



**VTT**

# Matkatiedon laatu ja sen mittaaminen

*Olli Pihlajamaa (VTT)*

*Tuomo Kinnunen (Matkahuolto)*

Väylät & Liikenne 2023, Hämeenlinna, 22.3.2023

# Matkatiedon olennaiset elementit

## Reitit

liityntäpisteiden ketju, geometria, matkat (GTFS: routes, trips, stop\_times, shapes)

## Pysäkit

sijainti, pysäkin nimentä ja kuvaustiedot (GTFS: stops)

## Aikataulut

pysäkkiaikataulut, lähtöjen aikataulut  
(GTFS: calendar, calendar\_dates, stop\_times)

## Hinta- & maksutapatiedot

(GTFS: fare\_rules, fare\_attributes)

## Palvelun kuvaus & saatavuus

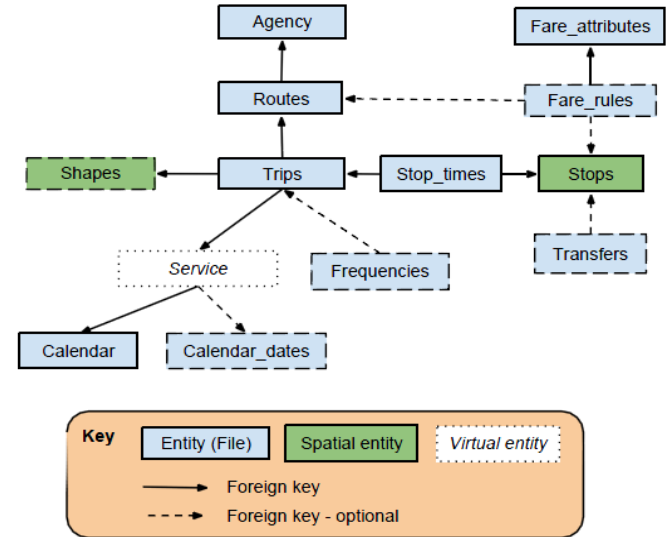
(NAP, GTFS: agency, calendar, calendar\_dates)

## Esteettömyys

pysäkit, kulkuneuvot, matkaketju (GTFS: trips, stops)

## Reaaliaikatieto

kulkuneuvojen sijaintitieto, pysäkkiennusteet ja poikkeustiedot  
vapaaehtoinen tieto – ”annettava, jos saatavissa”

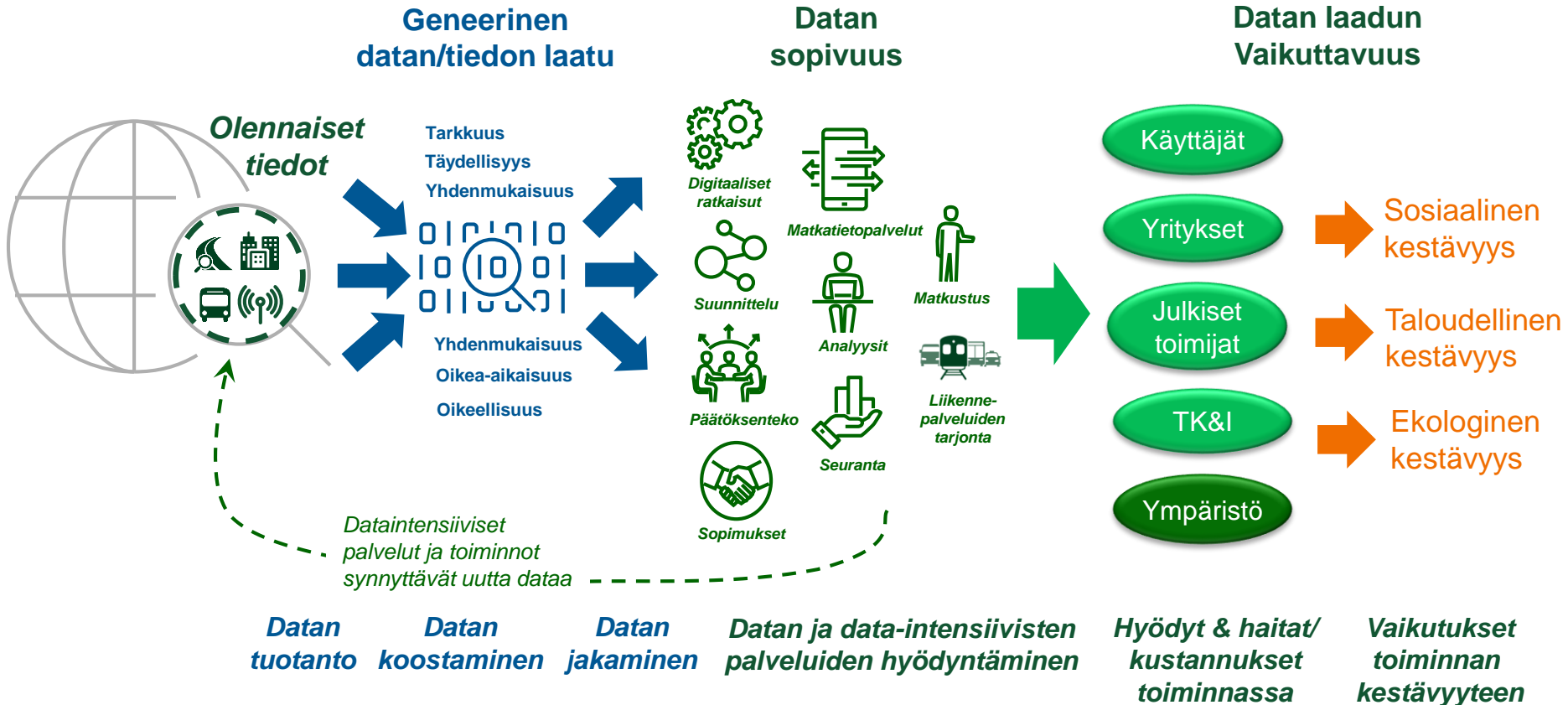


# Datan/tiedon geneerinen laatu

#	Datan laadun dimensio	Yleinen kuvaus
1	Täydellisyys (Completeness)	Tietojoukko antaa kaiken tarvittavan tiedon kuvattavasta kohteesta
2	Ainutlaatuisuus (Uniqueness)	Tietojoukossa ei esiinny sama tieto moneen kertaan (ei duplikaatteja)
3	Oikea-aikaisuus (Timeliness)	Tieto on ajantasaista ja oikea-aikaista käyttötarkoitukseen nähden
4	Oikeellisuus (Validity)	Tieto on virheetöntä ja syntaksin mukaista
5	Tarkkuus (Accuracy)	Tieto vastaa riittävän hyvin todellisuutta
6	Yhdenmukaisuus (Consistency)	Tieto on yhteismitallista, vertailukelpoista sekä yhteen toimivaa läpi tietojoukon

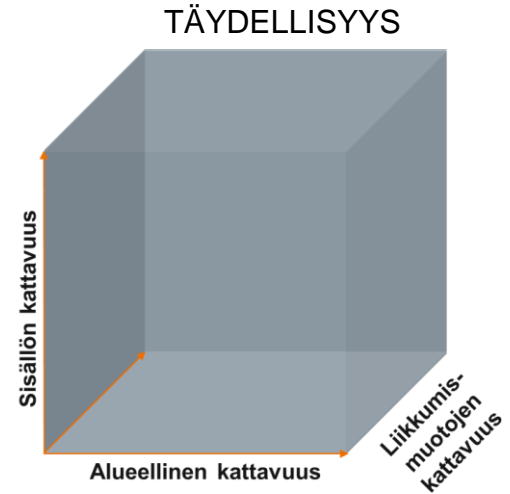
*Sovitettu matkatiedon kontekstiin lähteestä: DAMA (2013)*

# Matkatiedon laatu ja sen vaikutukset



# Esimerkkejä matkatiedon laatuongelmista

- **Täydellisyys:**
  - Kohdealueen liikkumispalveluita puuttuu
  - Osasta aineistoa puuttuu esteettömyystieto
- **Ainutlaatuisuus**
  - Eri järjestelmistä sama pysäkkitieto moneen kertaan
  - Saman linjan lähtötiedot osittain tai kokonaan duplikaattina
- **Oikea-aikaisuus**
  - Matkatieto vanhentunutta
  - Reaaliaikatieo ei ole ajantasaista
  - Matkatietoja ei saatavilla tarpeeksi pitkälle tulevaisuuteen
- **Oikeellisuus**
  - Matkatieto ei noudata sovittua standardia tai formaattia
- **Tarkkuus**
  - Pysäkkien sijainnissa virheitä tai reittigeometriat epätarkkoja
  - Hintatiedot liian ylimalkaisia
- **Yhdenmukaisuus**
  - Pysäkkien nimentäperiaatteet eivät noudata samaa ohjeistusta



# Matkatiedon laadun mittaristo

Alustava versio  
asiantuntijapalautteen pohjalta

# Mittareiden tärkeys – asiantuntijakysely

Olennaisten tietojen elementit	Laatu-kriteerit	Tarkkuus	Täydellisyys	Yhdenmukaisuus	Ainutlaatuisuus	Oikea-aikaisuus	Oikeellisuus
Reitit							
Pysäkit							
Aikataulut							
Hintatiedot							
Palvelun kuvaus & saatavuus							
Esteettömyys							
Dynaaminen tieto							

**Olennaisten tietojen elementeille pyrittiin löytämään mittareita eri datan laatukriteerien suhteen:**

- Matkatiedon laadun aspektit
- Esimerkkejä matkatiedon mittareista
- Mittarin yksikkö
- Mittarin tärkeys
- Perustelut & kommentit
- Pyyntö lisätä omia mittareita

# Mitattavien asioiden painotus asiantuntijapalautteessa

Olennaisten tietojen elementit \ Laatu-kriteeri	Tarkkuus	Täydellisyys	Yhden- mukaisuus	Ainut- laatuisuus	Oikea- aikaisuus	Oikeellisuus
Reiitit	2,0	0,7 EOS	0,8 EOS	4,0	4,0	2,0
Pysäkit	3,0	2,0 EOS	2,8	3,0	4,0	3,0
Aikataulut	2,5	1,5	3,0	2,0	2,3	2,0
Hintatiedot	0,3	1,0	1,0	1,0	1,0	2,0
Palvelun kuvaus & saatavuus	1,0	1,6	1,0	0,0	2,0	1,5
Esteettömyys	2,7	1,0	1,0	0,0	2,0	3,0
Dynaaminen tieto	1,0	0,7	1,0	1,0	0,0	1,0

Olennaisten tietojen elementit \ Laatu-kriteeri	Tarkkuus	Täydellisyys	Yhden- mukaisuus	Ainut- laatuisuus	Oikea- aikaisuus	Oikeellisuus
Reiitit	3,3	2,4 EOS	2,4 EOS	4,5	4,0	3,0
Pysäkit	4,0	2,4 EOS	3,5	3,7	4,0	3,5
Aikataulut	3,5	2,5	4,0	3,0	3,0	3,0
Hintatiedot	1,7	3,0	2,0	2,0	2,0	3,0
Palvelun kuvaus & saatavuus	2,5	2,6	2,0	0,0	3,0	1,5
Esteettömyys	3,3	2,0	2,0	0,0	3,0	3,0
Dynaaminen tieto	2,8	2,7	1,0	1,0	2,0	1,0

## Tarkastelu: ”Tärkeys” – ”Ei tärkeys”

- Tärkeä = +1, Ei tärkeä = -1, n = ~5
- Keskimääräinen ”Tärkeä” - ”Tärkeä” -erotus joka olennaisten tietojen elementille ja laatu-kriteerille
- Tässä tarkastelussa ”Ei tärkeä” -äänten määrä laskee tärkeyttä

## Tarkastelu: ”Tärkeys”

- Tärkeä = +1 n = ~5
- Keskimääräinen ”Tärkeä” -äänten summa joka olennaisten tietojen elementille ja laatu-kriteerille
- Tässä tarkastelussa huomioidaan vaan ”Tärkeä” -äänten osuus



# Alustavasti valitut matkatiedon mittarit (1)

## Reitit

- Reitin muotomallin (shape file) kuljetun matkan poikkeama reitin (ajoneuvon ajamasta) todellisesta pituudesta **Ta**
- Niiden lähtöjen määrä, joiden pysäkkilista on virheellinen **Ta** **O**
- Puuttuvien reitti-/lähtökuvausten määrä **Ta**
- Duplikaattien ja päällekkäisyyksien määrä reiteissä/lähdöissä **A**

## Pysäkit

- Raja-arvon ylittävien sijaintipoikkeamien määrä pysäkkiaineistossa **Ta** **O**
- Puuttuvien liityntäpisteiden määrä **Ta**
- Ohjeistuksen ja käytetyn standardin mukaisesti/vastaisesti nimettyjen pysäkkien määrä **Y**
- Duplikaattipysäkkien määrä pysäkkikuvausaineistossa **A**
- Samoilla nimillä esiintyvien eri pysäkkien määrä aineistossa (lkm/prosenttiosuus kaikista pysäkeistä) **A**

## Aikataulut

- Pysäkkiaikataulujen saapumis- ja lähtöaikojen keskimääräinen poikkeama keskimääräisistä toteutuneista saapumis- ja lähtöajoista **Ta**
- Puuttuvien aikataulujen määrä **Ta**
- Duplikaattien tai päällekkäisyyksien määrä lähdoissä **A**
- Lähdön peräkkäisten pysäkkien saapumis- ja lähtöaikojen kasvusääntöä rikkovien lähtöjen määrä **O**

- Ta** = Tarkkuus
- Ta** = Täydellisyys
- Y** = Yhdenmukaisuus
- A** = Ainutlaatuisuus
- Oa** = Oikea-aikaisuus
- O** = Oikeellisuus

# Alustavasti valitut matkatiedon mittarit (2)

## Hinta ja maksutapatiedot

- Puuttuvien hintatietojen määrä Tä

## Palvelun kuvaus ja saatavuus

- Puutteellisten palvelukuvaustietojen määrä FINAPI:ssa Tä
- Puutteellisten palvelun tarjoajan tietojen määrä matkatietoaineistossa Tä

## Esteettömyys

- Puuttuvien esteettömyystietojen määrä pysäkkitiedoissa Tä
- Puuttuvien esteettömyystietojen määrä kulkuneuvotiedoissa Tä

## Dynaaminen tieto

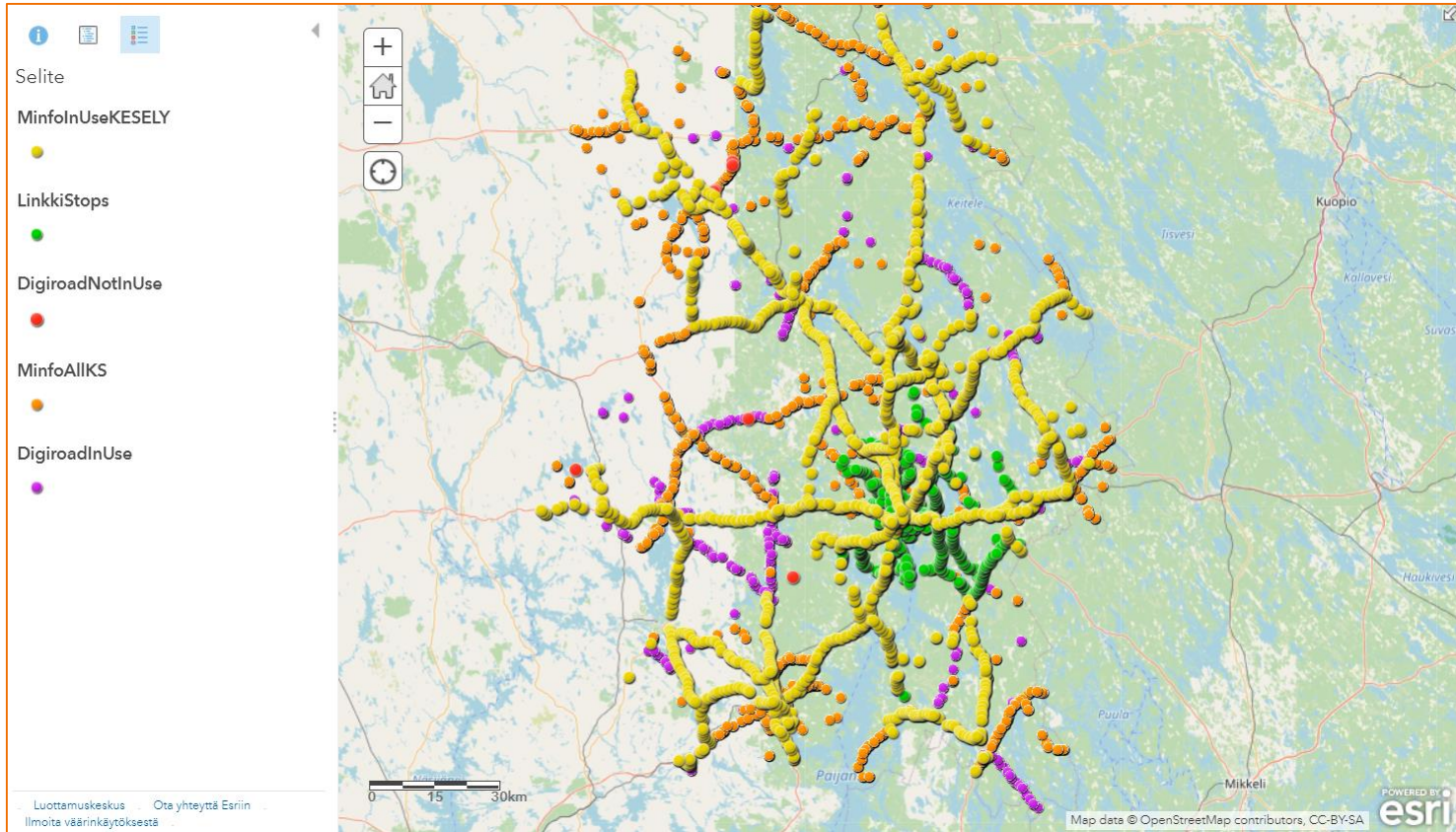
- Liikennevälineiden seurantatiedon keskimääräinen toteutunut näytteenotto-/raportointiväli Tä
- Reaaliaikaisen liikkumistiedon jakamisen piirissä olevien säännöllisen aikataulun mukaisen liikenteen lähtöjen määrä Tä

- Ta = Tarkkuus
- Tä = Täydellisyys
- Y = Yhdenmukaisuus
- A = Ainutlaatuisuus
- Oa = Oikea-aikaisuus
- O = Oikeellisuus

# Käytännön esimerkkejä:

## Keski-Suomen laatupalvelupilotti (*Tuomo Kinnunen, Matkahuolto*)

# Keski-Suomen pysäkkiaineisto kartalla

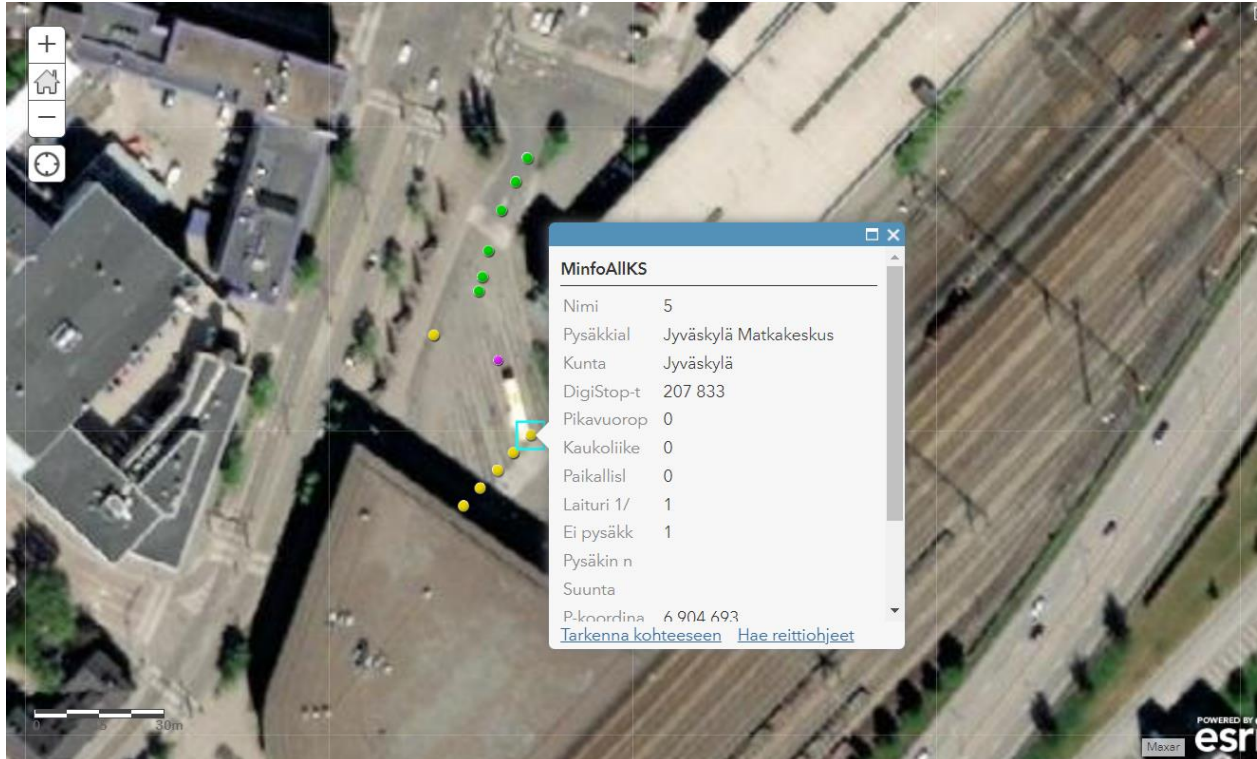


# Pysäkkien nimet eri tietoaineistoissa

<b>Pysäkkikannat: Matkahuolto vs. Digiroad</b>	<ul style="list-style-type: none"> <li>- Matkahuolto 5501, Digiroad 5792 pysäkkiä</li> <li>- 4649 pysäkillä Digistop ID vastaavuus Digiroadiin               <ul style="list-style-type: none"> <li>○ 2696 täsmälleen samaa pysäkin nimeä (63,9%)</li> <li>○ <b>1680 ainakin vähän eroavaa pysäkin nimeä (36,1%)</b></li> </ul> </li> </ul>
<b>Matkahuollon GTFS data vs. Digiroad</b>	<ul style="list-style-type: none"> <li>- 3051 pysäkkiä Matkahuollon GTFS datassa (16.2.2022)</li> <li>- 2852 pysäkin Digistop ID löytyy Digiroadista               <ul style="list-style-type: none"> <li>○ 1812 täsmälleen samaa pysäkin nimeä</li> <li>○ <b>1040 ainakin vähän eroavaa pysäkin nimeä</b></li> </ul> </li> </ul>
<b>Jyväskylä - Linkin GTFS data vs. Digiroad</b>	<ul style="list-style-type: none"> <li>- 2092 pysäkkiä Linkin GTFS datassa (16.2.2022)</li> <li>- 2014 pysäkin Stop ID löytyy Digiroadista               <ul style="list-style-type: none"> <li>○ 1925 täsmälleen samaa pysäkin nimeä</li> <li>○ <b>89 ainakin vähän eroavaa pysäkin nimeä</b></li> </ul> </li> </ul>

(Tuomo Kinnunen, Matkahuolto)

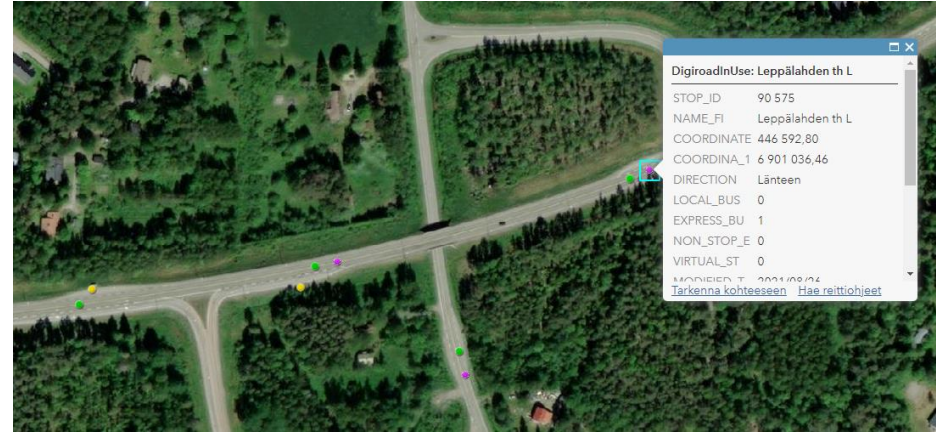
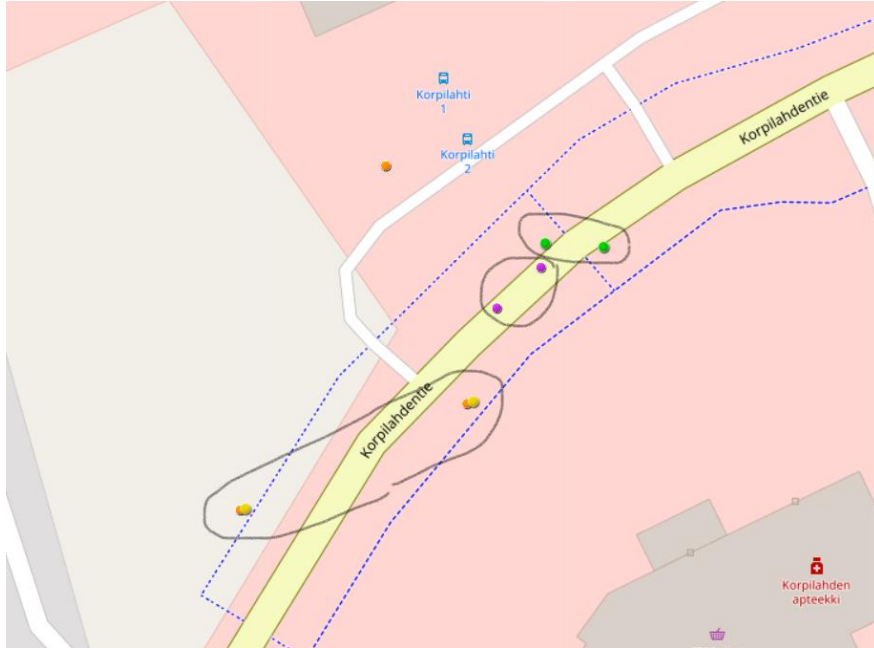
# Pysäkkien paikat: Digiroad ja pysäkin sidonta tien keskilinjaukselle



- Digiroad
- Matkahuolto

(Tuomo Kinnunen, Matkahuolto)

# Pysäkkien paikat: poikkeavuudet eri aineistoissa



- Jyväskylän liikenne (Linkki)
- Digiroad
- Matkahuolto

(Tuomo Kinnunen, Matkahuolto)

# Pysäkit: Eroavaisuudet eri aineistoissa

	Digiroad	Linkin liikennedata	Matkahuollon liikennedata
02/2022			
D-L-M	907	907	907
D-L	1108	1108	0
L-M	0	31	31
L	0	47	0
D-M	1935	0	1935
M	0	0	162
D	1860	0	0
<b>Yhteensä</b>	<b>5810</b>	<b>2093</b>	<b>3035</b>
<b>Käytössä</b>	3950		
	68,0 %		

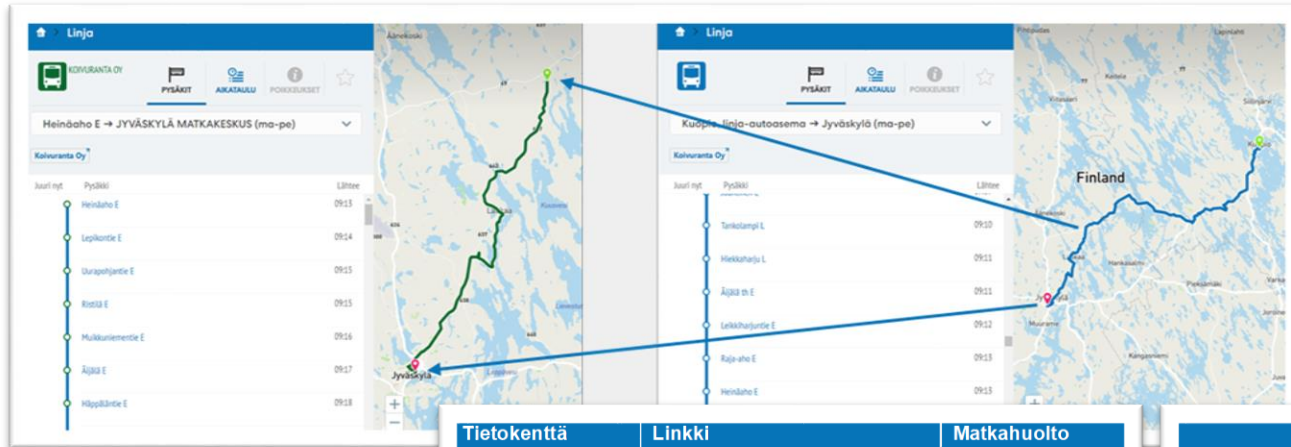
## 907 pysäkin otanta

- Sijainnin ero
  - 15 metriä keskiarvolta 11 metriä mediaani
  - Yli 50 metriä 27 pysäkillä
- Nimistön erot
  - 32 pysäkillä Digiroadin ja Linkin välillä
  - 197 pysäkillä Matkahuollon ja Linkin/Digiroadin välillä
    - Merkittävä ero nimissä löytyi noin 80 pysäkissä, joista 51 pysäkkiä käytiin tarkalla seulalla läpi.

(Tuomo Kinnunen, Matkahuolto)



# Esimerkki päällekkäisistä matkatiedoista



Tietokenttä	Linkki	Matkahuolto
Route_short_name	LAUKAA-JYVÄSKYLÄ	Kuopio - Konnevesi - Jyväskylä
Agency_ID	6743	235
Agency Name	Koivuranta Oy	Koivuranta Oy
Route_id	411	74479.235
Trip_id	SEUTU_M-P_talvi_411_1_091300_101500_0	74479.235
Arrival time (1.stop)	9:13 Heinäaho E	6:40 Kuopio linja-autoasema, laitur 8
Arrival time (last stop)	10:15 Jyväskylä linja-autoasema, tulolaituri	10:25 Keskussairaala 2

	Linkki	Matkahuolto
Reitti	LAUKAA-JYVÄSKYLÄ	Kuopio - Konnevesi - Jyväskylä
Pysäkkejä reitillä	81	219 (83)
Samoja pysäkkejä	79 / 85 (Linkillä yksi "väärän suunnan" pysäkki, eri ID)	
Täsmälleen sama pysähtymisaika	23 (29%)	
Eri pysähtymisaika	56 (71 %)	
Stop_id avaruus	Digiroad	Minfo

# Yhteenveto ja johtopäätökset

# Yhteenveto: matkatiedon laatu

- **Matkatietojen laatu** on monidimensioinen kenttä ja se on monin **paikoin vaikeasti mitattavaa**
- **Lainsäädännöstä vain raamit** vaadittavista matkatiedoista – Ohjeistus laadusta: *”Datan on oltava tarkkaa ja sen on oltava ajan tasalla.”*
- EU-lainsäädännössä **laatukriteerien määrittely on jätetty kansallisille toimivaltaisille viranomaisille** (MMTIS, Artikla 9, kohta 2a)
- Vain yksi EU-projektin julkaisu matkatiedon ehdotetuiksi laadun kriteereiksi
- Tässä selvityksessä **mittareita pyrittiin valitsemaan ja priorisoimaan** asiantuntijatyöpajan avulla ja samalla **keskittyen tiedossa oleviin käytännössä eniten haittaaviin laatuongelmiin**
- → **Kansallisten viranomaisten tulisi määritellä laatukriteerit sekä niiden mittarit matkatiedolle** – tämä selvitys ja Niculescu et al. (2019) auttavat työn suuntaamisessa

- **Lähtökohta** riittävän matkatiedon laadun takaamiselle on **mahdollistaa toimivat, ovelta-ovelle matkaketjut ja niitä tukevat digitaaliset matkatietopalvelut** kansalaisille
- Matkatiedon **laatuongelmat johtavat puutteisiin ja virheisiin matkatietoon perustuvissa palveluissa → seurauksia:**
  - → Laadukkaat matkatietopalvelun rakentaminen vaikeutuu tai estyy
  - → Matkustaja ei luota palveluun, jossa puutteita ja virheitä
  - → Matkustaja ei käytä joukkoliikennettä, koska ei löydä sopivaa kulkuyhteyttä
  - → Liikkumisen kulkumuotojakauman aiheuttamat kestävyysongelmat
  - → Liikennöitsijä, jolta ei tietoa tai virheellistä tietoa menettää asiakkaan
  - → Potentiaalisen matkaketjun kaikki liikennöitsijät menettävät asiakkaita matkaketjun osan puuttuessa
  - → Liiketoiminnan menetykset liikkumispalveluissa
  - → Liikkumispalveluiden tehottomuus
  - → Yhteiskunnan subvention tarpeen lisääntyminen
  - → Tehottomuuden aiheuttama kestävyden huonontuminen
  - → Eri toimialojen palveluiden saavutettavuuden huonontuminen
  - → Liikkumispalveluiden kokonaisvaltainen suunnittelu, seuranta ja analysointi vaikeutuu
  - → Eriarvoisuuden lisääntyminen niiltä osin, keneen matkatiedon puutteet vaikuttavat
- **Laadun vaikutukset** esim. kansalaisten palveluihin, kansalaisten toimintaan, liiketoimintaan sekä yleisemmin yhteiskunnan toimintoihin ja niiden kestävyteen syntyvät usein **monimutkaisten vaikutusketjujen ja –verkostojen kautta**
  - → Mittaamisen haasteet

# Yhteenveto: laadun ongelmien lähteitä

- Lainsäädännön toimeenpano **puutteellisella ja hajanaisella ohjeistuksella**
- **Puuttuva selkeä matkatiedon laadun kriteeristö**, jolla matkatiedon laadun valvontaa voidaan käytännössä tehdä
- Puutteellisesta matkatiedosta **sanktiointi puuttuu**
- Lukuisat tiedon tuottajat ja niiden **heterogeeniset yhteen sopimatonta matkatietoa tuottavat prosessit**
- **Päällekkäistä matkatietoa tuottavat toimijat ja prosessit**
  - Matkahuolto: ”1/3-laatua 3-kertaisella työllä”
- **Puutteelliset matkatiedon digitointityökalut, taidot ja resurssit**
- Eurooppalaisen standardoinnin mukaisen **tiedon tuottamisen tuen puute**
- **Kansallisen keskitetyn liityntäpistetietovarannon puuttuminen**
  - Digiroadin pysäkkirekisteri ei optimoitu matkatietoa vaan infran hallintaa varten
- **Kansallisen kaiken (perus)matkatiedon koontivarannon ja sen laadun hallinnan puuttuminen**

# Yhteenvedo: laadun parantamisen resepti

- **Yhtenäinen ja selkeä ohjeistus** matkatiedon digitaalisesta muodosta ja sisällöstä
- Matkatiedon **laadun kriteeristön kehittäminen** ja siihen perustuva **laadunvalvonta**
- Yhteinen **kansallinen liityntäpisterekisteri** (pysäkkirekisteri)
  - Multimodaalisen matkatiedon mahdollistava; ristiriidaton Digiroadin infra-tiedon kanssa
- **Yhteinen koontikanta**, johon kerätään kaikki kansallinen matkatieto perusosiltaan
- Toimintamalli, jossa **selkeä vastuutus ja resurssit jollekin toimijalle keskeisen matkatiedon koonnista ja sen häiriöttömästä jakamisesta** palveluiden ja toimintojen pohjaksi
- Yhtenäisen digitaalisen **matkatiedon tuottamisen välineet, palvelut ja tuki**
  - Saatavissa työkaluna itsedigitointiin tai puolestadigitoinnin palveluna
  - Digitointiratkaisujen tuottaman matkatiedon tulee läpäistä laadun kriteeristö (ks. yllä)
  - Eurooppalaisten standardien tuki
- **Viranomaisprosessien uudelleen organisointi, yhtenäistäminen sekä yhdistäminen liikenteenharjoittajien tiedontuotantoon**
- **NAP osaksi matkatiedon tuottamista ja koontia**

# bey<sup>0</sup>nd

## the obvious

Olli Pihlajamaa  
olli.pihlajamaa@vtt.fi