkoitukseen hyödyntää. Pohjan varsinainen tarkoitus on olla selkeä koontipaikka, jonne mittarien arvoja voi aluksi viedä samalla kun tiedon laatukriteereihin vielä tutustuu.

Mittaristopohjassa on useita täytettäväksi tarkoitettuja välilehtiä: tietoaaineiston tiedot, mittarit tietoaaineisto, mittarit ominaisuudet ja yleiset huomiot. Ensimmäisellä sisällysluettelo-välilehdellä on kooste mittareista ja hyvin tiivis yleisohje mittaristopohjan hyödyntämiseksi.

Mittaripohja on tarkoitettu käytettäväksi soveltamisohjeen rinnalla, mutta listauksiin on tiivistetty mittarin kuvaus ja sen arvon tuottaminen mahdollisimman selkeästi. Ajatuksemme on ollut, että mittaristopohjaa

pystyisi jatkossa käyttämään itsenäisemmin ilman muuta mittariston kuvailumateriaalia. Toisaalta mittaristopohjalta voi myös poimia sopivia ohjeistuksia omiin laatukuvausten materiaaleihin hyvin tiiviissä muodossa.

Mittaristopohjaa käyttävän on ajateltu täyttävän mittarit-välilehdiltä sarakkeet: mittarin saama arvo, sanallinen kuvaus, linkki lisätietoihin ja arvon muodostamistapa. Näistä viimeksi mainittu, arvon muodostamistapa, on sarake, joka on lisätty lähinnä käyttäjän omia muistiinpanoja varten. Projektin pilotoinnissa saimme palautetta, että muistiinpanot heti mittarin arvon rinnalla ja tuottamisen yhteydessä kirjattuna, toimivat jatkoa varten erinomaisina ohjeina seuraavaa laadun arviointia varten. Ajatuksena on ollut, että mittarin arvojen vertailukelpoisuus kierroksesta toiseen varmistuu, kun tuottamisen yksityiskohdat tulee kirjatuksi heti selkeästi.

Mittarin saama arvo -sarakeeseen haetaan varsinainen mittarin arvo. Tämä ilmoitetaan rajatulla esitysmuodolla. Mahdolliset lisätiedot annetusta arvosta kirjataan sarakeeseen sanallinen kuvaus. Mikäli tarkempaa tietoa mittarin arvosta tai muuta siihen liittyvää lisätietoa on saatavilla nettisivuilla, linkki lisätään linkki lisätietoihin -sarakeeseen.

2.4 Käytetyistä termeistä

Tiedon laatuksiteereistä ja erityisesti mittareista sekä niiden soveltamisesta puhuttaessa käytetään jotain ilmauksia, joiden merkitys ei ole kaikille ilmeistä. Laatuksiteereihin liittyvää terminologiaa on avattu tarkemmin tiedonhallintalautakunnan suosituksessa. Tässä avataan vain soveltamisohjeessa esiin nousevat uudet termit.

Rakenteellisella puuttuvuudella tarkoitetaan sellaisia arvoja, jotka ovat tietoaineistossa mahdottomia tai käytännössä mahdottomia. Esimerkkejä rakenteellisesta puuttuvuudesta voisivat olla lasten siviilisäätynä muu kuin naimaton tai Suomessa menestymättömien kasvien viljely tai rakennuksen tyyppinä asuntolaiva. Yleensä tällaisia arvoja merkitään aggregoidussa tiedossa nollaksi tai luokan arvot on jätetty suoraan ilmoittamatta.

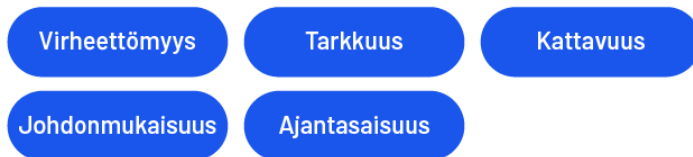
Tekstissä käytetään tietoa tarkastelevasta henkilöstä termejä **tiedon käyttäjä** ja **asiakas**. Näillä termeillä ei ole toisistaan eroavia konnotaatioita vaan ne ovat toistensa synonyymejä samalle asialle.

3 Laatuksriteriryhmä: Miten tieto kuvaa todellisuutta?

Ensimmäisen laatuksriteriryhmän kriteerit keskittyvät kuvaamaan sitä, mistä tietoaaineistossa on kysymys: miten hyvin tietoja on käytettävissä ja miten oikein ja tarkasti nämä tiedot kuvaavat todellista tilannetta. Tietojen ajantasaisuus kuuluu lisäksi tähän ryhmään.

Kuva 2. Laatuksriterit kysymyksen ”Miten tieto kuvaa todellisuutta?” alla.

Miten tieto kuvaa todellisuutta?



Miten tieto on kuvattu?



Miten tietoa voi käyttää?



3.1 Laatuksriteri: Virheettömyys

Synonyymit: **oikeellisuus**

Kuvaus: Virheettömyys kuvaa sitä, miten tietoaaineiston tiedot vastaavat todellisuutta. Tiedon oikeellisuutta tarkastelemalla voidaan saada kiinni myös systemaattisia vääristymiä tietoaaineistossa.

Esimerkit: Se tieto, mikä operatiiviseen päätökseen päätyy, on käytännössä paras käsitys siitä, mikä tieto on oikea. Tieto on oikeellista esimerkiksi tilanteessa, jossa verotuksessa kerrottu palkka on sama kuin todellisuudessa maksettu palkka.

3.1.1 Menetelmällisesti tuotetut arvot

Menetelmällisesti tuotetut arvot -mittari kuvaa tietoaineiston menetelmällisesti tai korvaavaa tietoa käyttäen tuotettujen ominaisuuden arvojen osuuden ominaisuuden kaikista arvoista.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia/ei relevantti

Mittarin arvo lasketaan menetelmällisesti eli laskennallisesti tai korvaavaa tietoa käyttäen tuotettujen ominaisuuden arvojen lukumäärän suhteena ominaisuuden kaikkien arvojen lukumäärään. Korvaavalla tiedolla tarkoitetaan arvoja, jotka on tuotettu käyttäen laskennallisia menetelmiä tai muuta korvaavaa tietoa, joka ei välttämättä ole täsmälleen kohdeyksikön aito arvo. Esimerkiksi edellisen vuoden aikana saatujen tietojen käyttäminen puuttuvien tietojen täydentämisessä on menetelmällisesti tuotettujen arvojen käyttämistä. Tiedon korjaaminen suoraan tiedonantajalta selvittämällä saadulla oikealla arvolla ei kuulu tämän mittarin tarkasteluun.

Mittarin soveltaminen:

Tietoaineistoon menetelmällisesti tuotetut arvot ovat yleensä merkitty aineistoon jollakin merkinnällä. Tarkastellaan ominaisuus kerrallaan näitä merkintöjä hyödyntäen kuinka suuri osuus kaikista ominaisuuden saamista arvoista on menetelmällisesti tuotettuja arvoja.

Esimerkit: Tulotiedon täydentäminen luovuttajaimputoinnilla käyttäen sukupuolta, ikää, koulutusta ja ammattia arvon luovuttajan valinnan kriteerinä. Pienten yritysten toimintaa kuvaavia tietoja kysytään harvemmin kuin kerran vuodessa ja arvoina käytetään aiemmassa tiedustelussa saatuja tietoja.

3.1.2 Virheelliset ominaisuusarvot

Virheelliset ominaisuusarvot -mittari kertoo tietoaineiston virheellisiä ominaisuusarvoja saavien kohdeyksiköiden osuuden kaikista kohdeyksiköistä.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia/ei relevantti
- Mittarin tausta: ISO 19157 (id 63 JHS 160)

Mittaria sovellettaessa tulee huomioida, että arvo ilmoitetaan virheellisen ominaisuusarvon saavien kohdeyksiköiden lukumäärän suhteena kaikkien ominaisuuden arvon saavien kohdeyksiköiden lukumäärään.

Mittarin arvon tuottamisessa voi hyödyntää vertailuun soveltuvaa tietoaineistoa tai tehdä laadun tarkastusta otosperusteisesti. Mikäli mittarin täsmällisen arvon laskeminen on haastavaa, voidaan antaa kokemukseen perustuva arvio prosenttiarvoksi.

Mittarin soveltaminen:

Virheelliset arvot pyritään korjaamaan tiedon käsittelyvaiheessa, mutta jos aineistossa on syystä tai toisesta virheellisyyttä, tulee sen osuutta kuvata.

Mittari lasketaan ominaisuuden saavien virheellisten arvojen osuutena yksikkötason aineiston kohdeyksiköiden lukumäärästä.

Esimerkit: Tietoaineistossa tiedetään olevan päivittämättömiä tietoja ja kokemuksen perusteella osataan arvioida tulevien päivitysten kohdistuvan 5 % ominaisuuden arvoista.

3.1.3 Väärinluokittelu

Väärinluokittelu-mittari kertoo tietoaineiston virheellisesti luokiteltuja ominaisuuden arvoja saavien kohdeyksiköiden osuuden kaikista kohdeyksiköistä.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia/ei relevantti
- Mittarin tausta: ISO 19157 (id 63 JHS 160)

Mittaria sovellettaessa tulee huomioida, että arvo ilmoitetaan väärinluokiteltuja ominaisuuden arvoja saavien kohdeyksiköiden lukumäärän suhteena kaikkien kohdeyksiköiden lukumäärään. Väärinluokitelluksi voidaan määritellä myös ne kohdeyksiköt, joille ei ole saatu korjattua puuttuvaa tietoa ja joiden puuttuvuus ei ole rakenteellista puuttuvuutta eli ominaisuus on kohdeyksikölle relevantti.

Mittarin arvon tuottamisessa voi hyödyntää vertailuun soveltuvaa tietoaineistoa tai tehdä laadun tarkastusta otosperusteisesti. Mikäli täsmällisen

arvon laskeminen on haastavaa, voidaan antaa kokemukseen perustuva arvio prosenttiarvoksi.

Mittarin soveltaminen:

Väärinluokiteltuja kohdeyksiköitä ja arvoja voidaan löytää tietoaaineiston loogisuustarkistusten avulla. Loogisuustarkistus paljastaa esimerkiksi yksiköt, joiden sijaintikunta- ja postinumerotieto eivät vastaa toisiaan.

Mittari lasketaan ominaisuuden saamien väärinluokiteltujen arvojen osuutena yksikkötason aineiston kohdeyksiköiden lukumäärästä (poislukien rakenteellinen puuttuvuus).

Esimerkit: Rakennuksen käyttötarkoitus on määritelty väärin n% kaikista rakennuksista.

3.2 Laatuksiteeri: Tarkkuus

Synonyymit: **harhattomuus**

Kuvaus: Tarkkuus kuvaa sitä, miten hyvin tietoaaineiston tiedot vastaavat sitä mitä tavoitellaan. Tarkkuus kuvaa sitä, kuinka hyvin tieto osuu oikeaan.

Esimerkit: Esimerkkejä tarkkuudesta ovat mitattujen arvojen hajonta ja tietoaaineiston poikkeavien havaintojen osuus sekä lisäksi luokittelun osuvuus ja mittaustaso esim. desimaalit, ajan tai koordinaattien mittaustaso.

3.2.1 Keskihajonta

Keskihajonta-mittari kuvaa ominaisuuksien arvojen vaihtelua keskiarvon ympärillä. Mittarin tarkoitus on antaa tiedon käyttäjälle käsitystä ominaisuuden arvojen hajonnasta.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: keskihajonta

Mittarin arvon on tarkoitus tuoda tiedon käyttäjälle tietoa siitä, miten laajaa ominaisuuden arvojen vaihtelu on.

Lisäksi on hyvä huomioida, että keskihajonnan arvon rinnalla olisi syytä olla myös keskiarvo, jotta arvojen vaihtelua voi hahmottaa paremmin. Ilman keskiarvoa vaihtelu jää asteikosta erilliseksi. Keskiarvo sinällään ei kuitenkaan kuvaa tarkkuutta, joten keskiarvoa ei ole tuotu mittaristoon erillisenä mittarina. Keskiarvo ja keskihajonnan arvojen tuottaminen on olennaista myös siksi, että sekä molempia tarvitaan poikkeavien havaintojen määrittämisessä.

Mittarin soveltaminen:

Keskihajonta voidaan laskea numeerisille, jatkuvia muuttujanarvoja saaville ominaisuuksille.

Esimerkit: Keskihajonnan ollessa pientä, ovat ominaisuuden arvot keskittyneet lähelle keskiarvoa ja vastaavasti suuren keskihajonnan tapauksessa arvot ovat enemmän hajallaan. Tämä voi liittyä epä-tarkkuuteen tai sitten olla vain ominaisuudelle tyypillistä vaihtelua. Esimerkiksi täysiaikaisena syntyneiden lasten pituuden keskihajonta on suhteellisen pientä, mutta iän myötä jo koulutulokkaiden pituuksissa on enemmän vaihtelua. Molemmat liittyvät lasten kasvamiseen eivätkä kerro tietoaineiston laadusta vaan pituus-ominaisuuden käyttäytymisestä.

3.2.2 Poikkeavat havainnot

Poikkeavat havainnot -mittari kuvaa tietoaineiston ominaisuuksien poikkeavien havaintojen osuutta kaikista kohdeyksiköistä.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia
- Mittarin tausta: ISO 25024

Mittaria sovellettaessa tulee huomioida, että mittarin arvo ilmoitetaan poikkeavien havaintojen lukumäärän suhteen kaikkien kohdeyksiköiden lukumäärään. Poikkeavalla havainnolla tarkoitetaan kohdeyksikköä, joka saa suuresti havaittujen arvojen enemmistöstä poikkeavan arvon.

Tyypillisenä raja-arvona poikkeavalle havainnolle pidetään 2,5 kertaa keskihajonnan verran keskiarvosta mitattuja arvoja ja tätä raja-arvoa käytetään myös tämän mittarin yhteydessä. Eli sellaiset arvot, jotka ovat pienempiä kuin arvo, joka saadaan, kun keskiarvosta vähennetään 2,5 kertaa keskihajonta tai suurempia kuin arvo, joka saadaan, kun keskiarvoon lisätään

2,5 kertaa keskihajonta, ovat poikkeavia havaintoja. Monissa tilastollisissa ohjelmistoissa poikkeavat havainnot on helppo määrittellä koneellisesti.

Lisäksi on hyvä huomioida, että poikkeava havainto voi olla oikea arvo tai virheellinen arvo.

Mittarin soveltaminen:

Mittari lasketaan kunkin ominaisuuden sisältämien poikkeavien havaintojen osuutena yksikkötason aineiston kohdeyksiköiden lukumäärästä.

Esimerkit: Tarkastellessa henkilöiden tulotietoja, huomattavan suurituloiset ovat usein henkilöaineistossa poikkeavia havaintoja. Huomattavan suurituloisten tulotiedot voivat aiheuttaa haastetta analyysissä ja siksi poikkeavia havaintoja poistetaan tietoaineistoista ennen analysointia.

3.3 Laatuksiteeri: Johdonmukaisuus

Synonyymit: **säännönmukaisuus, tiedon looginen eheys**

Kuvaus: Johdonmukaisuus kertoo siitä, että tietoaineisto on yhtenäinen ja ristiriidaton. Johdonmukaisuudella voidaan kuvata myös eri tietoa-ineistojen keskinäistä johdonmukaisuutta.

Esimerkit: Esimerkiksi seuraavat tilanteet eivät ole johdonmukaisia: asuinrakennuksella ei ole yhtään asuntoa tai henkilön avioliiton solmimispäivä on aikaisempi kuin syntymäpäivä. Johdonmukaisuutta voidaan tarkistaa tarkistussäännöillä/laatusäännöillä.

3.3.1 Tieto loogisuustarkastettu

Tieto loogisuustarkastettu -mittari kuvaa sitä, onko aineistoa tarkasteltu loogisuusehdoin tietoa-ineiston koostamisessa tai käsittelyssä.

- Arviointitaso: ominaisuus ja tietoa-ineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria sovellettaessa tulee kertoa, onko tietoja tarkistettu loogisuusehtoja tai laatusääntöjä käyttäen.

Lisäksi tulee huomioida, että erityisesti tietoaineistotasolla olisi hyvä kuvata tarkemmin miltä osin loogisuusehtoja on käytetty. Tiedon käyttäjää kiinnostaa myös yksityiskohdat käytetyistä loogisuusehdoista ja ne olisi hyvä tuoda tietoaineiston kuvauksessa esille.

Mittarin soveltaminen:

Tieto loogisuustarkastettu -mittarin kohdalla on hyvä huomata, että aiempi organisaatio on voinut tehdä loogisuustarkastuksia ennen kuin tietoaineisto on tullut oman organisaation käsittelyyn. Tätä ei kuitenkaan oteta huomioon mittarin arvoa määritettäessä, mutta aiempia tarkasteluja voi silti haluttaessa kuvata aineistokuvauksessa.

- Yksikkötason aineisto: Mittari saa arvon **kyllä**, mikäli huomattavasti suurempi osa tietoaineiston tiedoista on loogisuustarkastettu ainakin yhdellä loogisuuspäätelyllä. Mittari saa arvon ei, mikäli loogisuustarkastuksia ei ole tehty.
- Koostettu aineisto: Mittari saa arvon **kyllä**, mikäli koostettuun aineistoon on tehty loogisuustarkastuksia. Mittari saa arvon ei, mikäli loogisuustarkastuksia ei ole tehty lainkaan.
- Ominaisuus: Mittari saa arvon **kyllä**, mikäli ominaisuuden saamille arvoille on tehty yksi tai useampia loogisuustarkastuksia. Mittari saa arvon ei, mikäli ominaisuuden osalta ei ole tehty loogisuustarkastuksia.

Esimerkit: Ominaisuuden ”toimipaikan osoite” osalta on käytetty loogisuusehtoa, jossa tarkastellaan vastaako toimipaikan sijaintiosoitteen postinumero toimipaikan sijaintikuntaa.

3.4 Laatuksiteeri: Ajantasaisuus

Kuvaus: Ajantasaisuus kuvaa tietoaineiston tietojen täsmällisyyttä ajan suhteen. Ajantasaisuus on sitä parempi mitä lähempänä tiedon viiteajankohta on nykyhetkeä. Viiteajankohta on se ajankohta, jota tieto koskee.

Esimerkit: Tietoaineiston tietoon liittyvä viiteajankohta on kerrottu tiedon yhteydessä. Sen avulla pystytään päättelemään tiedon tuoreus. Viiteajankohta voi olla esimerkiksi vuoden alun ja lopun välinen aika tai

yhden tietyn päivän tilanne. Tiedon tuotannossa tarkastellaan erilaisia tiedon tarkastus- ja muutosajankohtia.

3.4.1 Viiteajankohta

Viiteajankohtaa -mittari kertoo ajankohdan, jolloin tietoaaineiston sisältämä tieto on kerätty eli ajankohdan, jota tieto koskee. Tietoaaineiston käsittely aiheuttaa viivettä ja tietoa on voitu kerätä aiemmin kuin tietoaaineisto on valmis tiedon käyttäjän käsiteltäväksi.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: ajankohta/ei relevantti

Mittarin arvo kertoo, kuinka paljon aikaisempaan ajankohtaan liittyviin tapahtumiin tietoaaineisto perustuu. Mittarin arvosta on mahdollista laskea tietoaaineiston käsittelyn aiheuttama viive tiedon julkaisuun tai käyttöön saamiseen.

Mittarin soveltaminen:

Mittari soveltuu parhaiten tietoaaineistoihin, joilla on selkeästi määritelty ajanhetki, jonka mukaista tilannetta tarkasteltavat tiedot kuvaavat. Vaikka tiedot kohdistuvat tietylle tarkastelujaksolle, tietoja on voitu täydentää myöhemminkin. Kuitenkin tavoitteena on ollut tuottaa tietoa nimenomaan viiteajankohdan tilanteesta.

Esimerkit: Tilaston viiteajankohta, joka voi olla useammankin kuukauden takainen, koska tietojenkeruu ja -käsittely vaatii aikansa ennen julkaisemista.

Tilastokeskuksen tuottama Suomen väkiluku ilmoitetaan vuoden viimeisen päivän tilanteen mukaan.

3.4.2 Luontiajankohta

Luontiajankohta-mittari kertoo päivämäärän, joka on kohdeyksikön tai ominaisuuden luontiajankohta. Tämän mittarin tarkoituksena on myös viestiä siitä, miten pitkältä ajanjaksolta tietoja on saatavilla.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: ajankohta

Mittarin arvo kertoo ominaisuuden tai tietoaineiston luontiajankohdan. Mittari vastaa myös kysymykseen mistä päivämäärästä lähtien ominaisuus on ollut mukana tietoaineistossa tai mistä päivämäärästä alkaen tietoaineistoa on koottu. Kohdeyksikön luontiajankohta ei välttämättä ole sama kuin tietoaineiston luontiajankohta.

Lisäksi tulee huomioida, että luontiajankohta on tieto, joka on yleensä osa metatietoja. Luontiajankohdan lisäksi aineiston kuvauksissa on tärkeää viestiä vertailukelpoisten tietojen ajanjaksot.

Mittarin soveltaminen:

Mittari kertoo mistä ajanhetkestä alkaen tietoja on koottu tarkasteltavaan tietoaineistoon tai ominaisuus on ollut mukana tietoaineistossa. Mikäli kaikki ominaisuudet ovat olleet tietoaineistossa sen luomisesta alkaen, niin kaikkien tietoaineistoon liittyvien luontiajankohdan mittareiden arvo on sama.

Monissa avoimen datan tiedoissa luontiajankohta on sama kuin päivitysajankohta ja tällöin luontiajankohta-tietoa on tulkittu toisin kuin tämän mittarin kohdalla. Päivitetty tietoaineisto nähdään omana kokonaisuutena, joka on luotu päivittämisen yhteydessä. Näissä tilanteissa aikasarjanäkökulmaa pitää tarkastella toisten annettujen tietojen kautta.

Ominaisuuden tai kohdeyksikön tasolla tapahtuva tarkastelu vastaa yleensä paremmin mittarin määritelmää. Metatiedoissa kuvataan usein mistä lähtien tai millä ajanjaksolla ominaisuus sisältyy tietoaineistoon.

Esimerkit: Aineistokatalogissa¹ datasetin kuvailu sisältää ominaisuuksien listauksen ja ominaisuuksien viiteajankohdan alkupäivä vastaa luontiajankohta-mittarin tarkoittamaa ominaisuuden luontiajankohtaa.

¹ Aineistokatalogin pääsivu: <https://aineistokatalogi.fi/catalog>

3.4.3 Tarkistusajankohta

Tarkistusajankohta-mittari kertoo päivämäärän, joka on kohdeyksikön tai ominaisuuden tarkistusajankohta.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: ajankohta/ei relevantti

Mittarin arvo kertoo ominaisuuden arvojen tai tietoaineiston tietojen viimeisimmän tarkistusajankohdan. Mittari vastaa myös kysymykseen, milloin tietoja on viimeksi tarkistettu.

Lisäksi kannattaa huomata, että tarkistusajankohta on tieto, joka on yleensä osa metatietoja.

Mittarin soveltaminen:

Tarkistusajankohdalla tarkoitetaan ajankohtaa, jolloin tietoaineisto tai ominaisuus on päivitetty, jos päivitystarvetta on ollut. Mikäli päivitystarvetta ei ole ollut, niin tiedot ovat pysyneet muuttumattomina tarkistuksesta huolimatta.

- Yksikkötason aineisto: Mittari kertoo sen päivämäärän, jolloin tietoaineistoa on tarkistettu joltain osin.
- Koostettu aineisto: Mittari kertoo sen päivämäärän, jolloin koostettua aineistoa on tarkistettu joltain osin.
- Ominaisuus: Mittari kertoo sen päivämäärän, jolloin ominaisuutta ja sen saamia arvoja on tarkistettu joltain osin.

Esimerkit: Maanmittauslaitoksen tietoihin tehdään säännöllisiä tarkistuksia maan muotojen, tiestön ja rakennusten muutosten varalta. Mikäli rakennus on edelleen pystyssä, eivät tiedot muutu, mutta käytettävissä oleva tietoaineisto on tarkistettu kyseisenä ajankohtana.

3.4.4 Muutosajankohta

Muutosajankohta-mittari kertoo päivämäärän, joka on kohdeyksikön tai ominaisuuden muutosajankohta.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: ajankohta/ ei relevantti

Mittarin arvo kertoo ominaisuuden tai tietoaineiston muutosajankohdan. Mittari vastaa myös kysymykseen, milloin tietoja on viimeksi päivitetty.

Lisäksi kannattaa huomata, että muutosajankohta on tieto, joka on yleensä osa metatietoja. Jatkuvasti tai usein päivittyvien tietoaineistojen tapauksessa mittari ei ole mielekäs, mutta näidenkin tapauksessa muutosajankohdat muuttuneille tiedoille on syytä olla saatavilla tiedon elinkaaren näkökulmasta.

Mittarin soveltaminen:

- Yksikkötason aineisto: Mittari kertoo sen päivämäärän, jolloin jotain tietoaineiston osaa on päivitetty. Muutosten ei tarvitse kohdistua kaikkiin tai suurimpaan osaan aineistoa vaan tarkoitus on kertoa, että jotain osaa tietoaineistosta on muutettu.
- Koostettu aineisto: Mittari kertoo sen päivämäärän, jolloin koostettuun aineistoon on tehty muutoksia. Muutokset voivat olla rakenteeseen tai sisältöön kohdistuvia.
- Ominaisuus: Mittari kertoo sen päivämäärän, jolloin jonkin kohdeyksikön ominaisuuden saama arvo on päivitetty. Merkinnät ovat tärkeitä tietojen myöhemmän jäljitettävyyden kannalta.

3.5 Laatuksiteeri: Kattavuus

Synonyymit: Täydellisyys

Kuvaus: Kattavuus kuvaa tietoaineiston tavoitellun ajallisen ja alueellisen kattavuuden sekä tavoitellut kohdeyksiköt ja ominaisuustiedot. Toisaalta kattavuus kertoo miltä osin tietoaineisto sisältää tavoiteltuja tietoja.

Esimerkit: Tietoaineisto kattaa määritellyn tarkastelualueen yksiköt, esimerkiksi kaikki Suomen yritykset. Alueellinen kattavuus kertoo, ovatko kaikki käsiteltävät alueet mukana, esimerkiksi kaikki Suomen kunnat tai kattaako tietoaineisto Ahvenanmaan. Ylimääräiset kohdeyksiköt eli ylipeitto kertoo, että tietoaineistossa on siihen kuulumattomia yksiköitä. Puuttuvat kohdeyksiköt eli alipeitto kertoo, että tietoaineistosta puuttuu siihen kuuluvia yksiköitä. Vastuskato on alipeittoa. Toisaalta kattavuus kertoo myös sisältääkö tietoaineisto kaikki määritellyt ominaisuustiedot tietoaineistossa oleville kohdeyksiköille, esimerkiksi tietoaineistossa

olevien Suomen kuntien väkiluku- ja pinta-aratiedot tai onko kaikilla yrityksillä tiedossa esimerkiksi osoite- tai liikevaihtotieto.

3.5.1 Ajallinen tavoitekattavuus

Ajallinen tavoitekattavuus -mittari kertoo, että tietoaineiston tavoiteltava ajallinen kattavuus ja tiheys on kuvattu. Ajallisella tavoitekattavuudella tarkoitetaan aikaväliä, jota tietoaineiston on tarkoitus kuvata, sekä tiheyttä, jolla ominaisuuksien arvoja on mitattu.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei; lisätieto/tarkennus voidaan kuvata

Mittarin arvo kertoo siitä, onko tietoaineiston ajallinen tavoitekattavuus kuvattu esimerkiksi aineistokuvauksessa.

Mittarin soveltaminen:

Mikäli tietoaineiston osalta asiakkaan saatavilla on kuvaus siitä, mitä ajanjaksoa tietoaineiston on tarkoitettu kuvaavan, saa mittari arvon **kyllä**. Jos tämä tieto ei ole saatavilla, mittari saa arvon ei.

Esimerkit: FinTerveys 2017 -tutkimuksen perusjoukko on vuonna 2017 Manner-Suomessa asuva vähintään 18 vuotta täyttänyt väestö. Tutkimuksen otokseen osuneita henkilöitä on tutkittu kerran, mutta FinTerveys 2017 -tutkimukselle toteutetaan myöhemmin erillinen seurantatutkimus.

3.5.2 Alueellinen tavoitekattavuus

Alueellinen tavoitekattavuus -mittari kertoo, että tietoaineiston tavoiteltava alueellinen kattavuus ja tiheys on kuvattu. Alueellisella tavoitekattavuudella tarkoitetaan sitä maantieteellistä aluetta, jonka tietoa tietoaineiston on tarkoitettu kuvaavan. Tiheydellä viitataan siihen, miten tarkasti aluetaso on tietoaineistossa kuvattu.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei; lisätieto/tarkennus voidaan kuvata

Mittarin arvo kertoo siitä, onko tietoaineiston alueellinen tavoitekattavuus kuvattu esimerkiksi aineistokuvauksessa.

Mittarin soveltaminen:

Mikäli tietoaineiston osalta asiakkaan saatavilla on kuvaus siitä, mitä aluetta tietoaineiston on tarkoitettu kuvaavan, saa mittari arvon **kyllä**. Jos tämä tieto ei ole saatavilla, mittari saa arvon **ei**.

Esimerkit: FinTerveys 2017 -tutkimuksen alueellinen kattavuus on Manner-Suomi. Tilastokeskuksen Ruututietokanta² sisältää keskeisiä tietoja alueen asukasrakenteesta, koulutuksesta, asumisesta, tuloista ja työpaikoista koko Suomen alueelta joko 250 m x 250 m-, 1 km x 1 km- tai 5 km x 5 km-karttaruuduittain.

3.5.3 Tavoitekohdeyksiköt

Tavoitekohdeyksiköt-mittari kertoo siitä, että tietoaineiston aineistokuvauksessa on selkeästi määritelty ajallisen ja alueellisen tavoitekattavuuden lisäksi muut tietoaineiston rajaukset.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei; lisätieto/tarkennus voidaan kuvata
- Mittarin tausta: ISO 19157

Mittaria sovellettaessa tulee huomioida, että tavoitekohdeyksiköiden kuvauksessa on kirjattu näkyviin olennaiset kohdeyksiköiden rajaukset. Ajallisen ja alueellisen kattavuuden näkökulma on mittaristossa omina mittareinaan.

Mittarin soveltaminen:

Mikäli tietoaineiston osalta asiakkaan saatavilla on kuvaus siitä, mitä kohdeyksiköitä tietoaineiston pitäisi sisältää, saa mittari arvon kyllä. Jos tämä tieto ei ole saatavilla, mittari saa arvon ei.

Esimerkit: Tietoaineisto kattaa tiettyjen toimialojen yritykset. Tietoaineisto sisältää tiedot rakennuksista, joiden koko on yli 10m².

FinTerveys 2017 -tutkimuksen perusjoukko on Manner-Suomessa asuva 18 vuotta vuonna 2017 täyttänyt väestö.

² Ruututietokannan kuvaus: <https://www.stat.fi/tup/ruututietokanta/index.html>

3.5.4 Tavoiteominaisuuksissa puutteita

Tavoiteominaisuuksissa puutteita-mittari kertoo, mikäli tietoaineiston ilmiön kannalta keskeisiä ominaisuuksia puuttuu arvioitavasta tietoaineistosta. Puutteellisuutta on tärkeä kuvata tarkemmin esimerkiksi aineistokuvauksessa.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria sovellettaessa tulee huomioida, että tietoaineiston laadun arvioinnin kannalta on tärkeä tunnistaa tietoaineiston kuvaamat ilmiöt ja näiden ilmiöiden kuvaamiseen tai mittaamiseen tarvittavat ominaisuudet. Mikäli tietoaineistosta puuttuu joku tietty tarkastelun näkökulma eli joitakin ominaisuuksia, on tämä kuvattava selkeästi tiedon käyttäjälle.

Mittarin soveltaminen:

Mikäli tietoaineiston kuvaaman ilmiön keskeisimmät ominaisuudet on tunnistettu ja ne ovat mukana tietoaineistossa, saa mittari arvon ei. Tavoiteominaisuuksissa ei tällöin ole tunnistettu puutteita. Jos tiedetään, että keskeisiä ominaisuuksia puuttuu kokonaan tai ominaisuuksien arvoissa on huomattavaa puuttuvuutta, mittari saa arvon **kyllä**.

Esimerkit: Kuvitteellinen esimerkki voisi olla tilanne, että tietoaineiston sisältö kuvaa henkilön tuloja, mutta ei silti sisällä etuus-tietoja. Kattavaa tulotietoa etsivälle kyseisessä aineistossa on puutteita tavoiteominaisuuksissa. Tällaisen tietoaineiston kohdalla on luonnollisesti mahdollista rajata tietoaineiston kuvausta tarkemmin esimerkiksi että tietoaineisto kattaa palkka- ja pääomatulot.

3.5.5 Puuttuvat kohdeyksiköt

Puuttuvat kohdeyksiköt -mittari kuvaa tietoaineiston alipeittoa eli tietoaineiston tavoitekohdejoukosta puuttuvien kohdeyksiköiden prosentuaalista osuutta.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: prosenttia

Mittarin tausta: ISO 19157

Mittaria sovellettaessa tulee huomioida, että arvo ilmoitetaan tietoa-aineistosta (tavoiteperusjoukosta) puuttuvien kohdeyksiköiden lukumäärän suhteena tietoaaineistoon kuuluvien kohdeyksiköiden lukumäärään.

Lisäksi tulee huomioida, että valmiiksi koostetun tiedon, esim. tilasto, osalta puuttuvuus on yleensä korjattu tilastollisin menetelmin. Tällöin voidaan olettaa, ettei tarkasteltavassa tietoaaineistossa ole puuttuvia kohdeyksiköitä. Puuttuvia kohdeyksiköitä voi kuitenkin halutessaan tarkastella raportointitasolla esimerkiksi toimiala-tasolla yritystason sijaan.

Mikäli täsmällisen arvon laskeminen on haastavaa, voidaan antaa kokemukseen perustuva arvio prosenttiarvoksi. Puuttuvuutta olisi tärkeää kuvata tietoaaineiston kuvaukseen, mikäli tietoaaineiston puuttuvista kohdeyksiköistä on riittävä käsitys.

Mittarin soveltaminen:

Kuvataan joko johonkin toiseen aineistoon vertaamalla todettu tai kokemuksen perusteella arvioitu tietoaaineiston puuttuvien kohdeyksiköiden osuus. Puuttuvuuden syiden kuvaaminen on hyvä sisällyttää aineistokuvaukseen.

Esimerkit: Harmaasta taloudesta voi seurata puuttuvia kohdeyksiköitä kansantaloutta kuvaavien tietojen yhteydessä. Tavoitteena on kuvata kaikkea taloudellista toimintaa, mutta osa siitä jää kirjaamatta. Harmaa talous voi tarkoittaa kokonaan puuttuvia kohdeyksiköitä tai kohdeyksiköiden puuttuvia ominaisuustietoja tietoaaineistossa. Harmaan talouden osuutta rakennusosalalla on arvioitu käytettyjen rakennustarvikkeiden määrää hyödyntäen, jolloin on saatu arvioita tarvittavasta työvoimasta.

Puuttuvia kohdeyksiköitä voi seurata otostutkimuksissa kadon seurauksena, jos jokin kohdeyksiköiden joukko jää kokonaan vastaamatta ja siten tuloksia ei voi laajentaa koskemaan koko tavoiteperusjoukkoa edes tilastollisin menetelmin käsittelemällä. Esimerkiksi, jos otostutkimusta tehtäisiin vain verkkotiedonkeruulla ja tavoiteperusjoukkoon kuuluisi myös hyvin iäkkäitä ihmisiä, voisi olla ettei iäkkäimmät vastaisi ollenkaan.

3.5.6 Ylimääräiset kohdeyksiköt

Ylimääräiset kohdeyksiköt -mittari kuvaa tietoaaineiston ylipeittoa eli tietoaaineistossa mukana olevien, mutta tietoaaineiston tavoiteperusjoukkoon kuulumattomien ja siten ylimääräisten kohdeyksiköiden prosentuaalista osuutta.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: prosenttia
- Mittarin tausta: ISO 19157

Mittaria sovellettaessa tulee huomioida, että arvo ilmoitetaan ylimääräisten kohdeyksiköiden lukumäärän suhteena tietoaaineiston tavoiteltavien kohdeyksiköiden lukumäärään.

Lisäksi tulee huomioida, että valmiiksi koostetun, esim. tilaston, tiedon osalta ylipeitto on yleensä pyritty jo korjaamaan. Mikäli täsmällisen arvon laskeminen on haastavaa, voidaan antaa kokemukseen perustuva arvio prosenttiarvoksi. Mikäli tietoaaineisto sisältää tietyntyyppeisiä ylimääräisiä kohdeyksiköitä ja tätä ylimääräisten kohdeyksiköiden joukkoa on mahdollista kuvata, kuvaus olisi tärkeää sisällyttää tietoaaineiston kuvaukseen.

Mittarin soveltaminen:

Arvioidaan tarkasteltavan tietoaaineiston ylimääräisten kohdeyksiköiden määrää vertaamalla johonkin toiseen tietoaaineistoon tai käyttäen otoksia näistä aineistoista. Vaihtoehtoisesti arvion voi tehdä kokemukseen perustuen. Mittarin arvoksi ilmoitetaan ylimääräisten kohdeyksiköiden prosentuaalinen osuus laskemalla ylimääräisten kohdeyksiköiden määrä suhteessa kaikkien kohdeyksiköiden määrään.

Ylimääräisten kohdeyksiköiden esiintymisen syiden kuvaaminen on hyvä sisällyttää aineistokuvaukseen.

Esimerkit: Tietoaaineistossa on samoja kohdeyksiköitä useamman kerran, koska tietoja on tuotu tietoaaineistoon useasta lähteestä ja kaikissa lähdeaineistoissa ei ole sama kohdeyksikön tunniste käytössä.

Otostutkimuksen kohdeyksiköiden otannassa on mukana maasta pois muuttaneita henkilöitä, koska osoitetiedot eivät ole päivittyneet otoksen pohjana olevaan tietokantaan.

3.5.7 Puutteelliset kohdeyksiköt

Puutteelliset kohdeyksiköt -mittari kuvaa tietoaineiston puuttuvia ominaisuustietoja omaavien kohdeyksiköiden osuuden kaikista kohdeyksiköistä.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: prosenttia

Mittaria sovellettaessa tulee huomioida, että mittarin arvo ilmoitetaan yhdenkin puuttuvan ominaisuustiedon omaavien kohdeyksiköiden lukumäärän suhteena kaikkien kohdeyksiköiden lukumäärään. Rakenteellista puuttuvuutta, eli tilannetta, jolloin jotain ominaisuuden arvoa ei kuulukaan tulla tietylle kohdeyksikölle, ei oteta tämän mittarin kohdalla huomioon. Tarkastelussa keskitytään puuttuvuuteen, jossa puuttuva arvo on kyseiselle kohdeyksikölle relevantti. Esimerkiksi lapsille ei kuulukaan olla ammattiluokituksen arvoa.

Lisäksi tulee huomioida, että tämä mittari ei välttämättä tuota mielekkäitä arvoja koostetun tiedon tilanteessa tietoaineistotasolla arvioitaessa. Vastaavasti laajojen yksikkötason tietoaineistojen tai pitkien aikasarjojen tapauksessa puutteellisten kohdeyksiköiden osuus voi kasvaa erittäin suureksi, koska yksittäisiä tietoja saattaa puuttua melkein jokaiselta kohdeyksiköltä. Ominaisuuksittaisen tarkastelun kohdalla on myös hyvä kuvata rakenteellista puuttuvuutta, jos se kohdistuu selvästi määriteltävään kohdejoukkoon.

Mittarin soveltaminen:

Mikäli kohdeyksikön yhdessäkin ominaisuudessa on puuttuvan arvon koodi (ei rakenteellista puuttuvuutta), kyseinen kohdeyksikkö on puutteellinen kohdeyksikkö.

Yksikkötason aineisto: Mittari lasketaan sellaisten kohdeyksiköiden, joilla on yksikin puuttuvan tiedon merkintä, osuutena koko yksikkötason aineiston kohdeyksiköistä.

Ominaisuus: Mittari lasketaan ominaisuuden saamien puuttuvan tiedon merkintöjen osuutena suhteessa koko aineistossa olevien kohdeyksiköiden lukumäärään.

Esimerkit: Puutteellisia kohdeyksiköitä tarkasteltiin yritysten tilastotiedoista tutkimustarkoituksiin kootusta valmisaineistosta³, joka sisältää yli kymmenen vuoden ja liki 70 ominaisuuden tietoja yrityksistä. Tietoaineiston pitkän tarkastelujakson takia tiedoista todettiin puutteita käytännössä jokaisen yrityksen kohdalla. Vaikka suuren aineistokokonaisuuden puutteellisten kohdeyksiköiden osuudeksi todettiin 100 %, ei aineisto ole hyödytön vaan päinvastoin sisältää suuren määrän pitkiä ja täydellisiä tarkastelujaksoja, kunhan tarkasteltavaa osajoukkoa eli yrityksiä ja näiden ominaisuuksia rajattiin tarkemmin.

3.5.8 Puutteelliset ominaisuustiedot

Puutteelliset ominaisuustiedot -mittari kuvaa kuinka suuri osuus kohdeyksiköistä saa puuttuvan arvon ominaisuuden suhteen tarkasteltuna. Mittari kuvaa sitä, kuinka kattavasti tietoaineistossa on arvoja tietyille ominaisuudelle.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: prosenttia

Mittaria sovellettaessa tulee huomioida, että mittarin arvo ilmaistaan puuttuvan ominaisuustiedon omaavien kohdeyksiköiden lukumäärän suhteen kaikkien kohdeyksiköiden lukumäärään ominaisuuksittain tarkasteltuna. Tarkastelussa keskitytään katsomaan aina yhtä ominaisuutta kerrallaan ja tuottamaan tietoa siitä kuinka laajasti kyseisen ominaisuuden arvoja sisältyy tietoaineistoon. Rakenteellista puuttuvuutta ei oteta tämän mittarin kohdalla huomioon, vaan tarkastellaan puuttuvuutta silloin, kun jokin arvo on kyseiselle kohdeyksikölle relevantti. Rakenteellisella puuttuvuudella tarkoitetaan tilannetta, jossa kohdeyksikkö ei missään tapauksessa voisi saada arvoa jollekin ominaisuudelle.

Mittarin soveltaminen:

- Yksikkötason aineisto: Kerrotaan suurin puuttuvien tietojen osuus, joka aineiston ominaisuuksista on havaittu. Mikäli puuttuvuuden

3 FIRM_ENTER Yritysrekisterin tilastotiedostot: yritykset (1982-) 2012 (YA221) -aineistokuvaus
https://taika.stat.fi/fi/aineistokuvaus.html#!?dataid=YA221_2012_jua_yrekyritys_003.xml

osuus vaihtelee huomattavasti aineistossa, on hyvä kuvata tarkemmin puuttuvuutta esimerkiksi aineistokuvauksessa.

- Koostettu aineisto: Kuvataan koostetussa aineistossa käytettävien ominaisuuksien osalta eli tarkastelussa ei palata tarkimmalle koodistotasolle, jos sitä ei koostetussa aineistossa raportoida.

Esimerkit: Kyselytutkimuksissa erityisesti vastaajan kannalta arka-luonteisiin kysymyksiin vastaamisessa on muita ominaisuuksia enemmän puuttuvuutta. Voi olla, että tietoaaineistossa on vastaajan ikä kaikille vastaajille, mutta tupakointi-tieto vain 80 % vastaajista.

4 Laatukriteeriryhmä: Miten tieto on kuvattu?

Toisen laatukriteeriryhmän kriteerit korostavat tiedon elinkaaren ja tiedon kuvaamisen tärkeyttä. Hyvin kuvatun tiedon käyttäminen on helpompaa ja vähentää virhetulkintoja.

Kuva 3. Laatukriteerit kysymyksen ”Miten tieto on kuvattu?” alla.



4.1 Laatukriteeri: Alkuperäisyys

Synonyymit: **jäljitettävyys, kiistämättömyys**

Kuvaus: Alkuperäisyys kertoo siitä, että tietoaaineistoon ja sen tietoihin tehdyt muutokset voidaan jäljittää. Tiedon alkuperä tunnetaan.

Esimerkit: Tiedon alkuperä ja muutostapahtumat on kuvattu, muutosten aikaleimat ovat saatavilla. Tieto voidaan todistaa kiistämättömäksi, tietoaaineiston tiedot voidaan todentaa uudelleen.

4.1.1 Tietolähde

Tietolähde-mittari kertoo kuinka suurella osalla kohdeyksiköitä tai ominaisuustietoja lähdetiedot ovat saatavilla.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: prosenttia

Mittaria sovellettaessa tulee huomioida, että mittarin arvo ilmoitetaan saatavilla olevien lähdetietojen suhteena kaikkiin tietoihin. Tarkastelussa keskitytään tarkastelemaan tietolähdettä, josta tiedot on suoraan saatu laatua arvioivalle taholle. Mahdolliset aiemmat tiedon vaiheet on tällöin mahdollista selvittää tiedon aiemman raporttoijan laaturaporttien kautta.

Lisäksi tulee huomioida, että tiedon käyttäjä kiinnostaa tietää tietolähteet ja miltä osin lähdetietoja puuttuu. Tietolähteen osalta on olennaisinta kertoa edellinen taho, joka tiedon on toimittanut, jolloin haluttaessa laaturarkastelussa voidaan ottaa huomioon ketjumaisesti tiedon laatuun liittyviä näkökulmia eri toimijoiden osalta. Jos tietolähteen vaillinaisuus koskee esimerkiksi tiettyä kohdeyksikköjoukkoa, olisi tietoaineiston kuvaustiedoissa hyvä kuvata puutteita. Tietolähteen lisäksi voidaan kuvata tietoihin tehtyjä muokkauksia, joiden avulla tiedon käyttäjä saa tietolähteen lisäksi lisätietoa tiedon prosessointihistoriasta.

Mittarin soveltaminen:

Lähtökohtaisesti tarkastellaan sitä tietolähdettä, joka on organisaation ulkopuolella, esimerkiksi toinen organisaatio tai suora tiedonkeruu. Mikäli tiedon alkuperäinen tietolähde on jokin oman organisaation osasto/yksikkö tai tiedon sisäisellä siirtymisellä on erityistä merkitystä, otetaan tämä tarkastelussa huomioon.

Tietolähteet ovat yleensä hyvin tunnistettu vähintään yleisellä tasolla, joten oletettavasti tämän mittarin arvot ovat yleensä lähellä 100 %. Puutteita saattaa olla vanhemmissa tiedoissa tai jos puuttuvia tietoja on täydennetty menetelmällisesti.

Yksittäisen kohdeyksikön tietolähdettä ei tarvitse lähteä tarkastelemaan vaan yleistason tieto siitä, mitä lähteitä on käytetty riittää mittarin tarkastelussa. Yksittäisen arvon lähde voi olla tärkeä silloin, jos tietojen lähde vaikuttaa esimerkiksi tietojen oikeellisuuteen. Lähdetieto on erityisen tärkeä silloin, jos tietoja on lisätty menetelmällisesti.

Esimerkit: Tietolähde voi olla tuntematon, jos tietoja on tuotettu menetelmällisesti. Kyseiset arvot eivät välttämättä perustu mihinkään yksilöitävissä olevaan tietolähteeseen vaan voivat olla mallinnuksen tulos. Tällaiset tiedot yleensä merkitään tietoaineistoon omin merkinnöin, mutta tietojen lähde ei ole rinnastettavissa suoraan tiedonkeruuseen tiedonantajalta tai esimerkiksi rekisteritietoon.

4.1.2 Tiedon elinkaari

Tiedon elinkaari -mittari kertoo, onko tietoaineiston elinkaari määritelty ja kuvaus siitä saatavilla.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittarin arvo kertoo, onko tiedon elinkaari määritelty ja kuvattu. Tiedon elinkaari kuvaa prosessin ja tiedon muutokset tiedon luomisesta aina tiedon arkistointiin tai poistamiseen ja sisältää tiedon lähteiden tai tuottamisen muutokset, esimerkiksi jonkin laskentatavan tai määritelmän muutokset.

Lisäksi tulee huomioida, että tiedon syntyyn ja tuottajaan liittyvät asiat ovat olennaisia tietoja tiedon käytettävyyden tarkastelussa.

Mittarin soveltaminen:

Tiedon elinkaaren voi yksinkertaistettua ajatella sisältävän seuraavat vaiheet:

- tiedon hankinta
- tiedon varastointi
- tiedon hyödyntäminen
- tiedon arkistointi tai poistaminen.

Tiedon elinkaaren⁴ määrittelemiseksi tarkastelet seuraavia kysymyksiä: Miten tarkasteltavaan tietoon liittyvät metatiedot eli tietoaineiston ja ominaisuuksien kuvailutiedot on tallennettu? Onko muutoshistoria dokumentoitu ja tallennettu? Onko tämä tieto asiakkaiden saatavilla?

⁴ Tiedon elinkaaren määrittelyn kuvauksia Kuntaliiton sivulla: <https://www.kuntaliitto.fi/kehittaminen-ja-digitalisaatio/tiedon-elinkaari>

Mittari saa arvon **kyllä**, mikäli tarkasteltavan aineiston elinkaari on määritelty ja kuvaus on saatavilla. Mittari saa arvon **ei**, mikäli elinkaarta ei ole määritelty tai määrittely ei ole tiedon käyttäjän saatavilla. Oletettavasti elinkaaren määrittely on tehty asianmukaisesti, mutta tiedot eivät välttämättä ole tiedon käyttäjän saatavilla.

Esimerkit: Tiedon elinkaarta kuvataan esimerkiksi otostutkimuksiin perustuvien tilastojen tiedonkeruupäätöksissä, joita voi tarkastella tarkemmin Tilastokeskuksen verkkosivujen kautta.

4.1.3 Muutosten hallinta

Muutosten hallinta -mittari kertoo, että ominaisuustiedoissa tapahtuvia rakenteellisia tai lähdetietojen muutoksia seurataan.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: kyllä/ei

Mittarin arvo kertoo siitä, tallennetaanko tieto ominaisuustietojen muutoksista muutoksia tehtäessä siten, että muutoksia on mahdollista tarkastella myöhemmin.

Lisäksi tulee huomioida, että on hyvä tuoda esille, miten muutoshistoriaa on mahdollista päästä käsittelemään. Muutoksia on hyvä kuvata yleisellä tasolla esimerkiksi metatietojen kautta.

Mittarin soveltaminen:

Tarkastelussa pysytään ylätasolla. Seurataanko tietoaaineiston kohdalla säädösten, standardien, käytänteiden ja yleisten ohjeistuksien muutoksia, jotka vaikuttavat ominaisuuksien tasolla? Entä seurataanko yleisesti aihepiiriin kuuluvia ilmiöitä ja lähdetiedoissa tapahtuvia muutoksia?

Mittari saa arvon **kyllä**, mikäli aineiston ominaisuuteen kohdistuvia muutoksia seurataan ja tiedot muutoksista on tallennettu myöhempää käyttöä varten. Mittari saa arvon **ei**, mikäli muutoksia ei seurata lainkaan.

Esimerkit: Tietoaaineistossa on otettu käyttöön luokituksen uusi versio tai jonkin ominaisuuden kuvausaluetta (esim. alueellista kattavuutta) on laajennettu tai supistettu. Näistä muutoksista pidetään lokia tai tieto on löydettävissä aineistokuvauksesta tai ominaisuuden metatietojen kautta.

4.2 Laatuksiteeri: Metatietojen ymmärrettävyys

Synonyymit: **ymmärrettävyys, tulkittavuus, käsitettävyys**

Kuvaus: Metatietojen ymmärrettävyys kuvaa sitä, miten kattavasti tietoaineistolla on sellaisia metatietoja, jotka auttavat tiedon ymmärtämisessä sitä käytettäessä.

Esimerkit: Tietoaineisto ja sen ominaisuustiedot on kuvattu metatietokuvauksessa riittävällä tasolla siten, että tietosisällön ja merkityksen ymmärtäminen helpottuu. Ominaisuustiedoissa käytetyt koodistot on kirjattu ja ne ovat datan kanssa yhteneväiset. Koodistojen kuvaukset ovat saatavilla esimerkiksi linkkien kautta. Olennaiset käsitteet on kuvattu ja linkit tarpeellisiin sanastoihin on liitetty metatietokuvauksiin.

4.2.1 Aineistokuvaus

Aineistokuvaus-mittari kertoo, onko tietoaineiston aineistokuvaus saatavilla ja mitä kieliversioita siitä on saatavilla.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kieliversiot

Mittaria sovellettaessa tulee huomioida, että arvo ilmaistaan kertomalla ne kielet, joilla aineistokuvaus on saatavilla.

Lisäksi tulee huomioida, että kieliversioiden voi mainita vaikkei kaikki kieliversiot olisikaan yhtä kattavia. Aineistokuvauksen osalta myös määritelmät ja yhtenäinen termistö ovat tärkeitä – siten käyttäjä voi olla varma, että tietoaineistot ovat vertailukelpoisia.

Mittarin soveltaminen:

Mikäli tietoaineistolle on tehty aineistokuvaus ja tämä on asiakkaan saatavilla, kerrotaan ne **kielet**, joilla aineistokuvaus on saatavilla.

Esimerkit: Monien tietoaineistojen aineistokuvaus on saatavilla Aineistokatalogin⁵ kautta. Aineistokatalogissa tietoaineistosta kuvataan

⁵ Aineistokatalogi: <https://aineistokatalogi.fi/catalog>

perusjoukko, viiteajankohta ja asiasanat sekä kuvataan sanallisesti aineiston sisältöä. Lisätietoja on mahdollista liittää linkkien kautta.

Avoimen datan aineistokuvauksia on listattu avoindata.fi⁶ -sivustolle.

4.2.2 Käsitteiden määritelmät

Käsitteiden määritelmät -mittari kertoo, onko tietoaaineiston käsitteet määritelty selkeästi ja saatavilla sekä mitä kieliversioita niistä on saatavilla. Käsitteellä tarkoitetaan niitä tietoaaineistoon liittyviä asioita, joita ei voida välttämättä suoraan mitata.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: kieliversiot

Mittaria sovellettaessa tulee huomioida, että arvo ilmaistaan kertomalla ne kielet, joilla käsitteiden kuvaukset ovat saatavilla. Käsitteiden määrittelemisen selkeästi on tärkeä osa tiedon kuvausta, jotta voidaan varmistua siitä, että samoilla sanoilla ymmärretään sama asia.

Lisäksi tulee huomioida, että kieliversion voi mainita vaikkei kaikki kieliversiot olisikaan yhtä kattavia.

Mittarin soveltaminen:

Mikäli tietoaaineiston keskeiset käsitteet on pääosin määritelty ja nämä määritelmät ovat asiakkaan saatavilla, kerrotaan ne **kielet**, joilla käsitteiden määritelmät ovat saatavilla.

Esimerkit: Tietoaaineistossa, jossa käsitellään hyvinvointia, kuvataan tietoaaineiston määritelmä hyvinvoinnille. Eli kuvataan ne näkökulmat, joiden kautta hyvinvointia mitataan käsiteltävässä tietoaaineistossa.

Henkilön tulot -käsite voidaan ymmärtää hyvin monella tapaa. Tietoaaineistoista riippuen tarkastelu voi keskittyä brutto- tai nettotuloihin. Lisäksi tarkasteluun voidaan rajata pelkät palkkatulot tai niiden lisäksi etuudet, pääomatulot tai esimerkiksi palkkiot.

⁶ Avoindata.fi: <https://www.avoindata.fi/fi>

Tilastokeskuksella on keskitetty julkinen käsitetietokanta⁷, josta Tilastokeskuksen käyttämiä käsitteitä voi tarkastella.

4.2.3 Ominaisuuksien tietokuvaukset

Ominaisuuksien tietokuvaukset -mittari kertoo, onko tietoaaineiston ominaisuuksien kuvaukset ja koodistot saatavilla ja mitä kieliversioita niistä on saatavilla. Ominaisuuksilla tarkoitetaan tässä yhteydessä tietoaaineiston mitattavia ja arvoja saavia ominaisuuksia.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: kieliversiot

Mittaria sovellettaessa tulee huomioida, että arvo ilmaistaan kertomalla ne kielet, joilla kuvaus on saatavilla.

Lisäksi tulee huomioida, että kieliversioiden voi mainita vaikkei kaikki kieliversiot olisikaan yhtä kattavia. Kuvauksien osalta myös määritelmät ja yhtenäinen termistö ovat tärkeitä – siten käyttäjä voi olla varma, että tietoaaineistot ovat vertailukelpoisia.

Mittarin soveltaminen:

Mikäli tietoaaineiston ominaisuuden tietokuvaukset on pääosin kuvattu ja nämä kuvaukset ovat asiakkaan saatavilla, kerrotaan ne kielet, joilla tietokuvaukset on saatavilla.

Esimerkit: Monien tietoaaineistojen ominaisuuksien tietokuvaukset ovat koottuna Aineistokatalogiin. Aineistokatalogista löydät tiiviin aineistokuvauksen ja lisäksi tietoaaineistoon sisältyvien ominaisuuksien tietokuvauksia. Kieliversioiden määrä riippuu tietoaaineistosta.

4.2.4 Ymmärrettävyyden asiakaspalaute

Ymmärrettävyyden asiakaspalaute -mittari kertoo, että tietoaaineiston ymmärrettävyydestä on mahdollisuus antaa palautetta olemassa olevan palauteväylän kautta tai kohdennetulla asiakaskyselyllä. Saatua palautetta

⁷ Käsitetietokanta <https://stat.fi/meta/kas/index.html>

seurataan ja palautteeseen myös reagoidaan.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria sovellettaessa tulee huomioida, että mittarin on tarkoitus kuvata sitä, onko tiedon käyttäjän mahdollista antaa palautetta ja saada vastauksia antamaansa palautteeseen. Lisäksi tärkeä näkökulma on, että saatua palautetta voidaan hyödyntää tietoaineiston kehittämisessä.

Lisäksi tulee huomioida, että palautteen yhteenveto tai palautteen aiheuttamat muutokset tietoaineistossa ovat hyödyllistä lisätietoa tiedon käyttäjille.

Mittarin soveltaminen:

Mikäli organisaatiossa on palauteväylä, josta on kerrottu aineiston yhteydessä tai muuten hyvin selkeästi ja palautetta hyödynnetään ja siihen reagoidaan, saa mittari arvon **kyllä**. Jos palauteväylää ei ole käytettävissä mittari saa arvon ei.

Esimerkit: Tietoaineistoon liittyvää palautetta kerätään sähköposti-osoitteen tai palautelomakkeen kautta. Saadusta palautteesta kootaan yhteenveto ja keskeiset kehittämis ehdotukset kootaan tietoaineiston käsittelijöiden hyödynnettäväksi.

4.3 Laatu kriteeri: Suositusten mukaisuus

Synonyymit: **yhteensopivuus, semanttinen yhdenmukaisuus, yhdenmukaisuus**

Kuvaus: Suositusten mukaisuus kertoo siitä, että tietoaineisto ja sen ominaisuustiedot noudattavat tunnettuja standardeja, käytäntöjä ja säädöksiä ja ne on kerrottu tietoaineiston yhteydessä.

Esimerkit: Kansallista yhdenmukaisuutta tukee muun muassa yhtenäisten kansallisten sanastojen ja koodistojen käyttö tietoaineistojen tietoja suunniteltaessa. Kansainvälistä yhdenmukaisuutta tukevat esimerkiksi EU:n vahvistamat standardiluokitukset ja ISO-kielikoodisto.

4.3.1 Noudatettavat säädökset ja standardit

Noudatettavat säädökset ja standardit -mittari kertoo ovatko tietoa-aineiston noudattamat säädökset, standardit, hyvät käytänteet ja suositukset listattuna tietoaaineiston kuvauksissa.

- Arviointitaso: ominaisuus ja tietoaaineisto
- Arvon esitysmuoto: kyllä/osittain/ei
- Mittarin tausta: INSPIRE/FAIR⁸ -periaatteet

Mittaria sovellettaessa tulee huomioida, että säädöksiä ja standardeja on monen tasoisia: yleisiä ja toimialakohtaisia sekä kansallisia ja kansainvälisiä. Näistä vähintään keskeisimmät tulisi luetella esimerkiksi aineistokuvauksen yhteydessä. Myös ominaisuuksien osalta yleisten koodistojen käyttö tulee ottaa huomioon. Tilanteessa, jossa yleisestä normistosta poiketaan, olisi tärkeä tuoda esiin tämä poikkeama.

Mittarin soveltaminen:

Tietoaaineiston kuvaaman ilmiön keskeisimmät säädökset, standardit, suositukset, ohjeet ja käytännöt sekä ominaisuuksien koodistot tulee olla tunnistettu. Mittari mittaa sitä, miltä osin näitä noudatetaan. Mikäli tietoaaineiston sisältö noudattaa näitä säädöksiä ja standardeja pääosin, saa mittari arvon **kyllä**. Mikäli standardeista noudatetaan vain osaa, saa mittari arvon osittain. Jos standardeja ei noudateta, mittari saa arvon **ei**.

Tiedot säädöksistä ja standardeista on tärkeä osa aineiston ja ominaisuuksien kuvaustietoja. Joiltakin osin käytettävissä voi olla vaihtoehtoisia standardeja tai koodistoja ja näiden osalta olisi tärkeää kuvata noudatettava vaihtoehto metatiedoissa. Yleensäkin poikkeamat standardeista on tärkeää kuvata aineistokuvauksessa.

Esimerkit: Ominaisuus noudattaa kansainvälistä koodistostandardia, mutta kyseistä koodistoa on täydennetty organisaatiokohtaisesti lisäluokilla. Lisäluokat on määritelty koodistostandardista erilliseksi siten, että ne täsmentävät standardia mutteivat ole ristiriidassa sen kanssa.

8 INSPIRE - the infrastructure for spatial information in Europe: <https://inspire.ec.europa.eu/>, FAIR-periaatteet: <https://www.fairdata.fi/tietoa-fairdatasta/fair-periaatteet/>

5 Laatuksiteeriryhmä: Miten tietoa voi käyttää?

Kolmas laatuksiteeriryhmä sisältää laatuksiteereitä, jotka kertovat tiedon käyttäjälle miten, milloin ja missä muodossa tieto on käytettävissä.

Kuva 4. Laatuksiteerit kysymyksen ”Miten tietoa voi käyttää?” alla



5.1 Laatuksiteeri: Koneluettavuus

Kuvaus: Koneluettavuus kuvaa, onko tietoaaineisto rakenteistettu siten, että sitä voidaan käsitellä koneellisesti ja käsittely on mahdollista eri tietojärjestelmissä.

Esimerkit: Tietoaaineisto on rakenteisessa formaatissa, esim. csv, json, xml. Tietoaaineiston rakenne on kuvattu, esim. skeema tai muu vastaava tietorakennemäärittäminen.

5.1.1 Tietoaaineiston tietomalli

Tietoaaineiston tietomalli -mittari kertoo, onko tietoaaineisto kuvattu rakenteisesti.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: kyllä/ei

Mittarin arvo kertoo, onko aineiston rakenne kuvattu tietomallin, skeeman tai vastaavan standardin mukaisesti. Mikäli tietoaaineisto on kuvattu tietomallin mukaisesti, se on koneluettava.

Lisäksi tietoaineiston tietomallin osalta olisi hyvä kertoa minkä tietomallin tai standardin mukaan tietoaineisto on kuvattu.

Mittarin soveltaminen:

Tietomalli kuvaa, mitä tietokenttiä tietoaineistossa on ja miten ne liittyvät toisiinsa. Hyvä esimerkki standardien mukaisesta tietomallista on xml-skeema, jonka käyttö parantaa eri tietojärjestelmien yhteentoimivuutta.

Mikäli tietoaineiston rakenne on kuvattu esimerkiksi tietomallin mukaisesti, saa mittari arvon **kyllä**. Jos tietoaineiston rakennetta ei ole kuvattu, mittari saa arvon **ei**. Tietomalli, skeema tai standardi, jota kuvauksessa on käytetty, on tärkeä kertoa kuvauksen yhteydessä tai aineistokuvauksessa.

Esimerkit: Kansallisen maastotietokannan rakennustietojen rakenne on kuvattu JHS 210-standardissa. Tietoaineiston rakenne on kuvattu skeemalla (esim. xml, json). Lisää esimerkkejä erilaisista käytössä olevista tietomalleista löydät tietomallit-työkalusta⁹.

5.1.2 Kohdeyksikön pysyvä tunniste

Kohdeyksikön pysyvä tunniste -mittari kertoo, että tietoaineiston kohdeyksiköillä on vähintään tietoaineistokohtainen pysyvä tunniste, jonka perusteella kohdeyksiköt voidaan erottaa toisistaan.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria sovellettaessa tulee huomioida, että tarkastelussa riittää, että pysyvä tunniste on vähintään tietoaineistokohtainen. Luonnollisesti yhtenäinen ja pysyvä tunniste kansallisesti tai jopa kansainvälisesti olisi laajemman hyödynnettävyyden kannalta eduksi. Tunnisteen pysyvyyden lisäksi myös tunnisteiden ainutlaatuisuuteen tulisi kiinnittää huomiota eli pyrkiä siihen, ettei tietoaineistossa olisi samaa kohdeyksikköä useampaan kertaan.

⁹ Tietomallit-työkalu: <https://tietomallit.suomi.fi/>

Mittarin soveltaminen:

Mikäli tietoaaineisto kohdeyksiköillä on vähintään tietoaaineiston sisäisesti pysyvä tunniste, saa mittari arvon **kyllä**. Jos pysyvää tunnistetta ei ole, mittari saa arvon **ei**.

Esimerkit: Yritys- ja yhteisötunnus¹⁰ eli Y-tunnus on PRH:n tai Verohallinnon yritykselle ja yhteisölle antama tunnus. Siinä on seitsemän numeroa, väliviiva ja tarkistusmerkki, eli se on muotoa 1234567-8. INSPIRE-soveltamissäädökset edellyttävät paikkatietokohteen tunnuksen julkaisemista HTTP URI-muodossa. (JHS 193)

5.1.3 Koneluettavuuden asiakaspalautte

Koneluettavuuden asiakaspalautte -mittari kertoo, että tietoaaineiston koneluettavuudesta on mahdollisuus antaa palautetta olemassa olevan palauteväylän kautta tai kohdennetulla asiakaskyselyllä. Saatua palautetta seurataan ja palautteeseen myös reagoidaan.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria sovellettaessa tulee huomioida, että mittarin on tarkoitus kuvata sitä, onko tiedon käyttäjän mahdollista antaa palautetta ja saada vastauksia antamaansa palautteeseen. Tämän lisäksi tärkeä näkökulma on, että saatua palautetta voidaan hyödyntää tietoaaineiston koneluettavuuden parantamiseksi. Lisäksi tulee huomioida, että palautteen yhteenvedo tai palautteen aiheuttamat muutokset ovat hyödyllistä lisätietoa tiedon käyttäjille.

Mittarin soveltaminen:

Mikäli organisaatiossa on palauteväylä, josta on kerrottu aineiston yhteydessä tai muuten hyvin selkeästi ja palautetta hyödynnetään ja siihen reagoidaan, saa mittari arvon **kyllä**. Jos palauteväylää ei ole käytettävissä mittari saa arvon **ei**.

Esimerkit: Tietoaaineiston koneluettavuuteen liittyvää palautetta kerätään sähköpostiosoitteen tai palautelomakkeen kautta. Saadusta palautteesta

¹⁰ Yritys- ja yhteisötunnuksesta ja sen pysyvyydestä tarkemmin: <https://www.ytj.fi/index/y-tunnus.html>

kootaan yhteenveto ja keskeiset kehittämisehdotukset kootaan hyödynnettäväksi.

5.2 Laatuksiteeri: Käyttöoikeudet

Kuvaus: Käyttöoikeudet kuvaa sitä, miten tietoaaineiston käyttöoikeus on määritelty ja mitä aineistolla voi tehdä eli mihin käyttötarkoituksiin tietoaaineistoa voi hyödyntää.

Esimerkit: Esimerkiksi tietoaaineisto on tietyin rajoituksin käytettävissä tieteelliseen tutkimukseen. Avoin data on lisensoitu.

5.2.1 Käyttöoikeus

Käyttöoikeus-mittari kertoo, miten tietoaaineiston käyttö on rajoitettu eli kenen käytettävissä tietoaaineisto on.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: käyttöoikeuden rajaus

Mittarin arvo kertoo, kenen on mahdollista saada tietoaaineisto käyttöönsä. Tietoaaineisto voi olla esimerkiksi avointa dataa, julkista tietoa tai määräaikaaisella käyttöluvalla, sopimusperusteisesti tai virkakäytön perusteella käytettävissä. Käyttöä voi rajoittaa tietosuojaanäkikulmat tai tiedonkeruussa tiedonantajalle annettu informaatio. Henkilötietojen käyttöä rajoittaa erityisesti yleinen tietosuoja-asetus (2016/679).

Mittarin soveltaminen:

Kuvataan tietoaaineiston käyttöoikeuden rajaus siitä näkökulmasta kuka tai mikä taho voi tietoaaineistoa käsitellä.

Suoralla tiedonkeruulla kerättävän henkilötiedon käsittely perustuu yleensä suostumukseen, jonka tiedonantaja antaa vastatessaan kyselyyn. Tiedonantajalle toimitetaan suostumuksen antamista varten tarkka

informaatio¹¹ siitä, mihin tarkoitukseen tietoja kerätään, mitä muita tietoja vastaajasta tullaan yhdistämään nyt kerättäviin tietoihin, mikä taho näitä tietoja tulee käsittelemään ja kuinka kauan. Tämä informointi voi rajata jotain uusia käyttötarkoituksia tai koottuun tietoaaineistoon liitettävien muiden tietojen käyttöä huomattavasti.

Tiukempien rajausten osalta on tärkeää kuvata tilannetta tarkemmin esimerkiksi tietoaaineiston aineistokuvauksessa tai linkin kautta. Esimerkiksi tieto- tai käyttöluvan myöntämisen ehdot on tärkeä tuoda esille.

Esimerkit: Tilastotieto on julkista tietoa. Monet Tilastokeskuksen yksikköaineistot ovat käyttöluvalla käytettävissä esimerkiksi tutkimustarkoituksessa.

5.2.2 Käytön rajoitukset

Käytön rajoitukset -mittari kertoo, mihin kaikkiin käyttötarkoituksiin tietoaaineiston tietoa on luvallista käyttää.

- Arviointitaso: tietoaaineisto
- Arvon esitysmuoto: käyttötarkoituksen rajaus

Mittarin arvo kertoo, millaiseen käyttötarkoitukseen tietoaaineistoa saa käyttää. Esimerkiksi käyttö voi olla virkakäyttöä päätöksentekoon tai virkakäytössä referenssinä, käyttöluvan tai sopimusehtojen rajoissa tai avoimen datan lisenssin rajaamaa käyttöä.

Lisäksi olisi hyvä lyhyesti kuvata mahdollisia rajoituksia tiedon käytön osalta. Suorissa tiedonkeruissa tiedonantajalle kuvatut tietojen käyttötarkoitukset rajaavat tietojen käyttöä ja yhdistämistä mahdollisiin muihin tietoaaineistoihin.

11 Tiedonantajien informoinnista tarkemmin esimerkiksi Tietoarkiston ohjeissa: <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/tutkittavien-informointi/>

Mittarin soveltaminen:

Tietoaineiston käytön rajoitukset kuvataan tiiviisti. Avoimen datan tapauksessa kerrotaan avoimen datan lisenssi¹² (esim. Creative Commons CC BY 4.0 tai CC0 -lisenssit), muussa tapauksessa rajaus tiiviisti tai linkki tarkempaan tietoon.

Esimerkit: Tilastokeskuksen käyttöluvallisista yksikköaineistoista saa tuottaa anonyymejä, summattuja tutkimustuloksia. Julkaistua tietoa saa käyttää vapaasti kunhan lähde mainitaan.

5.3 Laatuksiteeri: Oikea-aikaisuus

Synonyymit: **Täsmällisyys**

Kuvaus: Täsmällisyys tarkoittaa sitä, että tietoaineisto on käytettävissä ilmoitettuna ajankohtana ja riittävän tiheästi tietoaineistossa tapahtuviin muutoksiin nähden.

Esimerkit: Tietojen julkistamisaika ja julkistamistiheys on ilmoitettu. Julkistamisaikataulujen muutoksista kerrotaan etukäteen.

5.3.1 Määräpäivien noudattaminen

Määräpäivien noudattaminen -mittari kuvaa tietoaineiston suunnitellun toimitusaikataulun seuranta toteutuneeseen toimitusaikatauluun.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: toimitusviive/ei relevantti
- Mittarin tausta: Euroopan tilastojen käytännösäännöt

Mittaria sovellettaessa ilmoitetaan toimitetun aineiston viive suhteessa sovittuun määräpäivään eli ”toteutunut toimituksen päivämäärä – sovittu päivämäärä”.

12 Tietoa avoimesta datasta: <https://www.avoindata.fi/fi/tietoa-avoimesta-datasta/lisenssit>

Lisäksi viivetilanteessa tiedon käyttäjälle voidaan lisätietona kuvata viiveen syy.

Mittarin soveltaminen:

Mikäli tietoaineistolla on suunniteltu toimitusaika tai julkaisupäivä, niin mittarin arvo kertoo kuin paljon viivettä tietoaineiston toteutuneen toimitusajan ja suunnitellun toimitusajan välillä on. Lisätiedot esimerkiksi viiveen syystä voidaan lisätä tietoaineiston aineistokuvaukseen erityisesti silloin, jos viiveellä voi olla vaikutusta tietoaineiston käytölle.

5.3.2 Päivitystiheys

Päivitystiheys-mittari kuvaa tietoaineiston säännöllistä päivitystiheyttä.

- Arviointitaso: ominaisuus ja tietoaineisto

Arvon esitysmuoto: sanallinen ilmaus

- Mittarin tausta: ISO 19139 MDMaintenanceFrequencyCode ja maintenanceNote

Mittarin arvon tulee kuvata säännöllisen päivityksen tiheyttä. Annettavat ilmaukset esim. reaaliaikainen, jatkuva, viikoittain, kuukausittain, kerran vuodessa tai muu päivitysrytmi.

Mittarin soveltaminen:

Mittarin arvo kertoo tietoaineiston suunniteltujen päivitysten tiheyden. Tässä mittarissa ei oteta huomioon puutteiden tai virheiden aiheuttamia päivityksiä. Tietoaineiston tapauksessa käytetään sitä päivitystiheyttä, joka vaikuttaa tiedon käyttäjän saamaan aineistoon. Jos jotain ominaisuutta päivitetään esimerkiksi kerran viikossa ja tietoaineisto on saatavilla aina tuoreimman päivityksen mukana, niin tietoaineiston päivitystiheys on kerran viikossa vaikka pääosin päivitykset tapahtuvat harvemmin.

Esimerkit: Tiestöä ja nimistöä päivitetään **jatkuvasti**. Hallintorajoja ja rakennuksia päivitetään **vuosittain**. Muut kohteet päivitetään karttalehdittäin määräaikaisen ajantasaistusproessin yhteydessä **5–10 vuoden välein**.

5.3.3 Päivityksessä muuttuneet ominaisuustiedot

Päivityksessä muuttuneet ominaisuustiedot -mittari kuvaa muuttuneiden ominaisuuden arvojen osuutta kaikista tietoaaineiston ominaisuuden arvoista verrattaessa päivitettyä tietoaaineistoa edelliseen versioon.

Mittarin arvon tavoitteena on kuvata päivitysten aiheuttaman muutoksen suuruutta tietoaaineistossa.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia

Mittarin arvon on tarkoitus kertoa siitä, kuinka suuri vaikutus päivityksen odottamisella on mahdollisesti jo käytössä olevan tietoaaineiston sisältöön. Yksinkertaisimmillaan verrataan edellistä tietoaaineistoa ja päivitettyä tietoaaineistoa muuttuja kerrallaan ja todetaan muuttuneiden arvojen osuus päivitettyssä tietoaaineistossa.

Myös jatkuvasti päivittyvien tietojen kohdalla muutosten suuruuden kuvaaminen on hyödyllinen tieto tiedon käyttäjälle. Päivityksessä muuttuneiden ominaisuustietojen osuutta voi tarkastella esimerkiksi päivityksen kannalta mielekkään ajanjakson ajalta. Esimerkiksi kuukauden takaisen tietoaaineiston kanssa vertailtaessa todetaan, että ominaisuuden arvoista 35 % on muuttunut tarkastelujakson aikana.

Mittarin soveltaminen:

Mittarin arvo kuvaa päivityksen aineistoon aiheuttaman muutoksen suuruutta. Mittari lasketaan kunkin ominaisuuden muuttuneiden tietojen osuutena kaikista ominaisuuden arvoista kyseisessä tietoaaineistossa.

Esimerkit: Yritysrekisterin tietopalvelussa tietoaaineistotoimituksen jälkeen uuteen aineistotoimitukseen on lisätty uudet aloittaneet yksiköt ja niiden ominaisuustiedot sekä vanhojen yksiköiden osalta niiden ominaisuustiedoissa tapahtuneet muutokset sekä mahdolliset lopetustiedot.

