

## Tiedon laatukriteerien visualisointipilotin tulokset ja esimerkkejä

### Sisältö

Sisältö.....	1
Visualisointipilotin tavoite .....	1
Visualisointipilotin toteutus .....	1
Huomautukset materiaalin käytöstä .....	2
Visualisointipilotin johtopäätökset ja tulevaisuuden näkymät.....	2
Liite 1: Esimerkkejä pilotissa tuotetuista visualisoinneista .....	3
Liite 2: Esimerkkejä Power BI:n tietomallista .....	15

### Visualisointipilotin tavoite

Visualisointipilotin tarkoituksena on ollut tuottaa ensimmäisiä ideoita siitä, miten tiedon laatukriteerien ja mittariston arvoja/tuloksia voitaisiin kuvata visuaalisesti. Mittaristo on laaja ja sisältää monia mittareita, joten helposti selailtava ja tiivistetysti tietoa sisältävä esitys helpottaisi laatuarvioihin tutustumista. Visualisoinnille asetettiin seuraavat tavoitteet:

- Tarjota nopeasti omaksuttava kuvaus tiedon laadun tasosta.
- Tukea mittariston haltuunottoa, ymmärtämistä ja käyttöönottoa visualisoinnin kautta.
- Mahdollistaa erojen havaitseminen erityyppisten aineistojen välillä siinä, mitkä mittarit tai minkä tyyppinen visualisointi on tärkeää.
- Mahdollistaa yhden yhtenäisen visualisoinnin toteuttaminen case-visualisointien jälkeen.

### Visualisointipilotin toteutus

Visualisointipilotti (dashboard-tyyppinen) toteutettiin lokakuussa 2022 Microsoftin Power BI -sovelluksella yhteistyössä ohjelmistoyhtiö Pinja Oy:n kanssa. Pilotin taustamateriaalina käytettiin Tiedon laatukehikko -osahankkeen vuoden 2021 lopussa toteutettujen mittaristopilotointien tuloksia. Nämä tulokset olivat vuoden 2022 alussa toimineet syötteenä laatukriteerien viimeistelyssä ja uutta pilotointia ei enää viimeistelyn jälkeen tehty. Siksi nyt tehdyssä visualisointipilotoinnissa käytössä ollut laatukriteeristö eroaa ryhmittelyltään ja erityisesti joidenkin mittarien osalta suositusehdotuksesta. Visualisoinnin sanastossa ja mittareissa on tästä syystä näkyvillä eroja verrattuna Tiedon laatukriteerit ja mittaristo, ehdotus suositukseksi -dokumenttiin<sup>1</sup>.

Visualisointipilotissa keskityttiin tarkastelemaan kahta erillistä tilannetta:

<sup>1</sup> Dokumentti saatavilla sivustolla: [stat.fi/tiedonlaatu](http://stat.fi/tiedonlaatu)

- Ensimmäisessä harjoituksessa tarkastelimme niin kutsutun raakadatan eli tarkastamattoman ja muokkaamattoman datan arvoja verrattuna tarkastettuun ja korjattuun aineistoon. Näistä toinen on täysin keksittyä, mutta asiantuntemuksen mukaan uskottavia mittarin arvoja sisältävä tieto ja toinen perustuu löyhästi pilotointiin.
- Toisessa harjoituksessa hyödynsimme ajatuksen tasolla useampia pilotoinnissa olleita tietoja, mutta mittarin arvoja muutettiin tai lisättiin osin varsin rajusti visualisoinnin ehdoilla. Harjoituksessa päätimme käyttää tietoaineistoista merkintöjä Aineisto A – Aineisto F.

Käytimme visualisointipilotissa lähtökohtana mittaristo-exceliä. Kyseisen Excel-tiedoston rakenne vaati kuitenkin muokkausta harjoituksen edistyessä. Laatuksien soveltamisen tueksi tuotettu mittaristopohja ei siis toimi suoraan Power BI -sovelluksen kanssa yhteen. Visualisointipilotin lopputuotosten esittely toteutettiin 1.11.2022 tiedon laadun yhteistyöverkoston tilaisuudessa Tilastokeskuksen tiloissa Helsingin Kalasatamassa.

## Huomautukset materiaalin käytöstä

Tuotetut visualisoinnit eivät perustu aitoihin tietoaineistojen laatuarviointeihin. Visualisointeja tulee sen sijaan tarkastella esimerkkeinä siitä, miten mittariston monet mittarin arvot olisi mahdollista kuvata erilaisia visuaalisia tehostekeitä hyödyntäen. Voit lähettää visualisointeja tarkastellessa mieleen tulevat kehitysajat jatkoyöstettäväksi osoitteella [tiedonlaatu@stat.fi](mailto:tiedonlaatu@stat.fi).

## Visualisointipilotin johtopäätökset ja tulevaisuuden näkymät

Visualisoinnin todettiin olevan hyvä tapa esitellä tietoja:

- Nyt tiedetään mitä pitäisi tehdä eli visualisoinnin lisäarvo toteutuu vertailunäkökulman kautta ja vertailukohta on aiheesta ja kontekstista riippuvainen. Visualisoinnissa voi verrata toteutuneita arvoja esimerkiksi vertailuarvoihin tai tavoitearvoihin.
- Tiedetään, mitä ja miten kannattaa visualisoida - esimerkiksi yksinkertaisetkin visualisoinnit toimivat ja ovat helppoja lukea.
- Tiedetään ja omataan käsitystä, mitä visualisointivälineestä voi saada irti.
- Käytettyjä tietomalleja ja Power BI:n tietovarastoa eli sitä missä muodossa datan tulee olla, pitää vielä miettiä. Voisiko mittaristo-excelin pohja olla sellainen, että siitä olisi helppo viedä tietoja suoraan myös Power BI -sovellukseen?
- Visualisoinnissa voisi olla käytössä myös yksi koostenäkymä, jossa kaikkien kriteerien tulokset on koottu yhdelle sivulle.
- Visualisoinnissa voisi olla käytössä toiminnallisuus, jossa käyttäjä voi itse antaa haluamansa raja-arvot aineistojen tai kriteerien haussa.
- Visualisoinnissa pitäisi saada helposti käyttöön myös yhden kalvon raportteja ja enemmän erilaisia muotoiluja, esimerkiksi liikennevalotyyppejä visualisointeja.
- Power BI -sovelluksen käyttäminen vaatii yleensä peruskoulutuksen tai johdantokurssin pohjalle ennen kuin visualisointia pystyy tekemään kunnolla Power BI -sovelluksella.

Ideoita visualisoinnin kehittämiseen kerätään jatkossa myös tiedon laadun yhteistyöverkostolta.

## Liite 1: Esimerkkejä pilotissa tuotetuista visualisoinneista

Visualisointipilotissa käytettiin todellista tietoaineiston laatumittariston pilottien dataa ja kuvitteellista dataa. Visualisointipilotissa pyrittiin tuomaan esille vertailuja eri aineistojen kesken ja toteuman suhdetta tavoitearvoihin. Esimerkeissä on tuotu esille kahden aineiston vertailua, tässä raakadatan ja palveludatan eroja ja toisaalta viiden eri aineiston eroja. Raakadatatalla tarkoitetaan käsittelemätöntä dataa ja palveludatatalla käsiteltyä dataa.

Tiedon laatumittarit:



Kuva 1. Julkisen hallinnon tiedon laatumittarit ryhmittäin.

Esimerkissä on vertailtu kahta aineistoa, raaka- ja palveludataa, yksittäisen kriteerin näkökulmasta. Kuvaamisessa on käytetty kyllä/ei -vaihtoehtoja ja näitä kuvaavia symboleja, huomiomerkkiä sekä prosenttiosuuksia ja pylväsdiagrammeja.

Kuvissa 2 – 8 tarkastellaan esimerkkejä laatumittariryhmästä, miten tieto kuvaa todellisuutta. Kuvissa 9 – 10 on esimerkit ryhmän, miten tieto on kuvattu, visualisoinnista. Ja kuvissa 11 – 13 on visualisoitu laatumittariryhmää, miten tietoa voi käyttää.

## Kattavuus - Puutteelliset kohdeyksiköt

Kattavuus (koko aineisto)

Puutteelliset kohdeyksiköt

Puutteelliset ominaisuustiedot



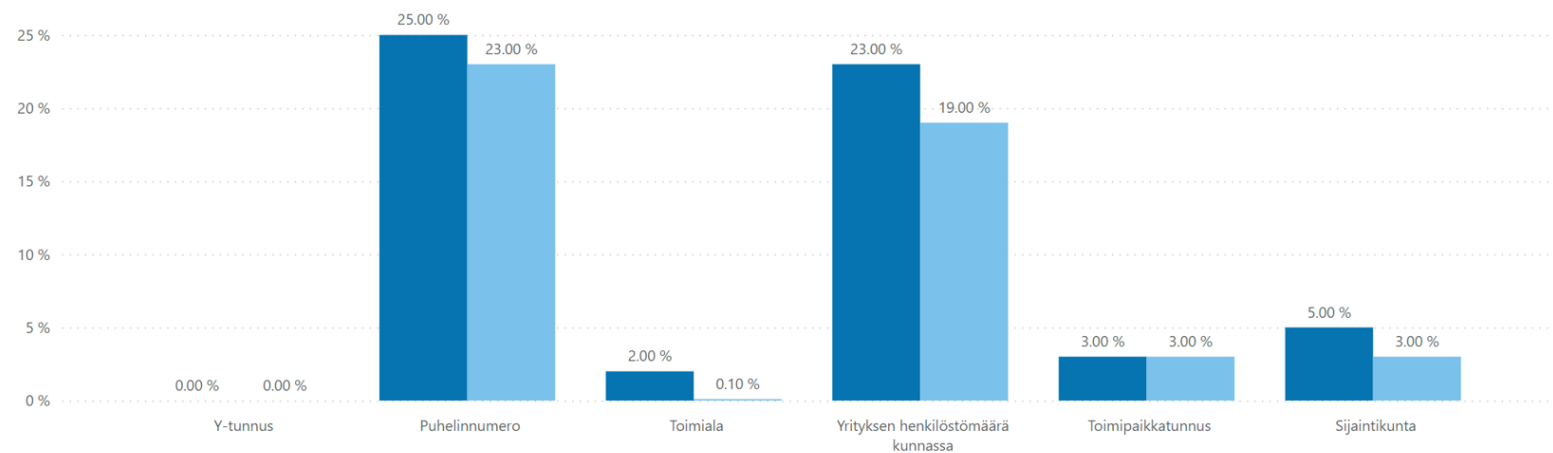
Ominaisuus	Raakadata	Palveludata
Y-tunnus	0.00 %	0.00 %
Puhelinnumero	25.00 %	23.00 %
Toimiala	2.00 %	0.10 %
Yrityksen henkilöstömäärä kunnassa	23.00 %	19.00 %
Toimipaikkatunnus	3.00 %	3.00 %
Sijaintikunta	5.00 %	3.00 %

Ominaisuus	Raakadata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 25.00 %
Toimiala	✓ 2.00 %
Yrityksen henkilöstömäärä kunnassa	✗ 23.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 5.00 %

Ominaisuus	Palveludata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 23.00 %
Toimiala	✓ 0.10 %
Yrityksen henkilöstömäärä kunnassa	✗ 19.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 3.00 %

### Puutteelliset kohdeyksiköt

Datan tyyppi ● Raakadata ● Palveludata



Kuva 2. Esimerkki laatukriteerin kattavuus visualisoinnista puutteellisten kohdeyksiköiden näkökulmasta kukin ominaisuus erikseen tarkasteltuna.

**Kattavuus - Puutteelliset ominaisuustiedot**

Kattavuus (koko aineisto)

Puutteelliset kohdeyksiköt

**Puutteelliset ominaisuustiedot**



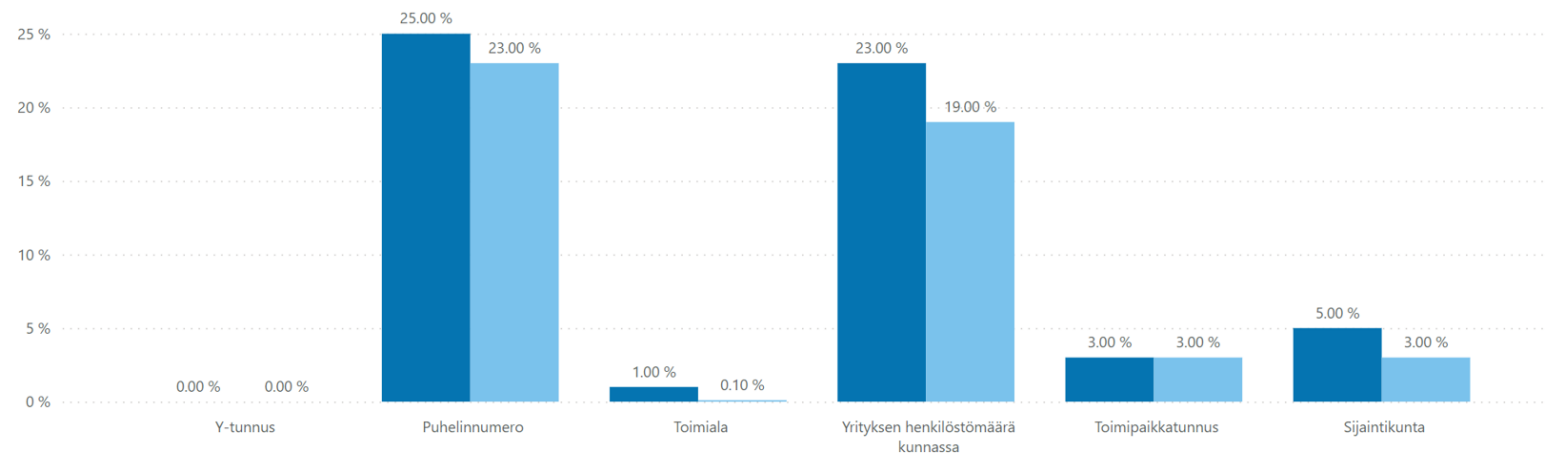
Ominaisuus	Raakadata	Palveludata
Y-tunnus	0.00 %	0.00 %
Puhelinnumero	25.00 %	23.00 %
Toimiala	1.00 %	0.10 %
Yrityksen henkilöstömäärä kunnassa	23.00 %	19.00 %
Toimipaikkatunnus	3.00 %	3.00 %
Sijaintikunta	5.00 %	3.00 %

Ominaisuus	Raakadata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 25.00 %
Toimiala	✓ 1.00 %
Yrityksen henkilöstömäärä kunnassa	✗ 23.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 5.00 %

Ominaisuus	Palveludata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 23.00 %
Toimiala	✓ 0.10 %
Yrityksen henkilöstömäärä kunnassa	✗ 19.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 3.00 %

**Puutteelliset kohdeyksiköt**

Datan tyyppi ● Raakadata ● Palveludata



Kuva 3. Esimerkki laatukriteerin kattavuus visualisoinnista puutteellisten ominaisuustietojen näkökulmasta kukin ominaisuus erikseen tarkasteltuna.

### Kattavuus - Puutteelliset kohdeyksiköt

Kattavuus (koko aineisto)

Puutteelliset kohdeyksiköt

Puutteelliset ominaisuustiedot



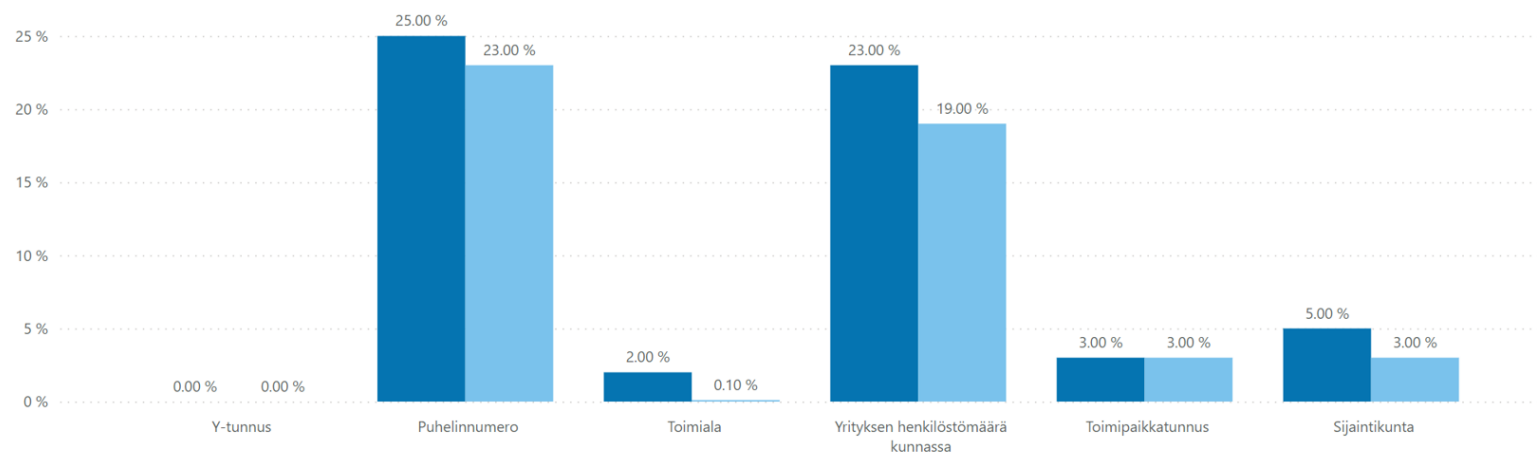
Ominaisuus	Raakadata	Palveludata
Y-tunnus	0.00 %	0.00 %
Puhelinnumero	25.00 %	23.00 %
Toimiala	2.00 %	0.10 %
Yrityksen henkilöstömäärä kunnassa	23.00 %	19.00 %
Toimipaikkatunnus	3.00 %	3.00 %
Sijaintikunta	5.00 %	3.00 %

Ominaisuus	Raakadata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 25.00 %
Toimiala	✓ 2.00 %
Yrityksen henkilöstömäärä kunnassa	✗ 23.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 5.00 %

Ominaisuus	Palveludata
Y-tunnus	✓ 0.00 %
Puhelinnumero	✗ 23.00 %
Toimiala	✓ 0.10 %
Yrityksen henkilöstömäärä kunnassa	✗ 19.00 %
Toimipaikkatunnus	✓ 3.00 %
Sijaintikunta	✓ 3.00 %

### Puutteelliset kohdeyksiköt

Datan tyyppi ● Raakadata ● Palveludata



Kuva 4. Esimerkki laatukriteerin kattavuus visualisoinnista yhden mittarin "Puutteelliset kohdeyksiköt" näkökulmasta.

19.12.2022

## Ajantasaisuus

Ajantasaisuus (koko aineisto)

Ajantasaisuus (ominaisuudet)

### Datan päivityspäivämäärä

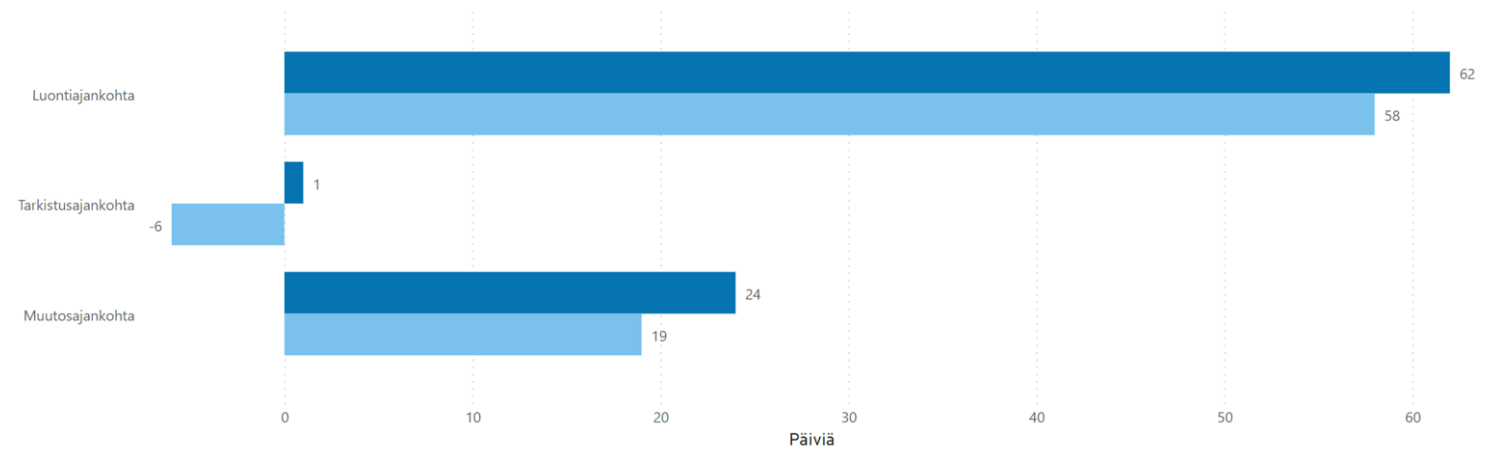
Mittarin nimi	Raakadata	Palveludata
Viiteajankohta	8/1/2022	8/1/2022
Luontiajankohta	8/23/2022	8/27/2022
Tarkistusajankohta	10/23/2022	10/30/2022
Muutosajankohta	9/30/2022	10/5/2022

### Montako päivää datan päivityksestä on?

Mittarin nimi	Raakadata	Palveludata
Luontiajankohta	62	58
Tarkistusajankohta	1	-6
Muutosajankohta	24	19

### Datan ajantasaisuus - montako päivää datan päivityksestä on

Datan tyyppi ● Raakadata ● Palveludata



Kuva 5. Esimerkki laatukriteerin ajantasaisuus visualisoinnista kaikki laatukriteerin mittarit huomioiden.

## Ajantasaisuus - ominaisuudet

Ajantasaisuus (koko aineisto)

Ajantasaisuus (ominaisuudet)

### Datan päivityspäivämäärä

Ominaisuus	Y-tunnus	
Mittarin nimi	Raakadata	Palveludata
Luontiajankohta	Tuesday, August 23, 2022	Saturday, August 27, 2022
Tarkistusajankohta	Sunday, October 23, 2022	Sunday, October 30, 2022
Muutosajankohta	Friday, September 30, 2022	Wednesday, October 05, 2022

Datan tyyppi

All

Mittarin nimi

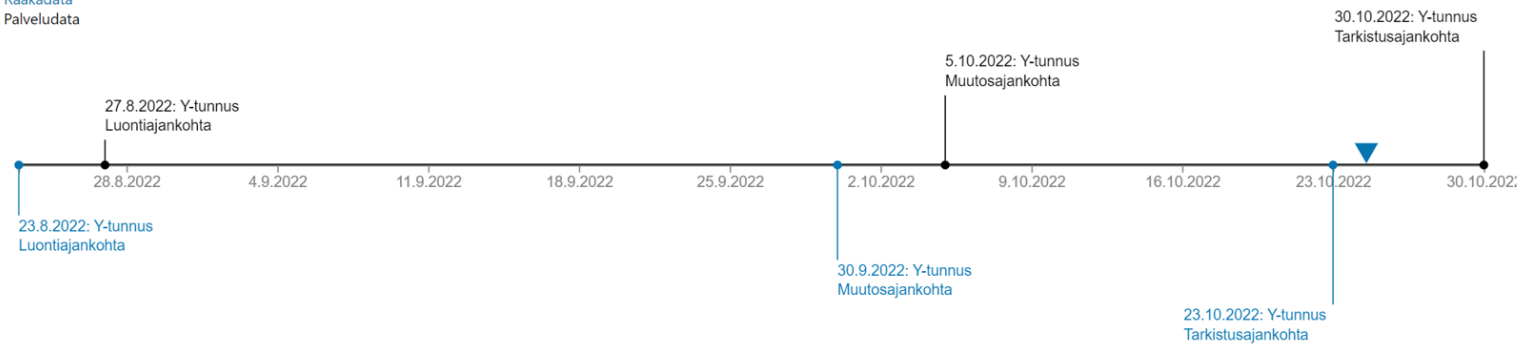
All

Ominaisuus

Y-tunnus

### Datan ajantasaisuus

- Raakadata
- Palveludata



Kuva 6. Esimerkki laatukriteerin ajantasaisuus visualisoinnista kukin ominaisuus erikseen tarkasteltuna.

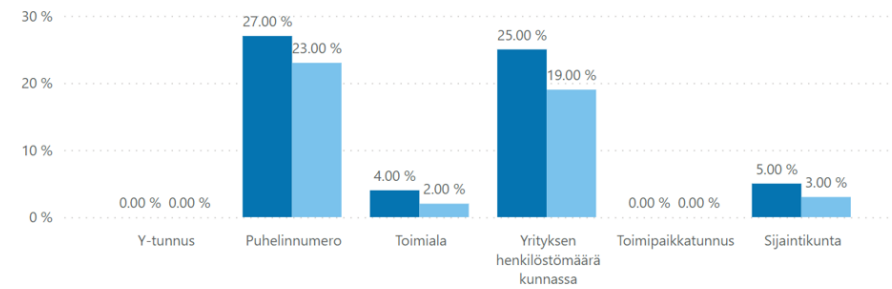


## Johdonmukaisuus ja virheettömyys

Laatukriteeri	Raakadata	Palveludata
<b>Johdonmukaisuus</b>		
<b>Tieto loogisuustarkastettu</b>	kyllä	kyllä
Y-tunnus	✓	✓
Puhelinnumero	✓	✓
Toimiala	✓	✓
Yrityksen henkilöstömäärä kunnassa	✓	✓
Toimipaikkatunnus	✓	✓
Sijaintikunta	✓	✓
<b>Virheettömyys</b>		
<b>Virheelliset ominaisuusarvot</b>	<b>8.00 %</b>	<b>4.00 %</b>
Y-tunnus	✓ 0.00 %	✓ 0.00 %
Puhelinnumero	✗ 27.00 %	✗ 23.00 %
Toimiala	⚠ 4.00 %	⚠ 2.00 %
Yrityksen henkilöstömäärä kunnassa	✗ 25.00 %	✗ 19.00 %
Toimipaikkatunnus	✓ 0.00 %	✓ 0.00 %
Sijaintikunta	✗ 5.00 %	⚠ 3.00 %
<b>Väärinluokittelu</b>	<b>8.00 %</b>	<b>4.00 %</b>
Y-tunnus	✓ 0.00 %	✓ 0.00 %
Puhelinnumero	ei relevantti	ei relevantti
Toimiala	✗ 5.00 %	⚠ 2.00 %
Yrityksen henkilöstömäärä kunnassa	ei relevantti	ei relevantti
Toimipaikkatunnus	✓ 0.00 %	✓ 0.00 %
Sijaintikunta	✗ 5.00 %	⚠ 3.00 %

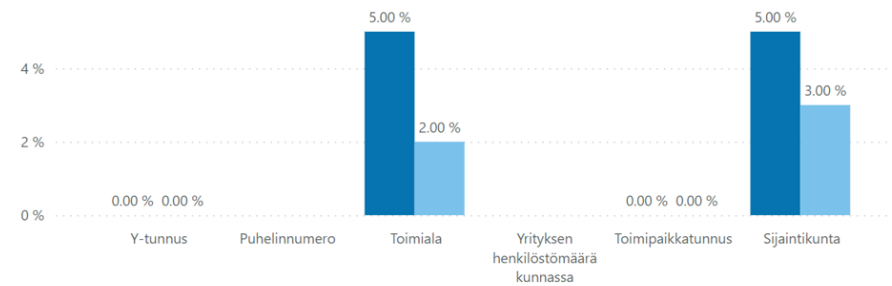
### Virheelliset ominaisuusarvot

Datan tyyppi ● Raakadata ● Palveludata



### Väärinluokittelu

Datan tyyppi ● Raakadata ● Palveludata



Kuva 7. Esimerkki johdonmukaisuuden ja virheettömyyden laatukriteerien visualisoinnista ominaisuuksittain tarkasteltuna.

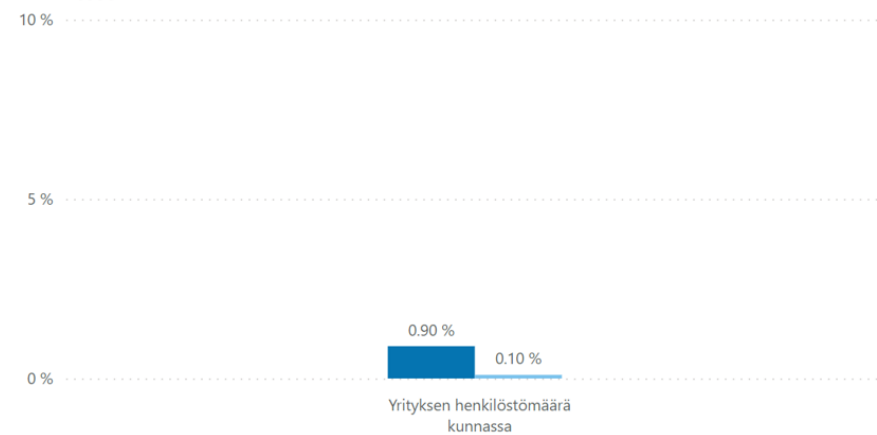
19.12.2022

## Tarkkuus

Laatukriteeri	Raakadata	Palveludata
<b>Tarkkuus</b>		
<b>Hajonta</b>		
Yrityksen henkilöstömäärä kunnassa	1500	1432
<b>Poikkeavat havainnot</b>		
Yrityksen henkilöstömäärä kunnassa	0.90 %	0.10 %

### Poikkeavien arvojen osuus

Datan tyyppi ● Raakadata ● Palveludata



Kuva 8. Esimerkki tarkkuuslaatukriteerin kuvaamisesta.

19.12.2022

## Suosituksenmukaisuus ja ymmärrettävyys

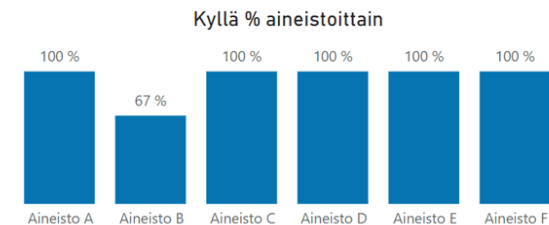
Aineisto

Select all Aineisto A Aineisto B Aineisto C Aineisto D Aineisto E Aineisto F

Kyllä/ei

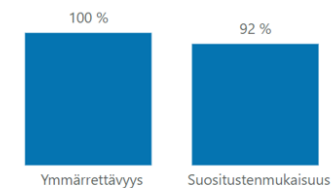
Select all ei kyllä

Laatukriteeri	Aineisto A	Aineisto B	Aineisto C	Aineisto D	Aineisto E	Aineisto F
<b>Suosituksenmukaisuus</b>						
Käytetyt standardit	✓ kyllä	✗ ei	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä
Standardienmukaisuus	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä
<b>Ymmärrettävyys</b>						
Ymmärrettävyyden asiakaspalaute	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä	✓ kyllä

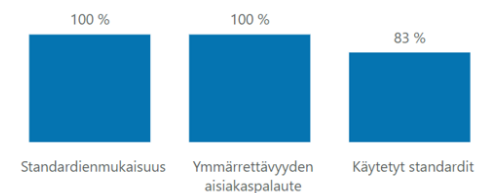


Laatukriteeri	Aineisto A	Aineisto B	Aineisto C	Aineisto D	Aineisto E	Aineisto F	Total
Suosituksenmukaisuus	100 %	50 %	100 %	100 %	100 %	100 %	92 %
Ymmärrettävyys	100 %	100 %	100 %	100 %	100 %	100 %	100 %
<b>Total</b>	<b>100 %</b>	<b>67 %</b>	<b>100 %</b>	<b>100 %</b>	<b>100 %</b>	<b>100 %</b>	<b>94 %</b>

### Kyllä % laatukriteereittäin



### Kyllä % mittareittain



Kuva 9. Esimerkki suosituksenmukaisuuden ja ymmärrettävyyden laatukriteerien visualisoinnista kuudessa eri aineistossa samanaikaisesti.

### Kieliversiot aineistossa

Aineisto

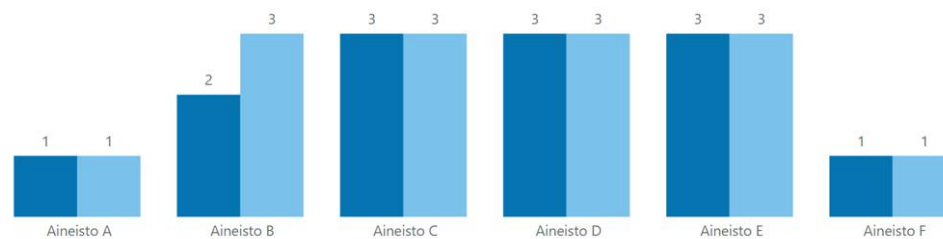
Kieli

### Kieliversiot, joilla aineisto on saatavilla

Mittarin nimi	Aineisto A	Aineisto B	Aineisto C	Aineisto D	Aineisto E	Aineisto F
Aineistokuvaus	Suomi	Suomi, Englanti	Suomi, Ruotsi, Englanti	Suomi, Ruotsi, Englanti	Suomi, Ruotsi, Englanti	Suomi
Käsitteiden määritelmät	Suomi	Suomi, Ruotsi, Englanti	Suomi, Ruotsi, Englanti	Suomi, Ruotsi, Englanti	Suomi, Ruotsi, Englanti	Suomi

### Kieliversioiden määrä eri aineistoissa

Mittarin nimi ● Aineistokuvaus ● Käsitteiden määritelmät



Kuva 10. Esimerkki aineistokuvausten ja käsitteiden määritelmien saatavuudesta eri kieliversioina.

19.12.2022

### Aineiston käyttöoikeudet

Aineisto

Select all Aineisto A Aineisto B Aineisto C Aineisto D Aineisto E Aineisto F

Käyttöoikeus

All

#### Aineiston käyttöoikeuden rajaukset

Mittarin nimi	Käyttöoikeus	Käytön rajoitukset		
Aineisto	Käyttöoikeuden rajaus	Sanallinen kuvaus	Käyttöoikeuden rajaus	Sanallinen kuvaus
Aineisto A	avoin	Julkaistu versio on avoin	ei rajattu	
Aineisto B	käyttölupa		rajattu	
Aineisto C	sopimus	Tietopalvelun tietoaineistolla on sekä maksullisia tietotoimituksia että viranomaistoimituksia, joten tietotoimitus pohjautuu yleensä sopimukseen.	rajattu	Tietotoimitussopimuksessa määritetään mihin käyttökohteisiin tietoaineistoa voi käyttää.
Aineisto D	julkinen	Summatason aineisto on julkinen, yksilötason aineistoon tarvitaan käyttölupa	ei rajausta	Summatason aineistossa ei rajoitusta, lähde mainittava
Aineisto E	julkinen	Julkinen tilasto ja tiedot vapaasti käytettävissä	julkinen	Julkinen tilasto ja tiedot vapaasti käytettävissä
Aineisto F	avoin		ei rajattu	

Kuva 11. Esimerkki aineiston käyttöoikeuksien kokoavasta kuvaamisesta.

### Aineiston koneluettavuus

Aineisto

Select all Aineisto A Aineisto B Aineisto C Aineisto D Aineisto E Aineisto F

Koneluettavuus (kyllä/ei)

Select all kyllä ei

#### Aineiston koneluettavuus

Aineisto	Kohdeyksikön pysyvä tunniste	Koneluettavuuden asiakaspalautte	Tietoaineiston tietomalli
Aineisto A	kyllä	kyllä	kyllä
Aineisto B	kyllä	kyllä	kyllä
Aineisto C	kyllä	kyllä	kyllä
Aineisto D	kyllä	kyllä	ei
Aineisto E	kyllä	kyllä	kyllä
Aineisto F	kyllä	ei	kyllä

Kuva 12. Esimerkki aineiston koneluettavuuden kokoavasta kuvaamisesta.

19.12.2022

## Aineiston täsmällisyys

Aineisto

Toimitusviive

All

### Aineiston täsmällisyys

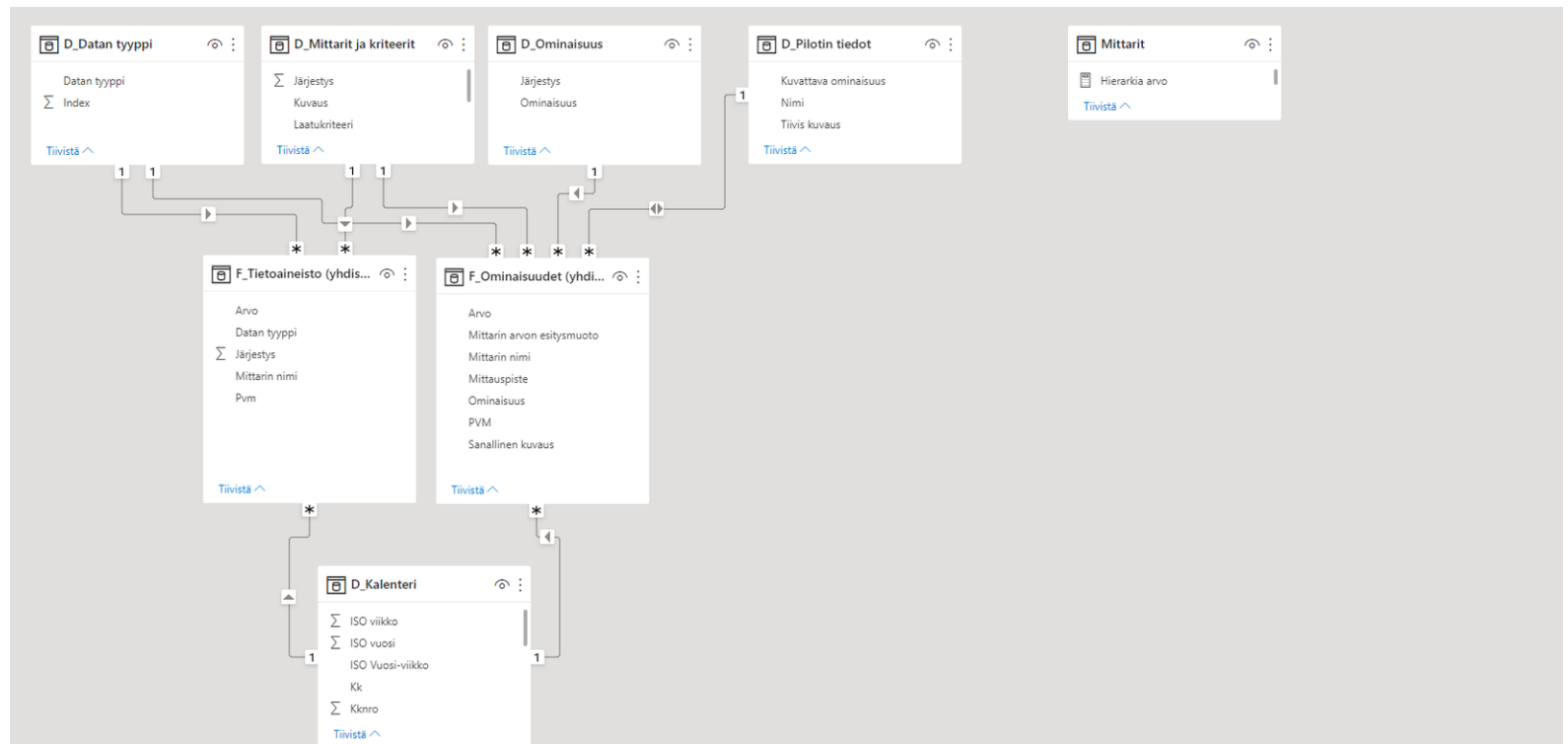
Mittarin nimi	Määräpäivien noudattaminen	Päivitystiheys	
Aineisto	Arvo	Sanallinen kuvaus	Arvo
		Sanallinen kuvaus	
Aineisto A	ei relevantti		jatkuva
			Perustietoja päivitetään jatkuvasti, lisätietoja vuosittain sekä muita tietoja noin 5–10 vuoden välein.
Aineisto B	ei relevantti		kerran vuodessa
Aineisto C	toimitusviive		kuukausittain
			Uusi tietoaineisto poimitaan/päivitetään kerran kuukaudessa.
Aineisto D	ei relevantti		kerran vuodessa
			Vuosiaineisto päivittyy kerran vuodessa
Aineisto E	ei relevantti		muu
			Ilmoitetaan julkistuskalenterissa, tilaston laatuselosteessa sekä tietokantataulujen tiedoissa.
Aineisto F	ei toimitusviivettä		muu

Päivitystiheys

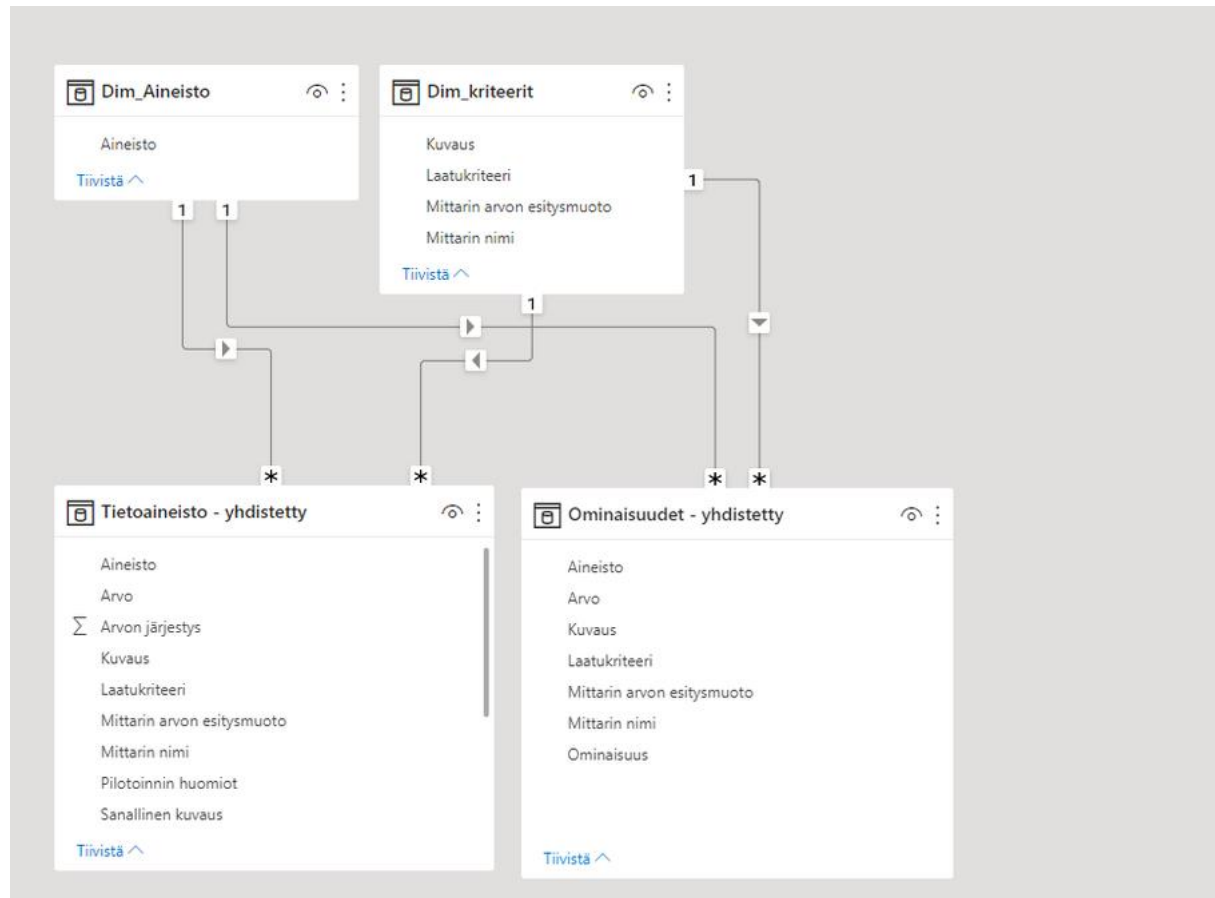
All

Kuva 13. Esimerkki aineiston täsmällisyyden kokoavasta kuvaamisesta.

Liite 2: Esimerkkejä Power BI:n tietomallista



Kuva 14. Miten tieto kuvaa todellisuutta? -laatuksiteeriryhmän kuvaamisessa käytetty tietomalli.



Kuva 15. Miten tieto on kuvattu? -laatuksiteeriryhmän kuvaamisessa käytetty tietomalli.