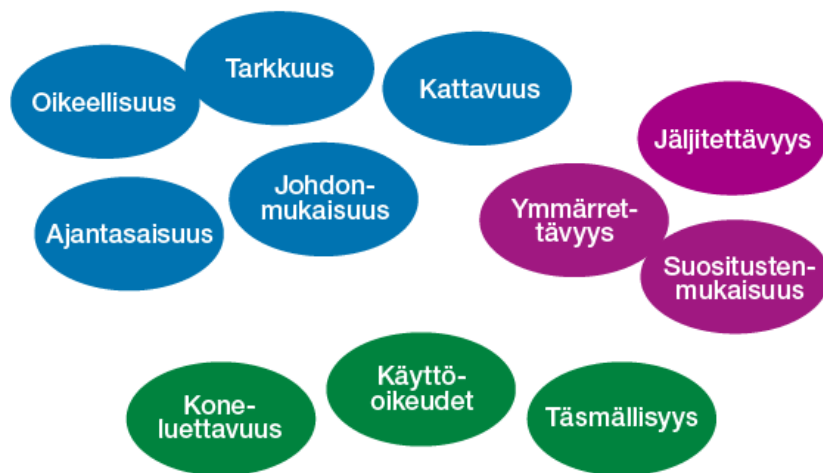


Tiedon laatukriteerit ja mittaristo, ehdotus suosituksiksi

Tiedon laatukehikko, TIHA TP3

11.2.2022



Sisälllys

1	Johdanto	3
2	Tiedon laadun kuvaaminen	4
2.1	Tavoitteet	4
2.2	Tunnistetut hyödyt	4
2.3	Rajaukset	4
2.4	Ymmärrys tiedon laadusta tietoaaineistojen laatukriteerien kautta	5
2.5	Huomio keskeisiin tekijöihin mittariston tuella	6
3	Laatukriteeriryhmä: Miten tieto kuvaa todellisuutta?	7
3.1	Laatukriteeri: Oikeellisuus	8
3.1.1	Menetelmällisesti tuotetut arvot	8
3.1.2	Virheelliset ominaisuusarvot	9
3.1.3	Väärinluokittelu	9
3.2	Laatukriteeri: Tarkkuus	9
3.2.1	Keskiahjonta	10
3.2.2	Poikkeavat havainnot	10
3.3	Laatukriteeri: Johdonmukaisuus	11
3.3.1	Tieto loogisuustarkastettu	11
3.4	Laatukriteeri: Ajantasaisuus	11
3.4.1	Viiteajankohta	11
3.4.2	Luontiajankohta	12
3.4.3	Tarkistusajankohta	12
3.4.4	Muutosajankohta	12
3.5	Laatukriteeri: Kattavuus	12
3.5.1	Ajallinen tavoitekattavuus	13
3.5.2	Alueellinen tavoitekattavuus	13
3.5.3	Tavoitekohdeyksiköt	13
3.5.4	Tavoiteominaisuuksissa puutteita	13
3.5.5	Puuttuvat kohdeyksiköt	14
3.5.6	Ylimääräiset kohdeyksiköt	14
3.5.7	Puutteelliset kohdeyksiköt	15
3.5.8	Puutteelliset ominaisuustiedot	15
4	Laatukriteeriryhmä: Miten tieto on kuvattu?	15
4.1	Laatukriteeri: Jäljitettävyys	16
4.1.1	Tietolähde	16
4.1.2	Tiedon elinkaari	16
4.1.3	Muutosten hallinta	17
4.2	Laatukriteeri: Ymmärrettävyys	17
4.2.1	Aineistokuvaus	17

4.2.2	Käsitteiden määritelmät.....	17
4.2.3	Ominaisuuksien tietokuvaukset.....	18
4.2.4	Ymmärrettävyyden asiakaspalaute	18
4.3	Laatukriteeri: Suositustenmukaisuus	18
4.3.1	Noudatettavat säädökset ja standardit	19
5	Laatukriteeriryhmä: Miten tietoa voi käyttää?	19
5.1	Laatukriteeri: Koneluettavuus	19
5.1.1	Tietoaineiston tietomalli.....	20
5.1.2	Kohdeyksikön pysyvä tunniste	20
5.1.3	Koneluettavuuden asiakaspalaute	20
5.2	Laatukriteeri: Käyttöoikeudet	21
5.2.1	Käyttöoikeus	21
5.2.2	Käytön rajoitukset.....	21
5.3	Laatukriteeri: Täsmällisyys	21
5.3.1	Määräpäivien noudattaminen	22
5.3.2	Päivitystiheys	22
5.3.3	Päivityksessä muuttuneet ominaisuustiedot.....	22
	Liitteet	23
	Liite 1 - Dokumentissa esiintyvien termien kuvaukset.....	23
	Liite 2 - Dokumentissa esiintyvät periaatteet ja standardit	26
	Liite 3 – Mittaristo.....	27

1 Johdanto

Tämä ehdotus suositukseksi julkishallinnon tiedon laadun kuvauksen ja arvioinnin laatukriteereistä ja mittaristosta on valmisteltu Valtiovarainministeriön asettaman Tiedon hyödyntämisen ja avaamisen hankkeen (TiHA, VN/5386/2020) työpaketissa kolme, hankkeen toimikaudella 30.4.2020–31.12.2022.

Suositusluonnos oli avoimesti kommentoivana julkisen lausuntopalvelun kautta 6.9.-4.10.2021 välisenä aikana.

Tässä suosituksessa kuvataan laatukriteerit ja mittaristo, jotka yhdessä tarjoavat selkeän ja ymmärrettävän työkalun julkishallinnon tietoaineistojen, ja nimenomaan rakenteisen datan, yhteismitallista ja käyttäjälähtöistä laadun kuvaamista ja tarkastelua varten. Tämä suositus on suunnattu erityisesti tiedon laadun kysymysten parissa työskenteleville tiedon tuottajille ja tiedon laadusta kiinnostuneille tiedon käyttäjille. Tavoitteena on tarjota, ensimmäistä kertaa, yhteinen kansallinen kieli ja termistö tiedon laadun käsittelyyn tiedon laadun asiantuntijoille ja siitä kiinnostuneille.

Tässä suosituksessa ei käsitellä laatukriteerien ja mittariston käyttöönottoa organisaatioissa eikä laatukriteerit ja mittaristo -työkalun ylläpitoa ja hallintaa. Tämän suosituksen laatukriteeri- ja mittarikuvaukset kertovat keskeisimmät soveltamisen näkökulmat, mutta tavoitteena on tarjota rinnalle lisäksi erillisiä soveltamisohjeita. Myöskään laatutiedon viestintä- ja jakeluratkaisut eivät sisälly tähän suositukseen. Lisäksi tulee huomioida, että laatuajattelun täytyy kulkea käsi kädessä tietotuote- ja järjestelmäkehityksen sekä asiakaspalvelun kehittämisen kanssa. Tietovarantojen monipuolinen käyttö ja yhdisteltävyys edellyttävät myös panostamista rajapintojen ja sähköisten järjestelmien käytettävyyteen ja palveluiden tunnettuuteen ja saatavuuteen. Näitä näkökulmia ei kuitenkaan käsitellä tässä suosituksessa.

Tiedon laatu on monitahoinen käsite. Laatukriteerit ja mittaristo pyrkivät yhdessä vastaamaan siihen, mitä laatu julkishallinnon tietoaineistojen osalta tarkoittaa ja mitä kaikkia osa-alueita tulee ottaa huomioon tiedon laatua kuvattaessa. Laatukriteerien määrittelyssä on tavoiteltu kattavaa mutta tiivistä kokoelmaa tiedon laadun eri näkökulmat huomioon ottaen. Tavoitteeseen pääsemisessä keskeisessä roolissa olivat projektiin osallistuneet henkilöt, jotka edustivat laajasti erilaisia tiedon suunnittelun, käsittelyn, luovutuksen, analysoinnin ja laadun tulokulmia. Myös mittaristo kehitettiin yhdessä ja testattiin konkreettisilla tietoaineistoilla erilaisissa piloteissa.

Nyt kehitetyt laatukriteerit pohjaavat pitkälti ISO 25012 -standardiin, jota käytetään myös useissa muissa maissa tietoaineistojen laadun kuvauksen taustamallina. Lisäksi laatukriteereihin on tuotu perustietovarannoille tärkeä oikeellisuuden kriteeri sekä asiakasnäkökulmaa. Kriteerit ovat myös linjassa Euroopan yhteentoimivuusperiaatteen (EIF), FAIR-periaatteiden ja Euroopan tilastojen käytännesääntöjen kanssa. Tiedon laatukriteerien soveltamista tukevat erillisten mittarien muodostama mittarikokoelma eli mittaristo. Tuotetuista laadun mittareista osalla on selkeämpi, osalla löyhempi kytkös yleisiin standardeihin tai muualla käytössä oleviin mittareihin.

Käytännön kehittämistyöhön osallistuivat Tilastokeskus, Maanmittauslaitos, Verohallinto, Tulli, Luonnonvarakeskus, Valtiokonttori ja Kansaneläkelaitos. Hankkeen laajan projektiryhmän kautta työhön ovat osallistuneet edellä lueteltujen lisäksi Digi- ja väestötietovirasto, Opetushallitus, Patentti- ja rekisterihallitus ja Työterveyslaitos. Lisäksi laatukriteerijä on esitelty TiHA-hankkeen tilaisuuksissa ja sidosryhmätapaamisissa muun muassa kuntasektorille ja ne ovat olleet avoimesti kommentoitavina Tilastokeskuksen sivuilla ennen pilotointia

2 Tiedon laadun kuvaaminen

2.1 Tavoitteet

Laatukriteerien ja mittariston muodostaman kokonaisuuden tavoitteena on tukea tietoaaineistojen, erityisesti rakenteisen datan laadun tunnistamista ja kuvaamista tiedon käyttäjille erilaisissa tiedon vaihdon tilanteissa sekä tiedon laadun yhtenäistä vertailua ja kehittämistä julkishallinnossa. Laatukriteerit yhdessä mittariston kanssa tavoittelevat siis yhdenmukaista kuvausta erilaisten tietoaaineistojen laadusta yli organisaatio- ja alarajojen ja auttavat siten käyttäjiä mm. arvioimaan, soveltuuko tietoaaineisto aiottuun käyttötarkoitukseen. Sisällöltään laadukkaan tiedon lisäksi kriteerien valinnan taustalla ovat vaikuttaneet tavoitteet tietojen helposta löydettävyydestä, yhdisteltävyydestä, yhteentoimivuudesta sekä sujuvasta käyttökokemuksesta.

Laatukriteerien ja mittariston suunnittelussa on laatua pyritty kuvaamaan ymmärrettävästi siten, että aineiston soveltuvuuden tarkastelu onnistuu myös ilman aiempaa kokemusta kyseisestä tietoaaineistosta tai tiedon laadun määrittämiseen liittyvää erityisosaamista. Tavoitteena on ollut myös, että kynnys laatukriteerien ja mittariston käyttöönottoon olisi mahdollisimman matala.

2.2 Tunnistetut hyödyt

Edellisessä kappaleessa kuvatut tavoitteet helpottavat tiedon soveltuvuuden arviointia eri käyttötarkoituksiin ja tukevat siten osaltaan julkishallinnon tietovarantojen nykyistä monipuolisempaa käyttöä. Näin laatukriteerit ja mittaristo yhdessä tarjoavat työkalun, joka osaltaan tukee tiedon avaamista, yhteentoimivuutta ja hyödyntämistä.

Työn aikana on tunnistettu myös muita hyötyjä kuten tiedon laadun seuraaminen ja parantaminen sekä prosessien kehittämisen ohjaaminen, kun mittareilla saadaan tuotettua ajassa vertailtavaa tietoa tietoaaineistojen laadusta. Mittariston odotetaan helpottavan tiedon laadun viestintää myös organisaatioiden sisällä.

Yhdessä kehitettävien laatukriteerien ja mittariston myötä myös kansallinen tiedon laadun ymmärrys ja osaaminen kasvaa ja samalla rakennetaan yhteistä kieltä, jolla tiedon laatua kuvataan ja siitä keskustellaan. Tiedon laatu voidaan nyt myös määrittää yhteisesti esimerkiksi kehityshankkeissa ja lainsäädännön valmistelussa. Laatukriteerit on tarkoitettu julkishallinnon työkaluksi, mutta niitä voidaan hyödyntää myös muualla yhteiskunnassa, kuten julkisissa hankinnoissa.

Pidemmällä aikavälillä kehittyvä laatukriteerien ja mittariston kokonaisuus voisi myös mahdollistaa julkishallinnon tiedon laadun kansallisen ohjauksen ja seurannan, mikäli se katsotaan tarpeelliseksi.

Vastaavaa mallia tiedon laadun kansalliseen määrittämiseen ei aiemmin ole ollut. Kansainvälisesti verraten tämä suositus on ensimmäisiä tiedon laatua määritteleviä kansallisia malleja. Se antaa siten mahdollisuuden osallistua omalla panoksella myös kansainväliseen keskusteluun ja kehittämiseen.

2.3 Rajaukset

Laatukriteerien ja mittariston kehittämisessä on tehty seuraavat rajaukset:

- Laatukriteereitä ja mittaristoa kehitetään ainoastaan rakenteiselle datalle.

- Kehitettävän mittariston ensisijainen tavoite on palvella erilaisia tiedonvaihdon tilanteita ja tiedon käyttäjistä puhuttaessa tarkastellaan lähinnä organisaatioiden ulkopuolisia käyttäjiä.
- Laatuksiteerit ja mittaristo eivät ota kantaa tavoiteltaviin laatusoihin.

2.4 Ymmärrys tiedon laadusta tietoineistojen laatuksiteerien kautta

Tiedon laatu on laaja ja moniulotteinen käsite. Tiedon laatu määritellään yleisesti sen perusteella, kuinka hyvin tieto soveltuu käyttäjän tarpeisiin. Nyt kehitetyt laatuksiteerit ja mittarit tarkastelevat tiedon laadun eri ulottuvuuksia sidottuna ensisijaisesti julkishallinnon tiedonvaihdon eri tilanteisiin ja siihen, mitkä laadun osa-alueet ovat keskeisiä tiedon käyttäjän näkökulmasta.



Kuva 1. Tietoineistojen laatuksiteerit ryhmittäin. Tiedon laatu tarkasteltaessa katsottiin keskeisiksi näkökulmat, jotka kertovat käyttäjälle siitä, mitä tieto koskee ja miten hyvin tarkasteltava tieto kuvaa todellisuutta. Tiedon käytettävyyden näkökulmasta olennaista on tietää miten tietoa voi käyttää ja miten hyvin tieto on kuvattu.

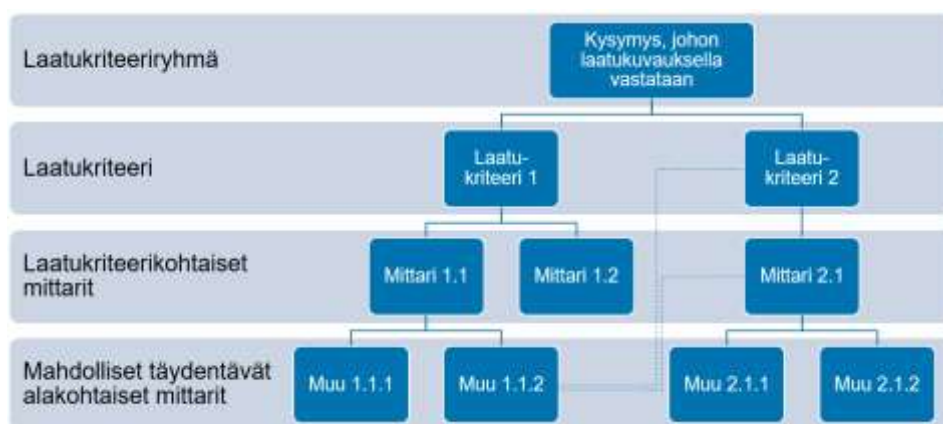
Tiedon vaihdon tilanteissa keskeiseksi haasteeksi nousee yhteisen ymmärryksen rakentaminen tiedon käyttäjien tietotarpeiden ja tiedon tuottajien tietolupausten välille. Laatuksiteerit auttavat tunnistamaan, mitkä tekijät ovat tiedon laadun näkökulmasta tärkeitä. Käyttäjälle on tärkeää saada kattava ja oikea kuva tietoineiston sisällöstä, sisällön ja sen kuvauksen laadusta ja osuvuudesta sekä tiedon käyttömahdollisuuksista. Seuraavassa tarkastellaan näitä eri näkökulmia käyttäjälähtöisten kysymysten ja kysymyksiin vastaavien laatuksiteerien kautta.

Laatuksiteerit kuvattuna ryhmittäin. Ryhmiä on kolme, joissa jokaisessa on kolmesta viiteen laatuksiteeriä. Miten tieto on kuvattu? -ryhmässä ovat laatuksiteerit jäljitettävyys, ymmärrettävyys ja suositustenmukaisuus. Miten tieto kuvaa todellisuutta? -ryhmässä ovat ajantasaisuus, johdonmukaisuus, oikeellisuus, tarkkuus ja kattavuus. Miten tietoa voi käyttää? -ryhmässä ovat kone-luettavuus, käyttö-oikeudet ja täsmällisyys.

Metatieto on tietoa, joka kuvailee toista tietojoukkoa. Tiedon laadun kuvaukset voidaan siten nähdä osana tietoineistojen (metatieto)kuvauksia. Laatuksiteerien ja mittarien tuottamat arvot tarvitsevat tuekseen myös muita tietoineistojen kuvausten tietoelementtejä, mm. tavoitekohdejoukon kuvauksen ja sen sijainnin. Tällainen kattava kuvailutieto on olennaista, jotta laatumittarien antamia arvoja voi tulkita ja ymmärtää. Siksi tähän laatuksiteerien ensimmäiseen versioon on nähty tarpeelliseksi sisällyttää myös sellaisia kuvaus-elementtejä, jotka laajemmassa metatietomallissa, kuten Rekisteritiedon metatietomallissa tai organisaatioiden omissa metatietomalleissa saattavat sijoittua luontevasti jonkin toisen metatietoelementin kuin laadun alle. Esimerkkinä tällaisesta voisi ajatella joitakin ajantasaisuuden laatuksiteeriin liitettyjä mittareita. Tämän lisäksi mittaristossa on mukana useampia mittareita, jotka peräänkuuluttavat

keskeisimpien metatietojen määrittelyä ja saatavuutta. Vähintään nämä metatiedot tulisi tarjota tiedon käyttäjälle laatukuvauksen rinnalla.

Laatukuvauksen rakenne on tiukka hierarkkinen malli (Kuva 3). Se on selkeä ja mittaristo tukee näin parhaiten laatukriteerien ymmärtämistä. Laatukriteerien ja erityisesti mittariston työstämisessä on tunnistettu paljon organisaatio-, aineistotyyppi- ja alakohtaisia laadun kuvaamisen eroja. Mittariston peruskokonaisuudeksi on kuitenkin koostettu joukko sellaisia mittareita, jotka soveltuvat käyttöön mahdollisimman laajasti. Laadun kuvaamista voidaan kuitenkin täydentää toimialakohtaisilla ja muilla mahdollisilla täydentävillä mittareilla. Näille mittareille on annettu enemmän vapausasteita hierarkiassa.



Kuva 2. Kaaviokuva laatukuvauksen rakenteesta.

Hierarkkisesta kuvaamistavasta huolimatta laatu, jota kriteerit ja mittarit kuvaavat, on kokonaisuus, jossa laadun eri elementit vaikuttavat toisiinsa. Laadun parannus jonkin laatukriteerin suhteen voi jopa heikentää toisen laatukriteerin avulla kuvattua tiedon laatua. Esimerkiksi jos tavoitellaan tietoaineiston täydellistä kattavuutta tai ominaisuustietojen erityisen hyvää tarkkuutta niin tiedon ajantasaisuus yleensä heikkenee.

Laatukriteerit ja niihin liittyvät mittarit on esitelty tarkemmin laatukriteeriryhmittäin kappaleissa 3 - 5. Lisäksi mittarit on kerätty taulukkoon Liitteeseen 3. Liitteeseen 1 kootun tämän dokumentin sanaston rooli on tärkeä ja sitä tulee kehittää edelleen. Sanaston kehittämisellä ja tarkalla sisällön kuvauksella päästään lähemmäs tilannetta, jossa asioista puhutaan samoilla käsitteillä ja ne ymmärretään mahdollisimman samalla tavalla. Yhteisymmärrys on tärkeää laatukriteerien ja mittarien jatkokehittämisen sekä laadun vertailemisen kannalta.

2.5 Huomio keskeisiin tekijöihin mittariston tuella

Laatukriteereitä täydentävä mittaristo tukee laatukriteerien ymmärtämistä ja tiedon laadun tunnistamista luomalla kriteereille konkreettisen sisällön. Mittarinen määrä vaihtelee kriteerien välillä, kun osassa kriteereistä on päästy suhteellisen helposti laadun kuvaamisen ytimeen ja toisissa on koettu tarvetta lähestyä kuvaamista useammasta suunnasta. Mittaristo sisältää sekä määrällisiä mittareita että tiedon sisällön ja kuvauksen toteavia mittareita (kyllä/ei). Toteavien mittarien osalta oletetaan, että lisätieto mittarin aiheesta on tiedon käyttäjän löydettävissä muusta tietoaineiston dokumentaatiosta. Tavoitteena on, että mittaristo tarjoaa tiedon käyttäjälle mahdollisuuden tarkastella yhteismitallisesti julkishallinnon tietoaineistojen laatua.

Mittarinen tunnistamisessa on huomioitu erityisesti seuraavat näkökulmat:

- Mittarit kuvaavat yksityiskohtaisemmin laatuksiteerien mukaista laatu näkökulmaa käsiteltävästä tietoaineistosta.
- Mittari kuvaa laatua ja erityisesti laatuksiteeriä selkeästi, yksikäsitteisesti ja ymmärrettävästi.
- Mittarien tulee olla yhteisiä, kaikille käyttäjille soveltuvia ja hyvin erilaisista tietoaineistoista kohtalaisen helposti tuotettavia.
- Jokaisella laatuksiteerillä on vähintään yksi mittari.
- Kullekin mittarille on sovittu yksi yhteinen suositeltu esittämismuoto tai tunnusluku.

Kappaleissa 3 – 5 on kuvattu laatuksiteerit sekä kuhunkin kriteeriin liittyvät mittarit. Tämän suosituksen kuvaukset ovat tiiviit ja esittelevät, mutta kertovat mittareistakin keskeisimmät soveltamisen näkökulmat, joita esimerkit havainnollistavat. Lisäksi mittariston soveltamisessa tulee huomioida seuraavat yleiset periaatteet ja neuvot:

- Laatuksiteerit ja mittaristo on tarkoitettu joustavaksi työkaluksi; kaikki kriteerit ja siten mittarit eivät välttämättä ole merkityksellisiä kaikissa tilanteissa. Laatua suositellaan kuitenkin tarkasteltavan tuotetun kokonaisuuden kautta mahdollisesti uusistakin näkökulmista ja tiedon käyttäjän tarpeista lähtien.
- Tiedon käyttäjät eli asiakkaan tunnistaminen: Eri hallinnonaloilla on erilaisia asiakkaita. Pohdi keitä tietoaineiston asiakkaat ovat ja millä perusteilla asiakasnäkökulmaa voidaan tuoda laadun tarkasteluun? Onko tunnistettavissa asiakasryhmiä, joilla on erityisiä laatuun liittyviä tarpeita?
- Mittariston soveltaminen koostettuun tietoon toteutetaan sillä tasolla millä tietoja on kuvattu. Esimerkiksi jos yritystiedot on summattu toimialatasolle, niin puuttuvien kohdeyksiköiden osuus -mittaria sovelletaan puuttuvien toimialojen tasolla.
- Mikäli tarkan kuvauksen, esim. numeroarvon tuottaminen on hankalaa, voi määrittelyssä käyttää myös arviota. Mieluummin siis arvio kuin jätetään kokonaan täyttämättä tarkasteltavan tiedon laadun kannalta merkityksellinen mittari. Jos kyseessä on arvio, on se hyvä mainita ja perustella.
- Laatuksiteerit ja mittaristo ovat tiivis kokonaisuus ja eri elementit linkittyvät toisiinsa siten, että laadun parantuminen yhtäällä voi merkitä heikkenemistä toisaalla. Esimerkiksi jos tavoitellaan tietoaineiston täydellistä kattavuutta tai ominaisuustietojen erityisen hyvää tarkkuutta niin tiedon ajantasaisuus yleensä heikkenee.
- Mittaristoa täydentävien mittarien käyttö on sallittua ja jopa suotavaa.

3 Laatuksiteeriryhmä: Miten tieto kuvaa todellisuutta?

Tiedon tarkastelu lähtee liikkeelle ilmiöstä, jota koskevaa tietoa halutaan käyttää. Saatavilla olevan tiedon hyödynnettävyyteen liittyy vahvasti tiedon sisällölliset tavoitteet eli mitä kaikkia näkökulmia tarkasteltavasta ilmiöstä tiedon halutaan kuvaavan. Tiedon laatua kuvaa se, miten hyvin tieto vastaa tätä tavoiteltua sisältökokonaisuutta. Laadukkaan tiedon tulisi kuvata tavoiteltua ilmiötä mahdollisimman tarkasti ja oikeellisesti. Käytettävissä olevan tiedon tulisi olla mahdollisimman ajantasaista.

Tiedon kattavuuden kautta kuvataan toisaalta kohdejoukkoa, jonka ominaisuuksia tietoaaineiston on tarkoitus kuvata ja toisaalta sitä, miten hyvin tavoiteltu kohdejoukko on tietoaaineistossa mukana. Yleensä tietoaaineiston aineistokuvaus (metatiedot) sisältää laajasti tietoa siitä, mikä tietoaaineiston tavoiteltu sisältö on. Tiedon laatukuvauksessa on kuitenkin haluttu korostaa keskeisimmät tarkasteltavat näkökulmat tästä tavoitetilasta. Kattavuus-laatukriteeriin sisältyy myös näkökulma siitä, miten laajasti tavoiteltuja ominaisuustietoja tietoaaineistoon sisältyy.

Ajantasaisuus-laatukriteerissä on mukana useita näkökulmia tiedon tuoreuden tarkasteluun. Tiedon tulisi olla mahdollisimman tuoretta suhteessa viiteajankohtaan eli ajankohtaan, jota tieto kuvaa. Toisaalta päivittämätönkään tieto ei välttämättä ole huonolaatuista, jos ominaisuuden arvoissa ei ole tapahtunut muutoksia.

Laadukas tieto kuvaa todellisuutta tarkasti ja oikeellisesti eli systemaattiset vääristymät tai muut virhelähteet on tunnistettu ja niiden vaikutus on korjattu. Johdonmukainen tieto ei sisällä sisäisiä ristiriitoja.

Miten tieto kuvaa todellisuutta?



Kuva 3. Laatukriteerien mittarit kysymyksen ”Miten tieto kuvaa todellisuutta?” alla.

3.1 Laatukriteeri: Oikeellisuus

Synonyymit: *virheettömyys*

Kuvaus: Oikeellisuus 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 tietoaaineiston 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. Tässä tarkoitetaan sellaisella korvaavalla tiedolla tai laskennallisella menetelmällä tuotettuja arvoja, jotka eivät ole täsmällisesti kohdeyksikön saamia oikeita arvoja. Esimerkiksi edellisen vuoden tietojen käyttäminen puuttuvien tietojen täydentämisessä, on menetelmällisesti tuotettujen arvojen käyttämistä. Tiedon korjaaminen suoraan tiedonantajalta selvittämällä saadulla oikealla arvolla ei kuulu tämän mittarin tarkasteluun.

Esimerkit: Tulotiedon täydentäminen luovuttajaimputoinnilla käyttäen sukupuolta, ikää, koulutusta ja ammattia arvon luovuttajan valinnan kriteerinä. Pienten yritysten toimintaa kuvaavia tietoja tarkastellaan 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 soveltaessa 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.

Esimerkit: Tietoaineistossa tiedetään olevan päivittämättömiä tietoja ja kokemuksen perusteella osataan arvioida 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 soveltaessa 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.

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

3.2 Laatuksiteeri: Tarkkuus

Synonyymit: *harhattomuus*

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

Esimerkit: Esimerkkejä tarkkuudesta ovat mitattujen arvojen hajonta ja tietoaineiston 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 voisi paremmin hahmottaa. Ilman keskiarvoa vaihtelu jää asteikosta erilliseksi. Keskiarvo sinällään ei kuitenkaan kuvaa tarkkuutta, joten siksi arvo ei ole mukana erillisenä mittarina. Arvon tuottaminen on olennasta myös siksi, että sekä keskiarvoa että keskihajonnan arvoa tarvitaan myös poikkeavien havaintojen määrittämisessä.

Esimerkit: Keskihajonnan ollessa pientä, ovat ominaisuuden arvot keskittyneet lähelle keskiarvoa ja vastaavasti suuren keskihajonnan tapauksessa arvot ovat enemmän hajallaan. Tämä voi liittyä epätarkkuuteen tai sitten olla vain ominaisuudelle tyypillistä vaihtelua.

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 soveltaessa tulee huomioida, että mittarin arvo ilmoitetaan poikkeavien havaintojen lukumäärän suhteena 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 määritellä koneellisesti.

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

Esimerkit: Tarkastellessa henkilöiden tulotietoja, huomattavan suurituloiset ovat henkilöaineistossa poikkeavia havaintoja. Huomattavan suurituloisten tulotiedot voivat aiheuttaa haastetta analyseissä.

3.3 Laatuksiteeri: Johdonmukaisuus

Synonyymit: *säännönmukaisuus, tiedon looginen eheys*

Kuvaus: Johdonmukaisuus kertoo siitä, että tietoaineisto on yhtenäinen ja ristiriidaton. Johdonmukaisuudella voidaan kuvata myös eri tietoaineistojen 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 tietoaineiston koostamisessa tai käsittelyssä.

- Arviointitaso: ominaisuus ja tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria soveltaessa 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.

Esimerkit: Ominaisuuden ”toimipaikan osoite” osalta on käytetty loogisuusehtoa, että 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 päättämään 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 tietoaineiston sisältämä tieto on kerätty eli ajankohdan, jota tieto koskee. Tietoaineiston käsittely aiheuttaa viivettä ja tietoa on voitu kerätä aiemmin kuin tietoaineisto on valmis tiedon käyttäjän käsiteltäväksi.

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

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

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

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 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 tärkeä viestiä vertailukelpoisten tietojen ajanjaksot.

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.

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.

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. Ylipeitto kertoo, että tietoaineistossa on siihen kuulumattomia

yksiköitä. 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-alatiedot 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.

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

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.

Esimerkit: FinTerveys 2017 alueellinen kattavuus on Manner-Suomi.

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

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

3.5.4 Tavoiteominaisuuksissa puutteita

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

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria soveltaessa 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.

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 soveltaessa tulee huomioida, että arvo ilmoitetaan tietoaineistosta (tavoiteperusjoukosta) puuttuvien kohdeyksiköiden lukumäärän suhteena tietoaineistoon kuuluvien kohdeyksiköiden lukumäärään.

Lisäksi tulee huomioida, että valmiiksi koostetun tiedon, esim. tilasto, osalta puuttuvuus on yleensä korjattu tilastollisin menetelmin. Tällöin voidaan olettaa, ettei tarkasteltavassa tietoaineistossa 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 kuvauksen tietoaineiston kuvaukseen, mikäli tietoaineiston puuttuvista kohdeyksiköistä on riittävä käsitys.

3.5.6 Ylimääräiset kohdeyksiköt

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

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

Mittaria soveltaessa tulee huomioida, että arvo ilmoitetaan ylimääräisten kohdeyksiköiden lukumäärän suhteena tietoaineiston 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 tietoaineisto sisältää tietäntyyppisiä ylimääräisiä kohdeyksiköitä ja tätä ylimääräisten kohdeyksiköiden joukkoa on mahdollista kuvata, kuvaus olisi tärkeää sisällyttää tietoaineiston kuvaukseen.

Esimerkit: Tietoaineistossa on samoja kohdeyksiköitä useamman kerran, koska tietoja on tuotu tietoaineistoon useasta lähteestä ja kaikilla ei ole sama kohdeyksikön tunniste käytössä. Otostutkimuksen kohdeyksiköiden otannassa on

mukana maasta 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 soveltaessa 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 tiettyyn kohdejoukkoon.

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 tietylle ominaisuudelle.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: prosenttia

Mittaria soveltaessa tulee huomioida, että mittarin arvo ilmaistaan puuttuvan ominaisuustiedon omaavien kohdeyksiköiden lukumäärän suhteena kaikkien kohdeyksiköiden lukumäärään ominaisuuksittain tarkasteltuna. 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 kuten esimerkiksi tuloeria sellaisen tulon osalta, jota tarkasteltava henkilö ei saa.

4 Laatuksiteeriryhmä: Miten tieto on kuvattu?

Laadukkainkin tietoaineisto ilman tiedon kuvausta voi olla pahimmillaan jopa käyttökeltvotonta. Tiedon käytettävyys vaatii aineistokuvauksen ja ominaisuuksien kuvaukset, jotta tietoaineistolle ja ominaisuuksien arvoille on mielekäs ja ymmärrettävä tulkinta. Tietojen yhteensovittamisen kannalta erilaisten suosituksen: standardien, käytäntöjen ja säädösten noudattaminen on olennaista.

Jäljitettävyys taas kertoo siitä, onko tiedon alkuperä ja muutokset tiedossa ja jälkikäteen tarkasteltavissa.

Miten tieto on kuvattu?



Kuva 4. Laatuksiteerien mittarit kysymyksen ”Miten tieto on kuvattu?” alla.

4.1 Laatuksiteeri: Jäljitettävyys

Synonyymit: *kiistämättömyys*

Kuvaus: Jäljitettävyys kertoo siitä, että tietoaineistoon 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, tietoaineiston 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 soveltaessa 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 raportoinnin laaturaporttien kautta.

Lisäksi tulee huomioida, että tiedon käyttäjää kiinnostaa tietää tietolähteet ja miltä osin lähdetietoja puuttuu. Tietolähteen osalta on olennaisinta kertoa edellinen taho, joka tiedon on toimittanut, jolloin haluttaessa laatuksitarkastelussa 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 tiedonkäyttäjä saa tietolähteen lisäksi lisätietoa tiedon prosessointihistoriasta.

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 tiedon muutokset luomisesta tiedon arkistointiin tai poistamiseen ja sisältää tiedon lähteiden tai tuottamisen muutokset, esimerkiksi laskentatavan muutokset.

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

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 myöhemminkin.

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

Esimerkit: Tietoaineiston ominaisuuksien muutoksista pidetään lokia tai tieto on löydettävissä metatietojen kautta.

4.2 Laatuksiteeri: Ymmärrettävyys

Synonyymit: *tulkittavuus, käsitettävyyys*

Kuvaus: 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 soveltaessa tulee huomioida, että arvo ilmaistaan kertomalla ne kielet, joilla aineistokuvaus on saatavilla.

Lisäksi tulee huomioida, että kieliversio 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.

4.2.2 Käsitteiden määritelmät

Käsitteiden määritelmät -mittari kertoo, onko tietoaineiston käsitteet määritelty selkeästi ja saatavilla sekä mitä kieliversioita niistä on saatavilla. Käsitteellä

tarkoitetaan niitä tietoaineistoon liittyviä asioita, joita ei voida välttämättä suoraan mitata.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kieliversiot

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

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

Esimerkit: Hyvinvointi-käsitteen osalta kuvataan tietoaaineiston määritelmä hyvinvoinnille. Eli kuvataan ne näkökulmat, joiden kautta hyvinvointia mitataan käsiteltävässä tietoaaineistossa.

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

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

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 seurataan ja palautteeseen myös reagoidaan.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria soveltaessa 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 tietoaaineiston kehittämisessä.

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

4.3 Laatuksiteeri: Suositusten mukaisuus

Synonyymit: *yhteensopivuus, semanttinen yhdenmukaisuus, yhdenmukaisuus*

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

Esimerkit: Kansallista yhdenmukaisuutta tukee muun muassa yhtenäisten kansallisten sanastojen ja koodistojen käyttö tietoaisteistojen 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 tietoaisteiston noudattamat säädökset, standardit, hyvät käytänteet ja suositukset listattuna tietoaisteiston kuvauksissa.

- Arviointitaso: ominaisuus ja tietoaisteisto
- Arvon esitysmuoto: kyllä/osittain/ei
- Mittarin tausta: INSPIRE/FAIR -periaatteet

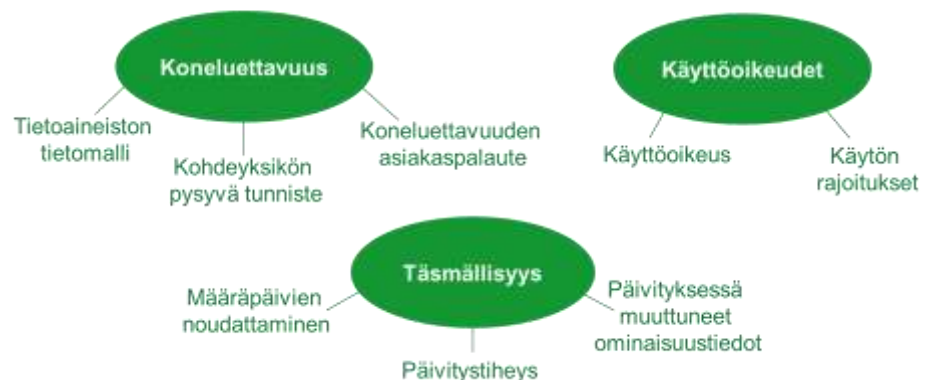
Mittaria soveltaessa tulee huomioida, että säädöksiä ja standardeja on monen tasoisia: yleisiä ja toimialakohtaisia sekä kansallisia ja kansainvälisiä. Näistä vähintään keskeisimmät tulisi luetella esimerkiksi aineistokuvauksen yhteydessä. Myös ominaisuuksien osalta yleisten koodistojen käyttö tulee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

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.

5 Laatuksiteeriryhmä: Miten tietoa voi käyttää?

Tietoaisteistojen käytölle on rajoituksia sen suhteen mikä taho tietoja voi tarkastella, mihin käyttötarkoitukseen niiden käyttö on sallittu ja missä muodossa tiedot ovat saatavilla. Rajoitteita aiheuttavat esimerkiksi tiedonkeruussa tehty tiedonantajaa informointi ja tietosuoja. Olennaista on myös varmistua, että tietoa on saatavilla silloin, kun se on luvattu. Tiedon käytettävyyttä tarkastellaan myös erityisesti koneuettavuuden näkökulmasta.

Miten tietoa voi käyttää?



Kuva 5. Laatuksiteerien mittarit kysymyksen ”Miten tietoa voi käyttää?” alla.

5.1 Laatuksiteeri: Koneuettavuus

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

Esimerkit: Tietoaineisto on rakenteisessa formaatissa, esim. csv, json, xml. Tietoaineiston rakenne on kuvattu, esim. skeema.

5.1.1 Tietoaineiston tietomalli

Tietoaineiston tietomalli -mittari kertoo, onko tietoaineisto kuvattu rakenteisesti.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

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

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

Esimerkit: Kansallisen maastotietokannan rakennustietojen rakenne on kuvattu JHS 210-standardissa. Tietoaineiston rakenne on kuvattu skeemalla (esim. xml, json).

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 soveltaessa 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.

Esimerkit: INSPIRE-soveltamissäädökset edellyttävät paikkatietokohteen tunnuksen julkaisemista HTTP URI-muodossa. (JHS 193)

5.1.3 Koneluettavuuden asiakaspalaute

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

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: kyllä/ei

Mittaria soveltaessa 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 tietoaineiston koneluettavuuden parantamiseksi. Lisäksi tulee huomioida, että palautteen yhteenvedo tai palautteen aiheuttamat muutokset ovat hyödyllistä lisätietoa tiedon käyttäjille.

5.2 Laatuksiteeri: Käyttöoikeudet

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

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

5.2.1 Käyttöoikeus

Käyttöoikeus-mittari kertoo, miten tietoaineiston käyttö on rajoitettu eli kenen käytettävissä tietoaineisto on.

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: käyttöoikeuden rajaus

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

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

5.2.2 Käytön rajoitukset

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

- Arviointitaso: tietoaineisto
- Arvon esitysmuoto: käyttötarkoituksen rajaus

Mittarin arvo kertoo, millaiseen käyttötarkoitukseen tietoaineistoa 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 tietoaineistoihin.

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

5.3 Laatuksiteeri: Täsmällisyys

Synonyymit: *Oikea-aikaisuus*

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 seurantaan toteutuneeseen toimitusaikatauluun.

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

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

Lisäksi viivetilanteessa tiedon käyttäjälle voidaan lisätietona kuvata viiveen syy.

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.

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

5.3.3 Päivityksessä muuttuneet ominaisuustiedot

Päivityksessä muuttuneet ominaisuustiedot -mittari kuvaa muuttuneiden ominaisuuden arvojen osuutta kaikista tietoaineiston ominaisuuden arvoista verrattaessa päivitettyä tietoaineistoa edelliseen versioon. Mittarin arvon tavoitteena on kuvata päivitysten aiheuttaman muutoksen suuruutta tietoaineistossa.

- Arviointitaso: ominaisuus
- Arvon esitysmuoto: prosenttia

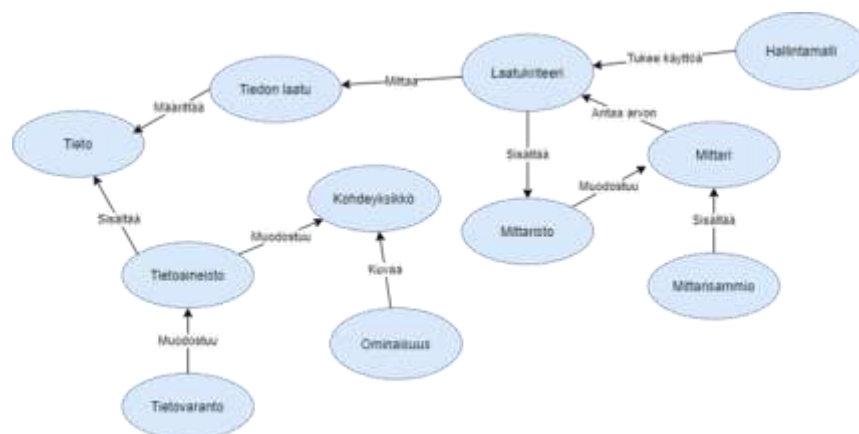
Mittarin arvoa tuottaessa ominaisuustasolla on mahdollista määritellä muuttuneiden arvojen osuus ominaisuuksittain tarkasteltuna. Tietoaineistotasolla muutosta lienee helpompi kuvata sanallisesti. Mittarin arvon on tarkoitus kertoa siitä, kuinka suuri vaikutus esimerkiksi päivityksen odottamisella on mahdollisesti jo käytössä olevan tietoaineiston sisältöön. Lisäksi on hyvä huomata, että 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 tarkasteltuna.

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

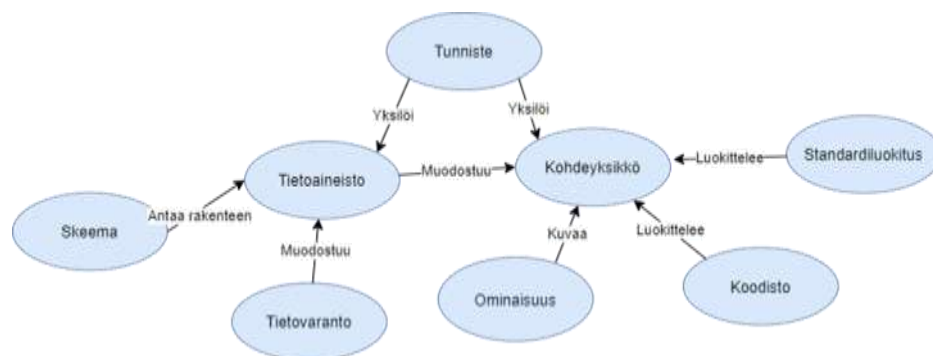
Liitteet

Liite 1 - Dokumentissa esiintyvien termien kuvaukset

Tässä liitteessä on kuvattu termit siten kuin ne on käytetty tässä suosituksessa.



Kuva 6. Kuvitus tämän dokumentin tiedon laadun kuvaukseen liittyvistä termeistä.



Kuva 7. Kuvitus tämän dokumentin rakenteiseen dataan liittyvistä termeistä.

Taulukko 1. Dokumentissa esiintyvien termien kuvaukset.

Termi (synonyymi)	Kuvaus
kohdeyksikkö (luokka, tilastoyksikkö, yksikkö, kohde)	Tietoaineistossa tarkasteltava yksikkö.
koodisto	Kokoelma ryhmiä, jotka eroavat toisistaan tietyiltä piirteiltään.
laatukriteeri	Tekijä, joka kuvaa laatukuvauksen laatua tietyistä näkökulmista.
mittari	Laatukriteerin näkökulmaa konkretisoiva ja tietoaineiston ominaisuutta mittaava mittari. Kukin mittari viittaa tiettyyn laatukriteeriin. Yhdelle laatukriteerille voidaan soveltaa useita mittareita.
mittarisammio	Mittariaihoiden, mittariehdotusten muodostama kokonaisuus, josta hankkeen aikana poimittiin sopivin kokonaisuus laatukriteerien mittaristoksi. Jatkossa mittarisammio toimii paitsi mittariehdotuksien myös mittaristoa täydentävien mittareiden kokoelmana.
mittaristo	Tiedon laatukriteerien soveltamiseen tuotettu erillisten mittarien muodostama mittarikokoelma.
ominaisuus (ominaisuus, attribuutti, kohde, muuttuja)	Kohdeyksikköä kuvaava tieto.
perustietovaranto	Perustietovaranto on alemman jalostusasteen tietovaranto tai jalostetun tiedon taustalla oleva tietovaranto.
rakenteinen data	Koneluettava tieto, johon on liitetty metatietojen avulla määritetty rakenne, jonka avulla tietoa voidaan jäsentää.
skeema	Tietoaineiston rakenteen määrämuotoinen esitys.
standardiluokitus (luokitusstandardi)	Luokitus suositus, joka perustuu esimerkiksi EU-direktiiveillä vahvistettuihin kansainvälisiin standardeihin.
tiedon käyttäjä	Tietoa käyttävä henkilö. Tiedon käsittelyprosessin näkökulmasta prosessin viimeinen, tietoa todella käyttävä henkilö. Tiedon käyttäjä voi myös edelleen jatkojalostaa tietoa, mutta tämä on uusi käsittelyprosessi verrattuna tiedon laadun kuvaukseen.
tiedon laatu	Se, missä määrin jonkin tietoaineiston ominaisuudet täyttävät vaatimuksia tai tavoitteita. Tiedon soveltuvuus siihen käyttötarkoitukseen, johon käyttäjä aikoo sitä käyttää ja tiedon tuottaja sitä tarjoaa.
tieto	Tiedolla voi viitata monenlaisiin asioihin: merkkijono, viesti, tosiasia, havainto, tulkinta tai käsitys. Tässä tiedolla tarkoitetaan ilmiöistä koottua alimman jalostusasteen tietoa tai siitä jalostamalla tuotettua informaatiota.
tietoaineisto	Aineisto, joka sisältää tietovälineelle tallennettuja tietoja.
tietovaranto	Tiettyä tarkoitusta varten muodostettu, loogisesti tai fyysisesti yhteenkuuluvien tietojen muodostama tietoaineisto tai tietoaineistojen kokoelma.
tunniste	Yksilöimiseen käytettävä merkkijono.

Taulukko 2. Dokumentissa esiintyvien tilastollisten termien kuvaukset.

Termi (synonyymi)	Kuvaus
Alipeitto, tilastomenetelmäkäsité	Alipeitto liittyy tutkimuksessa käytettyyn perusjoukkoon. Perusjoukosta täytyy olla käytettävissä kehikko eli luettelo kohteista, joita koskevia tietoja halutaan otantatutkimuksella kerätä. Alipeitolla tarkoitetaan sitä, että käytettävissä olevasta kehikosta puuttuu osa perusjoukon eli tutkimuksen kohdejoukon alkioista. Esim. henkilöt, joilla ei ole puhelinta, puuttuvat puhelinhaastattelun kehikosta.
Imputointi (paikkaus), tilastomenetelmä	(paikkaus) tarkoittaa havaintoaineistossa olevan puutteen tai poikkeavan havainnon korvaamista jollain imputointimenetelmällä. Imputointimenetelmiä ovat mm. looginen imputointi (korjataan loogisesti mahdolliset virheet, esim. lapsi ei voi olla vanhempi kuin vanhempansa), hot deck -imputointi, jossa puutteelle haetaan arvo toiselta vastaajalta, cold deck -imputointi, jossa havainto haetaan saman vastaajan edellisestä vastauksesta, regressio- ja muut malliperusteiset menetelmät, jossa käytetään tilastollista mallia ennustamaan arvo puutteelle.
Perusjoukko (populaatio), tilastomenetelmäkäsité	Perusjoukko on tutkimuksen kohteena oleva ryhmä, jota koskevia tietoja halutaan kerätä, esim. äänestysikäiset kansalaiset. Siitä käytetään myös tarkempaa nimitystä kohdeperusjoukko. Siitä erillään on kehikkoperusjoukko, jolla tarkoitetaan tutkimuksen käytössä olevan rekisterin tai muun luettelon peittämää kohdejoukkoa, joka ei kuitenkaan aina vastaa yksi yhteen kohdeperusjoukkoa (ks. ylipeitto, alipeitto).
Poikkeava havainto (outlier), tilastomenetelmäkäsité	poikkeava havainto (outlier) on sellainen muuttujan arvo, joka poikkeaa suuresti havaittujen arvojen valtaenemmistöstä ja se voi olla joko oikea tai virheellinen. Poikkeavilla arvoilla voi olla merkittävästi vääristävä vaikutus käytettyihin tilastollisiin tunnuslukuihin, kuten keskiarvoon, hajontaan, regressiosuoraan
Ylipeitto, tilastomenetelmäkäsité	ylipeitto tarkoittaa otoskehikkoon sisältyviä kohdeyksiköitä, jotka eivät enää kuulu kohdeperusjoukkoon, esim. laitoksiin siirtyneitä, kuolleita, maasta muuttaneita. Tällaisia tapauksia esiintyy aina jonkin verran otosyksiköiden joukossa, koska rekisterit, joista otokset poimitaan, eivät ole täysin ajantasaisia.

Liite 2 - Dokumentissa esiintyvät periaatteet ja standardit

Nimi	Selite
EIF	Euroopan yhteentoimivuusperiaate
Euroopan tilastojen käytäntesäännöt	The European Statistics Code of Practise
FAIR	FAIR-periaatteet: Tiedot ovat Löydettäviä (Findable), Saavutettavia (Accessible), Yhteentoimivia (Interoperable) ja Uudelleenkäytettäviä (Re-usable)
ISO 19139	Geographic information - Metadata - XML schema implementation
ISO 19157	Geographic information - Data quality
ISO 25012	Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model
ISO 25024	Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Measurement of data quality
ISO-kielikoodisto	ISO 639 määrittelemä kielikoodis

Liite 3 – Mittaristo

Taulukko 3. Miten tieto kuvaa todellisuutta? -kysymykseen vastaaviin laatukriteereihin liitetyt mittarit.

Nimi	Kuvaus	Esitysmuoto	Arviointitaso	Laatukriteeri
Menetelmällisesti tuotetut arvot	Menetelmällisesti tai korvaavaa tietoa käyttäen tuotettujen arvojen osuus ominaisuuden kaikista arvoista	prosenttia	Ominaisuus	Oikeellisuus
Virheelliset ominaisuus-arvot	Virheellisiä ominaisuusarvoja saavien kohdeyksiköiden osuus kaikista kohdeyksiköistä	prosenttia	Ominaisuus	Oikeellisuus
Väärinluokittelu	Virheellisesti luokiteltuja ominaisuuden arvoja saavien kohdeyksiköiden osuus kaikista kohdeyksiköistä	prosenttia	Ominaisuus	Oikeellisuus
Keskihajonta	Ominaisuuksien arvojen vaihtelu keskiarvon ympärillä	Keskihajonta	Ominaisuus	Tarkkuus
Poikkeavat havainnot	Poikkeavien havaintojen lukumäärän suhde kaikkien kohdeyksiköiden lukumäärään	prosenttia	Ominaisuus	Tarkkuus
Tieto loogisuus-tarkastettu	Tiedon keräämisessä, koostamisessa tai käsittelyssä on käytetty loogisuusehtoja	kyllä/ei	Ominaisuus ja tietoaaineisto	Johdonmukaisuus
Viiteajankohta	Ajankohta, jota tietoaaineisto kuvaa	ajankohta	Ominaisuus ja tietoaaineisto	Ajantasaisuus
Luonti-ajankohta	Kohdeyksikön tai ominaisuuden luontiajankohta	ajankohta	Ominaisuus ja tietoaaineisto	Ajantasaisuus
Tarkistus-ajankohta	Kohdeyksikön tai ominaisuuden tarkistusajankohta	ajankohta	Ominaisuus ja tietoaaineisto	Ajantasaisuus
Muutos-ajankohta	Kohdeyksikön tai ominaisuuden muutosajankohta	ajankohta	Ominaisuus ja tietoaaineisto	Ajantasaisuus
Ajallinen tavoite-kattavuus	Tietoaaineiston tavoiteltava ajallinen kattavuus ja tiheys on kuvattu.	kyllä/ei	Tietoaaineisto	Kattavuus
Alueellinen tavoite-kattavuus	Tietoaaineiston tavoiteltava alueellinen kattavuus ja tiheys on kuvattu.	kyllä/ei	Tietoaaineisto	Kattavuus
Tavoitekohdeyksiköt	Tietoaaineiston aineistokuvauksessa on selkeästi määritelty ajallisen ja alueellisen tavoitekattavuuden lisäksi muut tietoaaineiston rajaukset	kyllä/ei	Tietoaaineisto	Kattavuus

Tavoiteominaisuuksissa puutteita	Tietoaaineiston ilmiön kannalta keskeisiä ominaisuuksia puuttuu arvioitavasta tietoaaineistosta	kyllä/ei	Tietoaaineisto	Kattavuus
Puuttuvat kohdeyksiköt	Tietoaaineiston alipeitto eli tietoaaineiston tavoitekohdejoukosta puuttuvien kohdeyksiköiden prosentuaalinen osuus	prosenttia	Tietoaaineisto	Kattavuus
Ylimääräiset kohdeyksiköt	Tietoaaineiston ylipeitto eli tietoaaineistossa mukana olevien, mutta tietoaaineiston tavoiteperusjoukkoon kuulumattomien kohdeyksiköiden prosentuaalinen osuus	prosenttia	Tietoaaineisto	Kattavuus
Puutteelliset kohdeyksiköt	Yhdenkin puuttuvan ominaisuustiedon sisältävien kohdeyksiköiden osuus tietoaaineiston kaikista kohdeyksiköistä	prosenttia	Ominaisuus ja tietoaaineisto	Kattavuus
Puutteelliset ominaisuustiedot	Puuttuvan ominaisuustiedon omaavien kohdeyksiköiden osuus kaikista kohdeyksiköistä ominaisuusittain tarkasteltuna	prosenttia	Tietoaaineisto	Kattavuus

Taulukko 4. Miten tieto on kuvattu? -kysymykseen vastaaviin laatukriteereihin liitetyt mittarit.

Nimi	Kuvaus	Esitysmuoto	Arviointitaso	Laatukriteeri
Tietolähde	Tietoaaineiston, kohdeyksikön tai ominaisuuden lähdetieto on saatavilla	prosenttia	Ominaisuus ja tietoaaineisto	Jäljitettävyys
Tiedon elinkaari	Tietoaaineiston elinkaari määritelty ja kuvaus siitä saatavilla	kyllä/ei	Tietoaaineisto	Jäljitettävyys
Muutosten hallinta	Ominaisuustiedoissa tapahtuvia rakenteellisia tai lähdetietojen muutoksia seurataan	kyllä/ei	Ominaisuus	Jäljitettävyys
Aineistokuvaus	Mitä kieliversioita tietoaaineiston aineistokuvauksesta on saatavilla	kieliversiot	Tietoaaineisto	Ymmärrettävyys
Käsitteiden määritelmät	Kieliversiot, joilla keskeisten käsitteiden määritelmät ovat saatavilla	kieliversiot	Tietoaaineisto	Ymmärrettävyys
Ominaisuusien tietokuvaukset	Kieliversiot, joilla keskeisten käsitteiden määritelmät ovat saatavilla	kieliversiot	Ominaisuus	Ymmärrettävyys
Ymmärrettävyyden asiakaspalaute	Ymmärrettävyydestä on mahdollista antaa asiakaspalautetta, saatua	kyllä/ei	Tietoaaineisto	Ymmärrettävyys

	palautetta seurataan ja siihen reagoidaan			
Noudatettavat säädökset ja standardit	Tietoaineiston noudattamat säädökset, standardit, hyvät käytänteet ja suositukset ovat listattuna tietoaineiston kuvauksissa	kyllä/osittain/ei	Ominaisuus ja tietoaineisto	Suosittelun mukaisuus

Taulukko 5. Miten tietoa voi käyttää? -kysymykseen vastaaviin laatukriteereihin liitetyt mittarit.

Nimi	Kuvaus	Esitysmuoto	Arviointitaso	Laatukriteeri
Tietoaineiston tietomalli	Tietoaineisto on rakenteisesti kuvattu	kyllä/ei	Tietoaineisto	Koneluettavuus
Kohdeyksikön pysyvä tunniste	Tietoaineiston kohdeyksiköillä on vähintään tietoaineistokohtainen pysyvä tunniste	kyllä/ei	Tietoaineisto	Koneluettavuus
Koneluettavuuden asiakaspalaute	Koneluettavuudesta on mahdollista antaa asiakaspalautetta, saatua palautetta seurataan ja siihen reagoidaan	kyllä/ei	Tietoaineisto	Koneluettavuus
Käyttöoikeus	Kuvataan tietoaineiston käyttöoikeuden rajaukset: esim. käsittelijä, käsittely-ympäristö	käyttöoikeuden rajaus	Tietoaineisto	Käyttöoikeudet
Käytön rajoitukset	Kuvataan tietoaineiston käyttötarkoitusten rajoitukset: esim. avoimen datan lisenssit tai käyttöehtoja	käyttötarkoituksen rajaus	Tietoaineisto	Käyttöoikeudet
Määräpäivien noudattaminen	Tietoaineiston toimitusaikataulun seuranta toteutuneeseen toimitusaikatauluun verrattuna	toimitusviive	Tietoaineisto	Täsmällisyys
Päivitystiheys	Tietoaineiston säännöllinen päivitystiheys	sanallinen ilmaus	Ominaisuus ja tietoaineisto	Täsmällisyys
Päivityksessä muuttuneet ominaisuustiedot	Päivityksessä muuttuneiden ominaisuusarvojen osuus kaikista tietoaineiston ominaisuusarvoista	prosenttia	Ominaisuus	Täsmällisyys