

Joukkoistetun paikkatiedon laatu ja luotettavuus

Turun kaupunkiseudun pyöräilyverkosto OpenStreetMap-karttapalvelussa

Mikko Lahtinen

Maantieteen
pro gradu -tutkielma
Laajuus: 40 op

Ohjaaja:
Niina Käyhkö

27.10.2023

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Pääaine: Maantiede

Tekijä: Mikko Lahtinen

Otsikko: Joukkoistetun paikkatiedon laatu ja luotettavuus -

Turun kaupunkiseudun pyöräilyverkosto OpenStreetMap-karttapalvelussa

Ohjaaja: Niina Käyhkö

Sivumäärä: 130 sivua

Päivämäärä: 27.10.2023

Kansalaislähtöinen joukkoistettu paikkatieto on noussut varteenotettavaksi tietolähteeksi viranomaisten ja kaupallisen sektorin tuottaman auktoritatiivisen paikkatiedon rinnalle. Tuotantotavat poikkeavat toisistaan merkittävästi, mikä ilmenee joukkoistetun lopputuotteen laatuvaihteluna. Sitä aiheuttavat tavoitteiltaan ja taitotasoltaan kirjava kartoittajajoukko, tuotantoalustojen ominaisuudet ja avoin tuotantotapa, joka ei pakota vaatimustenmukaisuutta. Jotta joukkoistettua paikkatietoa voidaan luotettavasti hyödyntää auktoritatiivisessa yhteydessä, on ymmärrettävä sen laadullisia piirteitä. Niitä koskevaa kotimaista tutkimustietoa on vähän ja tiedon tarve on uusiutuva joukkoistetun paikkatiedon jatkuvasti päivittyessä ja rikastuessa teemoiltaan. Tutkielmassa tähän tarpeeseen pyrittiin vastaamaan vertaamalla joukkoistetun OpenStreetMap-karttapalvelun pyöräilyverkostotietoa Turun kaupungin ja sen ympäryskuntien ylläpitämän paikkatietoaineiston laatuvaatimuksiin. Laatu mitattiin ISO 19157:2013-standardin mukaisesti täydellisyyden, sijaintitarkkuuden ja topologisen eheyden laatuparametreille. Niitä määristettiin verkoston kokonaispituudella, keskimääräisellä etäisyydellä ja geometrisella yhdenmukaisuudella sekä kappalemääräisellä laskennalla. Alueellista vaihtelua arvioitiin visuaalisesti, keskihajonnalla ja tilastollisesti Anselinin paikallisella Moranin I -tunnusluvulla. Koska paikkatietoaineiston kokonaislaadun muodostaa kvantitatiivisen laadun lisäksi sopivuus käyttötarkoitukseen ja lisäksi kokonaislaadun molempien komponenttien yhteisvaikutus luotettavuuteen on aiemmassa tutkimuksessa jäänyt vähälle huomiolle, aineiston käytettävyyttä arvioitiin pyöräilyverkoston kartoitustarpeiden näkökulmasta laatuarvioon perustuen. Luotaava käytettävyyssmittaus painotti pyöräilyverkoston kartoitustarpeille kriittisiä laatuparametreja käytettävyyttä ilmaisevassa aggregaatissa. Joukkoistetun paikkatiedon laatuun vaikuttavat tekijät ovat osin tilasidonnaisia. Tätä tutkittiin laatu- ja käytettävyyssarvioiden tulosten alueellisen vaihtelun arviolla. OpenStreetMap-aineiston täydellisyyden ja sijaintitarkkuuden havaittiin olevan lähellä laatuvaatimusta ja osin täyttävän sen. Topologinen eheys osoittautui sen sijaan heikoksi. Käytettävyyssmittaus puolestaan osoitti virheettömyyden vaatimuksen korostuvan sovellustarpeiden monimutkaistuesssa. Sama aineisto voi soveltua hyvin yhteen käyttötarkoitukseen, mutta olla kelvoton toiseen. Kvantitatiivisen laadun ja käytettävyyden suuri alueellinen vaihtelu heikensi selvästi aineiston luotettavuutta. Vaikka laajoja yhtenäisiä hyvätasoisia alueita paikallisesti havaittiin, aineistotason luotettavuus oli välttävä. Tulokset osoittavat joukkoistetun paikkatiedon lunastavan osan sen hyödyntämiseen liittyvistä odotuksista, mutta sen luotettavuutta on vaikea ennakoida. Tämä viittaa laadunvarmistuksen jatkuvaan välttämättömyyteen sekä tarpeeseen asettaa laatumittauksen tavoitteet tapauskohtaisesti. Tällöin myös laatuparametreiltään tai alueellisesti osittain heikkolaatuista aineistoa voidaan käyttää sovelluskohteiden niin salliessa ja joukkoistetun paikkatiedon temaattisesta moninaisuudesta päästään erityyppisissä sovellusyhteyksissä hyötymään.

Avainsanat: joukkoistettu paikkatieto, VGI, paikkatietoinfrastruktuuri, laatuarvio, käytettävyyys, ISO 19157:2013, GIS

Master's thesis

Subject: Geography

Author: Mikko Lahtinen

Title: The quality and reliability of crowdsourced spatial data -
Cycling network of the Turku urban area in the OpenStreetMap service

Supervisor: Niina Käyhkö

Number of pages: 130 pages

Date: 27.10.2023

Citizen-sourced volunteered geographic information (VGI) has emerged as a significant data source alongside authoritative spatial data produced by governmental agencies and the commercial sector. The modes of production differ substantially, resulting in variable quality in the crowdsourced final product. This variability stems from a diverse group of mappers in terms of objectives and skill levels, platform attributes, and an open production method that does not enforce conformity to standards. To reliably utilize crowdsourced spatial data in authoritative contexts, one must comprehend its quality attributes. There's a paucity of domestic research on this topic, and the need for information is recurrent, with VGI continually updating and diversifying in themes. This study aimed to address this gap by comparing the cycling network data from the crowdsourced OpenStreetMap service to the quality standards of spatial datasets maintained by the city of Turku and its surrounding municipalities. Quality was assessed based on the ISO 19157:2013 standard, focusing on completeness, positional accuracy, and topological integrity quality parameters. These were quantified by the total network length, average distance, geometric consistency, and count-based calculations. The regional variability was assessed both visually and statistically, using standard deviation and Anselin's local Moran's I index. Given that the overall quality of spatial datasets is not only defined by quantitative quality but also its suitability for intended use, and considering previous research has often overlooked the combined impact of these components on reliability, the usability of the dataset was evaluated from the perspective of cycling network mapping needs based on the quality assessment. This exploratory usability assessment emphasized quality parameters critical to cycling network mapping in an aggregate indicating usability. Factors influencing the quality of crowdsourced spatial data are partly location dependent. This was examined by analyzing the regional variability in the results of quality and usability assessments. The completeness and positional accuracy of the OpenStreetMap dataset were found to be close to the quality requirements and in some cases meeting them. However, topological integrity was notably weak. The usability assessment indicated an increased emphasis on the requirement for accuracy as application needs become more complex. The same dataset might be well-suited for one purpose but unsuitable for another. The significant regional variability in quantitative quality and usability markedly reduced the dataset's reliability. Even though extensive, consistently high-quality areas were observed locally, the dataset's overall reliability was found to be subpar. The results suggest that crowdsourced spatial data meets some of the expectations related to its utilization but remains a volatile data source in terms of reliability. This underscores the continuous necessity for quality assurance and the need to set quality assessment objectives on a case-by-case basis. This approach allows for the use of datasets that may be subpar in quality parameters or regional attributes, given the application permits, capitalizing on the thematic diversity of crowdsourced spatial data in various application contexts.

Key words: crowdsourced spatial data, VGI (Volunteered Geographic Information), spatial data infrastructure, quality assessment, usability, ISO 19157:2013, GIS

Sisällysluettelo

1	Johdanto	7
2	Tutkimuksen tausta ja teoreettinen viitekehys	10
2.1	Joukkoistettu paikkatieto	10
2.1.1	Kulttuurinen tausta	10
2.1.2	Tyypit ja teemat	13
2.1.3	OpenStreetMap	16
2.1.4	Käyttäjät / kartoittajat	17
2.1.5	Kartoitus	19
2.2	Joukkoistetun ja auktoritatiivisen paikkatiedon kytkennät	20
2.2.1	INSPIRE-paikkatietoinfrastruktuurin ja OpenStreetMapin vertailu	20
2.2.2	Joukkoistetun paikkatiedon hyödyntäminen	23
2.2.3	Hyödyntäminen Suomessa	24
2.2.4	Joukkoistettu paikkatieto ja paikkatietoinfrastruktuurien muutos	25
2.3	Paikkatiedon laatu ja luotettavuus	26
2.3.1	Epävarmuus	27
2.3.2	Virheet	28
2.3.3	Laadun osatekijät	29
2.3.4	Paikkatiedon laatustandardi ISO 19157	31
2.3.5	Sisäinen ja ulkoinen laatu	33
2.3.6	Paikkatiedon laadunhallinta	34
2.4	Laadun arviointi	36
2.4.1	Menetelmien esivaatimukset	41
2.4.2	Täydellisyys	42
2.4.3	Sijaintitarkkuus	45
2.4.4	Looginen eheys	51
2.4.5	Käytettävyys	53
3	Aineistot ja menetelmät	55
3.1	Tutkimusalue ja paikkatietoaineistojen valinta	55
3.2	Tutkimuksen lähestymistapa	59
3.2.1	Laatuparametrien valinta	61
3.2.2	Laatumittarien valinta	62
3.3	Aineistojen esikäsittely	64
3.3.1	Vertailtavien paikkatietokohteiden valinta	65
3.3.2	Digiroad-ylikulutiedon valmistelu	66
3.3.3	Pää- ja lähiverkoston yhdistäminen vertailuaineistossa	66

3.3.4	Vertailuaineistosta poikkeavan testiaineiston erottaminen	66
3.3.5	Spatiaalisen painottuneisuuden analyysi	68
3.4	Täydellisyyden mittaus	69
3.5	Ulkoisen sijaintitarkkuuden mittaus	70
3.6	Topologisen eheyden mittaus	74
3.7	Laadun ja käytettävyyden arviointi	77
4	Tulokset	80
4.1	Aineiston esikäsittely	80
4.1.1	Optimaalisen puskurisäteen pituus	80
4.1.2	Spatiaalinen painottuneisuus	80
4.2	OSM-aineiston laatu ja sen alueellinen vaihtelu	82
4.2.1	Täydellisyys	82
4.2.2	Sijaintitarkkuus	85
4.2.3	Topologinen eheys	91
4.2.4	Vertailu laatuvaatimukseen	93
4.3	OSM-aineiston käytettävyys ja sen alueellinen vaihtelu	94
4.3.1	Käytettävyys peruskartoitukseen	95
4.3.2	Reititys- ja navigointikäyttö	95
4.3.3	Käytettävyys uuden tiedon lähteenä verkostosuunnittelussa	96
5	Tulosten tarkastelu	98
5.1	Kokonaislaadun osatekijät ja aiempi tutkimus	98
5.2	Menetelmälliset haasteet ja riskit	104
5.3	Jatkotutkimustarpeet	108
	Kiitokset	111
	Lähteet	112

1 Johdanto

Geospatiaalisen tiedon taloudellista arvoa on vaikea yliarvioida. The Economist julisti vuonna 2017, ettei maailman arvokkain resurssi ole enää öljy, vaan data (The worlds... 2017). Tuona vuonna talousalan konsultointipalvelu alphaBeta (2017) arvioi paikkatietopalvelujen tuottavan maailmanlaajuisesti 400 mrd. dollaria sekä säästävän aikaa ja polttoainetta 550 mrd. dollarin arvosta. Pelkästään GPS paransi tuottoja ja laski kustannuksia viidellä prosentilla sektoreilla, jotka muodostivat 70 % globaalista bruttokansantuotteesta. Jopa 78 % arkipäiväisesti käyttämästämme informaatiosta on luonteeltaan geospatiaalista (Hahmann & Burghardt 2013).

Taloudellista etua luovaa tietoperusteista päätöksentekoa tukee laadukas paikkatieto. Ollakseen laadukasta, on sen kuvattava todellisuuden ilmiöiden suhteita, sijaintia ja ominaisuuksia täsmällisesti ja loogisesti tietomalliaan seuraten (Aalders 2002; Van Oort 2006: 4–6). Täyttääkseen nämä vaatimukset, on paikkatiedon tuottajan huomioitava maantieteellisen tiedon perusluonne, teknisten ja inhimillisten virhelähteiden vaikutus sekä tuotetun tiedon soveltuvuus vaihteleviin käyttötarkoituksiin (Azouzi 2000; Fisher ym. 2006; Frank 2007; Shi 2010: 15–26; Longley ym. 2015).

Viimeaikainen tekninen kehitys on mahdollistanut kansalaislähtöisen paikkatiedon kasvamisen hallinnon, tutkimuksen ja liike-elämän hyödyntämäksi tietovarannoksi (Zhang 2021). Tietoa louhitaan massadatana kansalaisten digitaalisista jäljistä tai sitä kerätään aktiivisella yhteistyöllä määrättyyn käyttötarkoitukseen (Kitchin & McArdle 2016; See ym. 2017). Paikkatietotutkimuksen uranuurtaja Michael Goodchild (2007) nimesi tämän tiedon vapaaehtoisesti kerätyksi paikkatiedoksi (volunteered geographic information, VGI). Hän määritteli sen vapaaehtoisten amatöörien yhteistoiminnassa keräämäksi maantieteelliseksi informaatioksi paikallisesta ympäristöstään. Tässä tutkielmassa sitä tarkastellaan termillä joukkoistettu paikkatieto.

Joukkoistuksen ytimessä on digitalisaation ja yhteiskunnan suhdetta tarkastelleen filosofi Pierre Lévy'n (1999) käsittelemä ajatus yhteistä sosiaalista tietoa tuottavasta joukkojen kollektiivisesta älykkyydestä, joka nousee yksilöiden yhteisistä ponnistuksista ja kilpailusta, ilmeten konsensuspäättökseen. Ohjelmistokehittäjä Eric Raymond (1999) esitti tämän tyypisessä hajautetussa yhteistoiminnassa syntyvän tuloksen jalostuvan lopulta korkealaatuisemmaksi kuin keskitetysti tuotettu vastineensa. Avointa ohjelmistokehitystä

kuvatessaan, hän luonnehti joukkoistettua mallia toimintamuotona basaariksi. Perinteinen keskitetty ja hierarkkinen, standardien ja ennalta määriteltyjen käytänteiden ohjaama tuotanto puolestaan näyttäytyi hänelle prosessuaalisessa tinkimättömyydessään katedraalina.

Kansalliset paikkatieto-organisaatiot ja kaupallinen sektori ovat tunnistaneeet joukkoistetun paikkatiedon edut, mutta käyttöönnotot ovat viimeaikoihin saakka olleet haasteellisia (Olteanu-Raimond 2017; Karlsson 2018). Joukkoistamalla paikkatietotuotantoa on pyritty esimerkiksi pienentämään viranomaistietokantojen ajantasaisuuden kustannuksia, luotu karttapohjia erilaisiin palveluihin sekä rakennettu navigointisovelluksia. Valtionhallinnossa on tutkittu joukkoistetun ja auktoritatiivisen, eli viranomaisten tai muun luotettavana pidetyn toimijan tuottaman paikkatiedon kaksisuuntaista hyödynnettävyyttä sekä yhteisiä tiedontuotannon prosesseja (Tietoaineistojen... 2016).

Keskeisin syy integroinnin vaikeuksiin on joukkoistetun paikkatiedon laatuvaihtelu, joka aiheutuu joukkoistetun ja auktoritatiivisen tuotantotavan perustavaa laatua olevista eroista (Minghini ym. 2019). Toisin kuin auktoritatiivinen tieto, joka syntyy vakioitujen toimintaperiaatteiden mukaisesti (See ym. 2016; Ma ym. 2015), joukkoistetun tiedon tuotannossa vaikuttaa lukuisa määrä epävarmuustekijöitä, mikä tekee sen integroimisesta haastavaa (Roper 2023). Tämä on ongelma esimerkiksi paikkatietoinfrastruktuureissa, jotka edellyttävät tekniikan, tuotantoprosessien ja tietomallien yhteentoimivuutta (Bordogna ym. 2015).

Joukkoistetun paikkatiedon laatuongelmat ovat herättäneet kysymyksen Raymondin (1999) kuvaston hengessä: lasketaanko sitä hyödynnettäessä amatöörit sisään temppeliin (McCullagh & Jackson, 2013)? Toisin sanoen, voidaanko joukkojen kykyyn täyttää korkeat auktoritatiiviset tuotantoarvot luottaa? Tämän selvittämiseksi tutkimuskohteeksi rajataan joukkoistetun OpenStreetMap-karttapalvelun (OSM) aineistosta Turun kaupunkiseudun pyöräilyverkostoteema, jota verrataan kaupunkiseudun kuntayhtymän yhteiseen pyöräilyverkostomalliin. Tutkimusaiheen valintaa perustelevat paikkatiedon jatkuvasti kasvava merkitys datataloudessa ja tähän liittyvä pyrkimys käyttää joukkoistettua paikkatietoa hyödykkeenä (ERT position... 2017; Karlsson 2018; Kansallinen... 2021). Tieteellisesti tutkielmaa motivoi kotimaisessa kontekstissa vallitseva tiedon puute. Laatuun ja käytettävyyteen liittyvää vertaisarvioitua tutkimuskirjallisuutta joukkoistetusta paikkatiedosta, erityisesti OpenStreetMap-palvelusta, ei ole julkaistu.

Valittu paikkatietoteema puolestaan edustaa tyypillistä ihmisten arjen tarpeisiin paikallisista lähtökohdista vastaavaa kartoitustuotetta (See ym. 2017). Tämä korostuu, sillä joukkoistetun karttapaikkatiedon laadun katsotaan nojaavan osaksi paikallisen kartoittajayhteisön tietoon elinympäristöstään (Goodchild 2007, Fritz ym. 2017). Teemavalinta on myös ajankohtainen, sillä julkisen hallinnon tavoite luoda valtakunnallinen pyöräilyverkostomalli, sisältää joukkoistetun paikkatiedon hyödyntämiseen tähtääviä toimenpiteitä (Tietoaineistojen... 2016; Laitinen ym. 2015).

Tätä taustaa vasten tutkielman päätavoitteena on tutkia joukkoistetun paikkatiedon luotettavuutta, laatua ja soveltavaa käyttöä seuraavien tutkimuskysymysten avulla.

1) Kuinka laadukasta OSM-aineisto on ISO 19157-standardin avulla arvioituna? Tätä perustelee ISO 19157 (2013) -standardin vakiintunut asema laadun ontologiana paikkatietoalalla ja joukkoistetun paikkatiedon vertailevassa laatututkimuksessa (Paikkatietoalan... 2023; Antoniou & Skopeliti 2015). Laatuarviossa verrataan OSM-aineiston laatuparametreja auktoritatiiviseen vertailuaineistoon, jonka laatuvaatimuksia vastaan tehdään myös OSM-aineiston suorituskykymittaus täydellisyyden, sijaintitarkkuuden ja topologisen eheyden osa-alueilla.

2) Miten OpenStreetMap-aineiston laatu ja käytettävyys vaihtelevat alueellisesti?

Joukkoistetulle paikkatiedolle leimallinen tuotantoprosessista sekä tuottajien ja tuotantoalustojen ominaisuuksista johtuva lopputuotteen laatuvaihtelu, on usein yhteydessä sijaintiin ja paikkaan (Moradi ym. 2021). Tutkimuskysymys valottaa tämän tilasidonnaisen heterogeenisyyden vaikutusta kokonaislaatuun.

3) Kuinka käytettävää OSM-aineisto on pyöräilyverkon kartoitussovelluksissa?

Paikkatiedon kokonaislaadun muodostaa tuottajan asettamien laatuvaatimusten lisäksi sopivuus käyttötarkoitukseen (Devillers & Jeansoulin 2006: 35–39; Frank 2007). Tällöin tuottajan näkökulmasta määritelty universaali laatu voi saada vaihtelevia merkityksiä eri sovellusyhteyksissä.

2 Tutkimuksen tausta ja teoreettinen viitekehys

2.1 Joukkoistettu paikkatieto

Joukkoistettu paikkatieto viittaa georeferoituun tietoon, jonka lähteenä on yksityinen kansalainen. Sen ilmenemismuotojen vaihtelevia tuotanto- ja sovellustapoja kuvaavat yleisesti seuraavat piirteet.

Kansalaiset toimivat tuottajan ja käyttäjän kaksoisroolissa luoden runsaasti luonnon- ja sosiaalisen ympäristön ilmiöitä kuvaavaa informaatiota. Tuotantotavoille ovat tyypillisiä yhtäältä perinteistä asiantuntijatuotantoa vapaamuotoisemmat tuotantoprosessit, kuten laadunhallinta, mutta toisaalta edullisemmat tuotantokustannukset sekä aineiston suuri sovelluspotentiaali. Kansalaislähtöisen georeferoidun tiedon kuvaamat kohteet voivat poiketa asiantuntijatuotannon tyypillisistä teemoista. Lisäksi tuottajakunnan tietotaito vaihtelee ja osallistumisen motiivit ovat moninaiset. Näiden piirteiden yhteisvaikutus saa aikaan laatuvaihtelua, joka asettaa tiedon luotettavuuden sen hyödyntämisympyröiden kynnyskysymykseksi.

2.1.1 Kulttuurinen tausta

Joukkoistettu paikkatieto sijoittuu digitalisaation laajempaan kontekstiin osana *käyttäjän tuottaman tiedon* tekniskulttuurista ilmiötä (Goodchild 2007), jonka syntymiselle kahden viime vuosikymmenen tietotekninen kehitys on luonut edellytykset. Ilmiön mahdollistajina ovat olleet Web 2.0 -tekniikat, mobiililaitteet, paikannuspalvelut, yhteistoiminnalliset verkkoyhteisöt, avoin kehitysmalli ja avoin data (Capineri 2016).

Digitalisaatio on muuttanut myös maantieteen tutkimuksen kohteita ja lähestymistapoja huomattavalla tavalla, mikä sallii puhumisen sen piirissä tapahtuneesta “digitaalisesta käännöksestä” (Ash ym. 2018). Provokatiivisimmillaan käyttäjän tuottama massadata on nähty paradigmasiirtymän mahdollistavana muutostekijänä, jonka voi antaa puhua puolestaan ilman sitä kehystävää teoriaa (Anderson 2008). Näkemykseen on vastattu, ettei datasta muodostettu tieto ole vain faktoja ilman asiayhteyttä, vaan niistä tehtyjen tulkintojen on nojattava teoreettiseen taustaansa (Bryant & Raja 2014), jotta spatio-temporaaliset narratiivit saadaan esiin (Kitchin 2014).

Käyttäjän tuottamaan tietoon kiinnittyy nousevia yhteiskunnallisia trendejä, kuten yhteistoiminnallisuus, osallistaminen ja jakamistalous (Rifkin 2014; Taeihagh 2017). Ne

vaikuttavat osatekijöinä Lévy'n (1999) kuvaamassa *kollektiivisessa älykkyydessä*, jossa yksityinen ja kollektiivinen asemoituvat yhteiskunnallisissa päätöksentekoprosesseissa toisiinsa nähden uudelleen (Capineri 2016). Verkottunutta tietoyhteiskuntaa tutkinut sosiologi Manuel Castells (2009: 21–25) näkee tähän kehitykseen liittyvät yksilöiden ja yhteisön lomittuneet verkostot perustaksi, jolta yksilön sosiaalistumisen on mahdollista alati teknistyvässä kulttuurissa tapahtua.

Verkostoista nousevaa kollektiivista älykkyyttä ilmentää joukkoistetussa paikkatietotuotannossa avoin kehitysmalli. Sen keskeinen ajatus on monien kehittäjien vapaa ja hajautettu osanotto yhteiseen projektiin, jossa osallistujien kokonaispanos tuottaa laadukkaan lopputuloksen (Haklay ym. 2010; Raymond 1999). Joukkoistetun paikkatiedon suosion myötä paikkatietotuotannon painopisteen on nähty siirtyneen keskitetystä auktoritatiivisesta ympäristöstä kohti yksilötasolle hajautettua tilaa (Capineri 2016).

Lukuisten osallistujien suorittaman laatuvalidoinnin lisäksi joukkojen kollektiivista älykkyyttä on sen jäsenten paikallistuntemus omasta elinympäristöstään (Mooney & Minghini 2017; Fritz ym. 2017). Sen on havaittu vahvistavan tuotetun tiedon oikeellisuutta joukkoistetuissa kartoitusprojekteissa (Haklay ym. 2010). Havainto linjautuu maantieteilijä Waldo Toblerin (1970) niin kutsuttuun Maantieteen ensimmäiseen lakiin, jonka mukaan maantieteessä kaikki asiat ovat riippuvaisia toisistaan, mutta toisiaan lähempänä olevien asioiden riippuvuus on kauempana olevia vahvempaa.

Laatua parantavan kollektiivisen älykkyyden läsnäolo joukkoistetuissa projekteissa ei kuitenkaan ole automaattista, sillä se ei ole yhteistoiminnallisten ryhmien universaali ominaisuus, vaan nousee niissä suotuisissa olosuhteissa (Woolley ym. 2010). Datatieteilijä Seth Spielman (2014) kysyy mitä kollektiivinen älykkyys merkitsee joukkoistetun paikkatiedon kohdalla ja edelleen, toteutuvatko suotuisat olosuhteet sen ilmenemiselle joukkoistetuissa paikkatietojärjestelmissä? Hän kuvaa joukkoistettua paikkatietoa pikaruuan tieteelliseksi vastineeksi: sitä on helposti ja halvalla saatavilla, mutta sen kuluttaminen tutkimuskäytössä ei välttämättä ole terveellistä.

Spielman (2014) huomauttaa yksittäisten korkealaatuisten kartoituspanosten uhkana olevan regressio kohti keskiarvoa. Hän haastaa suoraviivaista käsitystä kollektiivisesta älykkyydestä joukkoistetussa toiminnassa seuraavasti. Jos sillä syntynyt lopputuote ei ole enemmän kuin osiensa summa, parempi kuin paras yksilötason suoritus, eivätkö joukkoistetut järjestelmät

vähennä laatua parantamisen sijaan? Jos taas kollektiivinen älykkyys tuottaa parhaan kartan, asiantuntijat ovat tarpeettomia.

Ilmetäkseen, kollektiivinen älykkyys tarvitsee tekijöikseen riippumattomat kartoittajat, mielipiteiden moninaisuuden, hajautetun toimintamallin ja suotuisan tavan yhdistää yksittäiset panokset (Surowiecki 2004). Spielmanin (2014) mukaan vain hajautettu toimintamalli toteutuu lähtökohtaisesti joukkoistetuissa kartoitusprojekteissa. Suotuisa tapa yhdistää yksittäiset panokset voi puolestaan olla jopa heikkolaatuisimpien kartoituspanosten ulosrajaaminen. Loput kaksi tekijää eivät myöskään ole itsestään selvyiksiä, kuten edempänä alaluvuissa 2.1.4 Käyttäjät ja Kartoitus 2.1.5 esitetään.

Verkottuneen tietoyhteiskunnan muutoksessa joukkoistetun paikkatiedon on katsottu liittyvän neogeografiaksi kutsuttuun ilmiöön. Paikkatietoasiantuntija Andrew Turner (2006) määrittelee sen yksityisten kansalaisten kyvyksi luoda ja käyttää omia karttojaan, omilla ehdoillaan. Hän kuvaa neogeografian käsittävän kansalaislähtöisen, jokaisen saatavilla olevan tekniikan mahdollistaman tilallisen tiedon tuotantomallin, jonka toimintakenttä lomittuu auktoritatiiviseen paikkatietotuotantoon ja akateemiseen tutkimukseen. Sen kohteet ovat lähellä kansalaisen henkilökohtaista elämää ja toiminnan motiivit mm. sosiaalisia, kaupallisia tai poliittisia. Turner (2006) näkee paikkatietotuotannon kansalaislähtöisyyden demokratisoivan itse tuotantoa ja edelleen yhteiskuntaa.

Maantieteilijä Mordechai Haklay (2013) huomauttaa kuitenkin uusia teknologioita hyödyntämään kykenevän joukon edustavan rajattua osaa väestöstä ja tiedon kontrollin olevan usein kaupallisten etujen ohjaama. Siksi keskeinen virhe demokratisaatioargumentissa on hänen mukaansa oletama, että suurempi poliittinen ja sosiaalinen kontrolli on väistämättä mahdollista saavuttaa pelkästään lisäämällä teknologioita käyttävien määrää. Hän vertaa Turnerin (2006) käsitystä neogeografian demokratisaatiopotentialista kriittisen ja osallistavan GIS-tutkimuksen näkemyksiin ja näkee sen toteutuvan jälkimmäisten piirissä täydemmin.

Myös Goodchild (2009) on esittänyt neogeografian, auktoritatiivisen paikkatietotuotannon ja paikkatietotutkimuksen välistä lomittuneisuutta koskevia varauksia (Wilson & Graham 2013). Hän huomauttaa, että paikkatietoa tuottavaan kansalaistoimintaan verrattuna asiantuntijoiden tuottamalla maantieteellisellä tiedolla on selkeä teoriaan nojaava menetelmällinen eronsa. Esimerkiksi geodesian sovellukset on perusteltua jättää juridisissa yhteyksissä koulutetuille

ammattilaisille. Samoin tutkimuksen piirissä tarvitaan edelleen havaintodataa informaatioksi syntetisoiva ja sitä tulkitseva maantieteilijä muodostamaan ilmiöistä tietoa asiayhteyksissään.

Jäsentäessään neogeografiaa Goodchild (2009) tekee eron sen yhteiskunnallisen ulottuvuuden ja sen piirissä tapahtuvan maantieteellisen informaatiotuotannon välille. Hän kutsuu jälkimmäistä *Vapaaehtoisesti kerätyksi paikkatiedoksi* (Volunteered Geographic Information, VGI), jonka tuotantotavat olennaisesti poikkeavat auktoritatiivisesta paikkatietotuotannosta.

2.1.2 Tyypit ja teemat

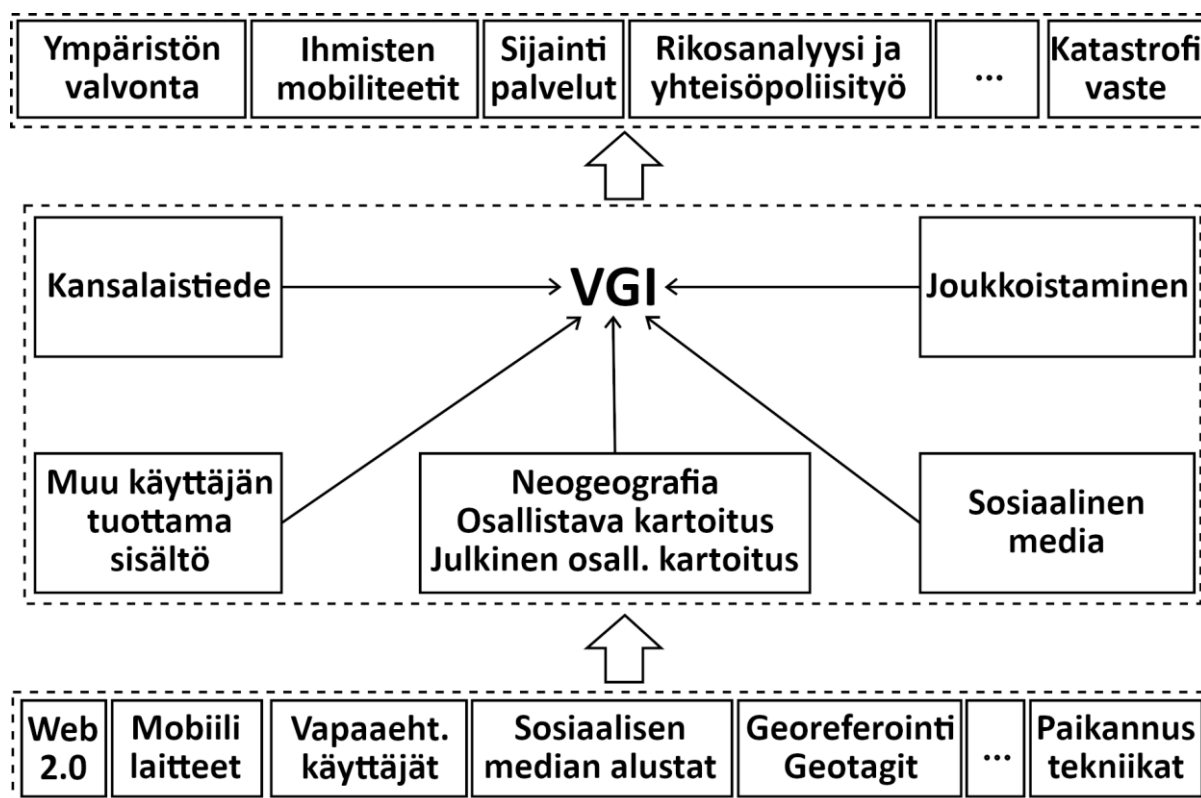
Tutkielmassa tarkasteluun valitaan kansalaislähtöisen paikkatiedon kentältä vapaaehtoisesti kerätty paikkatieto. Goodchild (2007) määrittelee sen *yksityisten kansalaisten muodostamien suurten joukkojen laajamittaiseksi toiminnaksi, jonka tavoitteena on kerätä maantieteellistä tietoa, ja jonka osanottajilla on usein vähäinen muodollinen pätevyys*. Seuraavassa tarkastellaan lähemmin sen osatekijöitä ja ominaisuuksia.

Goodchildin (2007) määritelmä kattaa useita rinnakkaisia joukkoistettuja ja osallistavia kartoituksen toimintatapoja, joista tässä mainitaan muutama esimerkki (kuva 1). Julkisella osallistavalla kartoituksella (PPGIS) on pyritty vahvistamaan kansalaisnäkökulmaa aluesuunnittelussa (Sieber, 2006; Fagerholm ym. 2012). Ekologian ja luonnonsuojelun alueella joukkoistusta on käytetty jo kauan ennen digitalisaatiota. Läntisellä pallonpuoliskolla edelleen jatkuva Audubon -yhdistyksen Christmas Bird Count on toiminut jo vuodesta 1900 (Echeverry-Galvis 2023). Vastaavaa *kansalaistieteellistä* toimintaa harjoitetaan myös monilla muilla tieteen aloilla (Wiggins & Crowston 2011). Liikemaailma on hyödyntänyt joukkoistusta ulkoistamalla toimintojaan kansalaisille (Taeuscher 2016). Esimerkiksi sosiaalisen median käyttäjän palvelulle jakama paikkatieto voi toimia kompensaationa yrityksen tarjoamista palveluista. Myös katastrofiavun tilannekuva ja riskinhallinta nojaa usein joukkoistettuun paikkatietotuotantoon (Albuquerque ym. 2016a).

See ym. (2016) luokittelevat joukkoistetun paikkatiedon verkkopalvelut ylätasollaan kolmeen ryhmään, joissa: 1. käyttäjä voi luoda ja jakaa kartan, 2. kerätään georeferoitua tietoa, 3. jaetaan asiantuntijoiden kansalaisavusteisesti tuottamaa tietoa.

Palvelut sisältävät See ym. (2016) luokituksessa monipuolisen joukon maantieteellisen tiedon teemoja, kuten: ekologia, koulutus, kartoitus, kalastus, nimirekisterit, maanpeite, vaellus ja ulkoilu, liikenne, sää, ympäristön valvonta, geokätköt, luonnonkatastrofit, matkailu ja

hakukonetieto. Sui ja Cinnamon (2017) ryhmittelevät nämä teemat löyhästi kolmeen ryhmään: kartoituksen peruspaikkatiedot, nimirekisterit ja temaattinen tieto.



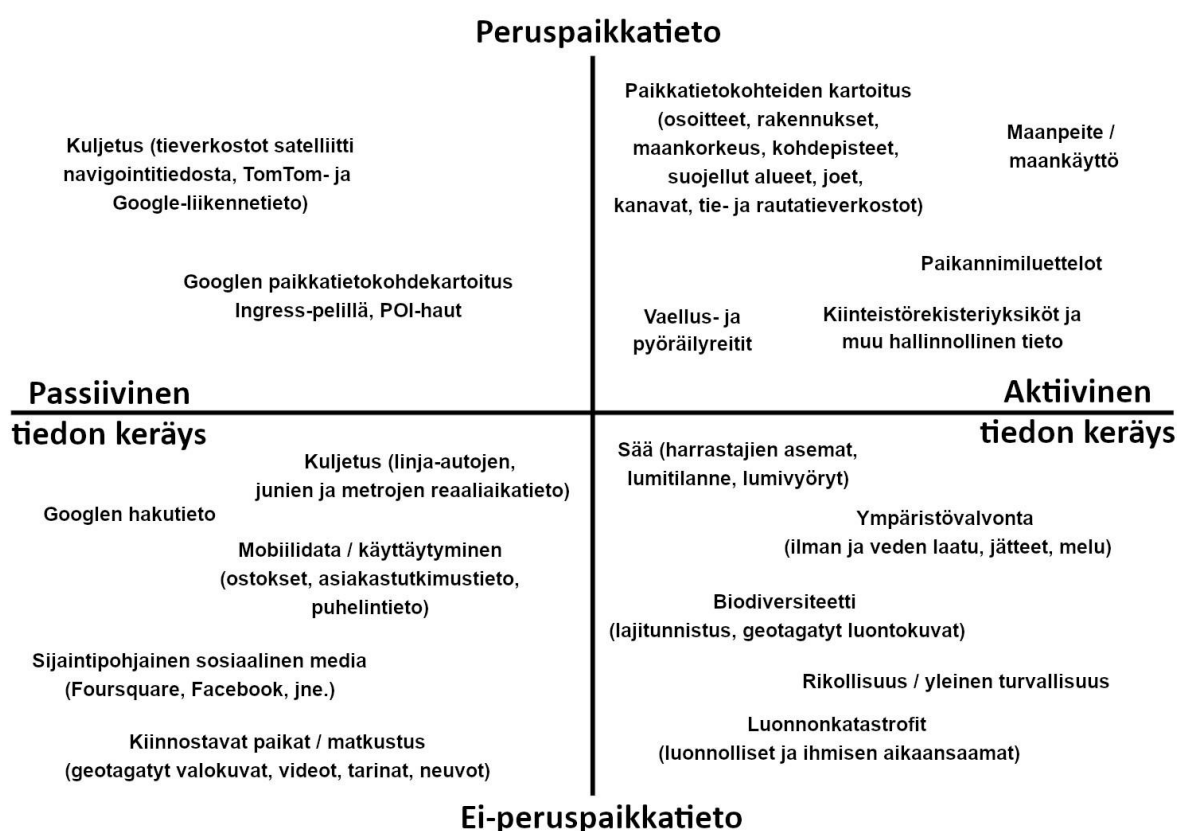
Kuva 1. Vapaaehtoisesti kerätyn paikkatiedon mahdollistavat osatekijät, lähteet ja sovellusalueet. Kuva on muokattu alkuperäisestä (Zhang 2021).

Tarkemmin kartoitukseen käytettävän joukkoistetun paikkatiedon ominaisuuksia See ym. (2017) jäsentävät kahdella kriteerillä. Ensimmäisenä niistä ilmaisee kuinka hyvin lähteet vastaavat kansallisten kartoitusorganisaatioiden paikkatietoinfrastruktuureissa ylläpitämiä peruspaikkatietoja. Näitä ovat geodeettinen-, geologinen-, maanpeite-, korkeus-, liikenne-, hydrografinen- ja kiinteistötieto sekä ortokuvat ja hallintoyksiköt (Directive... 2007). Peruspaikkatietoaineistot ovat asiantuntijoiden muodostamia ja niissä on minimimäärä virhettä. Niiden ajantasaisuus riippuu tuotannolle osoitetuista resursseista, mutta päivityksessä pyritään säännöllisyyteen.

Toinen See ym. (2017) kriteeri ilmaisee, onko tieto luovutettu aktiivisesti vai passiivisesti. Aktiivisesti tietoja luovuttaessaan kansalaiskartoittaja on tietoinen keräysprojektin agendasta ja kerättävistä havainnoista sekä osallistuu tuotantotehtäviin itse. Passiivisesti luovutetun tiedon lopullinen käyttötarkoitus voi jäädä käyttäjälle tuntemattomaksi. Hän on voinut tiedostamattaan sallia keräämisen, mutta ei ole tutustunut palvelun käyttöehtoihin tai estänyt

keräämistä asetuksin (Harvey 2013). Esimerkiksi sosiaalisesta mediasta kerättyä tietoa voidaan käyttää markkinointitarkoitukseen tai käyttäytymistutkimuksiin.

Aktiivinen tieto on helpommin prosessoitavissa ja vastaa käyttötarkoitustaan aineistotuotannon ennakkosuunnittelun ansiosta (See ym. (2017). Passiivisella tiedolla on aktiivista tietoa korostuneemmat massadatan piirteet, kuten suuri tuotantonopeus, moninaisuus ja määrä sekä problemaattinen luotettavuus (Kitchin & McArdle 2016). Se saattaa vaatia huomattavaa prosessointia ennen hyödyntämistä tai olla kokonaan täyttämättä aiotun sovelluskohteensa minimikriteerejä (See ym. 2017). Silti massadatan piirteistä arvo, voi hyödyntämisen onnistuessa olla suuri.



Kuva 2. Joukkoistetun paikkatiedon luokittelu sen mukaan, onko se kerätty aktiivisesti vai passiivisesti ja muistuttaako sen sisältö peruspaikkatietoja. Lähteiden suhteellinen sijainti kuvassa ei ole merkityksellinen, vaan maksimoi luotavuuden. Kuva on muokattu alkuperäisestä (See ym. 2017).

See ym. (2017) luokittelussa kuvan oikeaan yläneljännekseen sijoittuva joukkoistettu paikkatieto on lähtökohtaisesti yhteensopivaa kansallisten kartoitusorganisaatioiden tuottaman auktoritatiivisen peruspaikkatiedon kanssa (kuva 2). Se on kerätty kartoittajien aktiivisella panoksella ja sopii hyödynnettäväksi paikkatietoinfrastruktuureissa päivitys- ja

korjaustarkoituksessa. Neljännes alaoikealla sisältää lähteitä, joiden tietoa ei tavallisesti kerätä kartoitusorganisaatioiden toimesta. Niiden tarjoama tieto on aktiivisesti tuotettua ja on hyödynnettävissä muussa viranomaistoiminnassa tai tieteellisessä tutkimuksessa. Kuvan vasen puolisko sisältää esimerkiksi GNSS-palvelujen tai sosiaalisen median kautta passiivisesti kerätyn tiedon (kuva 2).

2.1.3 OpenStreetMap

Huomattavin esimerkki joukkoistetun paikkatiedon kartoituspalvelusta on OpenStreetMap (Yan ym. 2020). Suosittuna tutkimuskohteena (Mooney 2015) sitä on paikkatietotutkimuksen kentällä luonnehdittu jopa omaksi tutkimusalueekseen (Jokar Arsanjani ym. 2015).

OpenStreetMap on yhteisöllinen kartoitusprojekti, jonka tavoite on luoda avoimeen tietoon pohjautuva ilmainen maailmankartta. Vuonna 2004 perustettu palvelu on kahdessa vuosikymmenessä osoittanut elinvoimaisuutensa sekä joukoistettuna paikkatietokantana, että datan, tietojärjestelmien, sovellusohjelmistojen ja verkkopohjaisten tietovarastojen kasvavana ekosysteeminä (Mooney & Minghini 2017). Toukokuussa 2023 rekisteröityneitä jäseniä on yli 10 miljoonaa, joista aktiivisia kartoittajia on noin 50 000 käyttäjätiliä. Sekä jäsenten kokonaismäärän että aktiivisten jäsenten määrän kasvutrendi on ollut nouseva viimeisen vuosikymmenen (Neis 2023).

Palvelun syntymisen ja vahvan suosion kolme keskeistä toisiaan tukevaa tekijää ovat interaktiivinen Web 2.0, paikannustekniikka ja avoin kehitysmalli. Ne ovat luoneet pohjan tietoverkoissa toteutettaville suuren mittaluokan yhteistoiminnallisille projekteille (O'Reilly 2007). Toimintamuotoa on tukenut kuluttajamarkkinoille tullut halpa ja laadukas GNSS-tekniikka. Alun perin ohjelmistokehityksen tuotantomallina syntynyt avoin toimintatapa puolestaan mahdollistaa laajan yhteistoiminnan (Raymond 1999), jonka osanottajajoukolla on vaihteleva tekninen ja maantieteellinen osaaminen.

OpenStreetMap-palvelun sisältö kattaa laajan kirjon rakennetun ympäristön teemoja ja myös luonnonkohteita (Map features... 2023). Edustettuina ovat liikenneverkostojen ja reittitiedon lisäksi muiden muassa rakennukset, palvelut, turisti-, sotilas- ja historialliset kohteet, pelastustoiminta, terveydenhuolto ja maanpeite. Näissä teemaluokissa paikkatietokohteet jaetaan edelleen lukuisiin alaluokkiin.

Joukkoistetun paikkatiedon laatuongelma on sen tuotantotavan yleinen heterogeenisyys. Sillä tarkoitetaan vaihtelua käyttäjäkunnan ominaisuuksissa sekä puutteita järjestelmällisen

tuotantoprosessin soveltamisessa. Vaihtelu esiintyy erilaisina tuotannollisina ja osallistumiseen liittyvinä painotuksina, joista seuraava tasalaatuisuuden puute heikentää aineiston luotettavuutta ja rajaa näin sen sovelluspotentiaalia. Kun joukkoistetun paikkatiedon oikeellisuuden katsotaan teoriassa nojaavan Lévy'n (1999) joukkojen viisauteen, on joukkojen ominaisuuksista nousevan heterogeenisyyden syiden ymmärtäminen olennaista. Seuraavassa tarkastellaan OpenStreetMap-aineiston laatuun vaikuttavia tyypillisiä heterogeenisyystekijöitä.

2.1.4 Käyttäjät / kartoittajat

Käyttäjäkunnan motiivit, rakenne, toimintatavat, aktiivisuus, taitotaso ja kokemus vaikuttavat joukkoistetun paikkatiedon laatuun (Roper 2023). Pääsääntöisesti tietoa käyttäjien taustoista on niukasti saatavilla, sillä valtaosa palveluista ei kerää sitä (See ym. 2016).

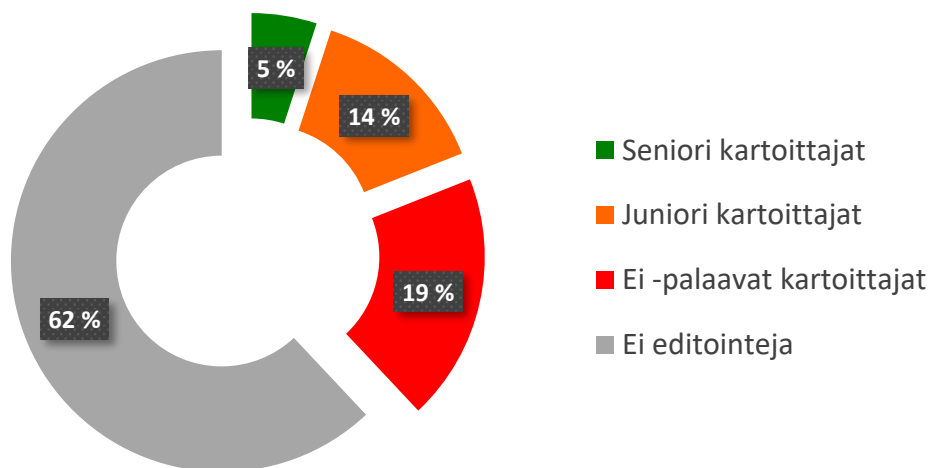
Käyttäjän tuottaman tiedon tyypillinen piirre on osallistumisepäsuhta, jossa pieni joukko aktiivisia tuottajia erottuu käyttäjäkunnan passiivisesta enemmistöstä. Osallistuminen seuraa Jacob Nielsenin (2006) heuristista 90:9:1-sääntöä, jonka mukaan 1 % käyttäjäkunnasta luo suurimman osan sisällöstä, 9 % osallistuu tuotantoon ajoittain ja 90 % hyödyntää aineistoa tuottamatta sitä (Roper 2023).

Neis ja Zipf (2012) luokittelivat OSM-käyttäjät kolmeen ryhmään heidän muokkausaktiivisuutensa perusteella (kuva 3). Yli 1 000 kohdetta luoneet käyttäjät olivat pitkäaikaisesti sitoutuneita aktiivisia Seniorikartoittajia. Juniorikartoittajat loivat yli 10 kohdetta ja Ei-palaavat käyttäjät alle 10. Valtaosa kaikista käyttäjistä ei tehnyt lainkaan muokkaustoimenpiteitä. Aktiivisen vähemmistön osuus käyttäjistä on noin 3–5 % (Neis & Zipf 2012; Yang ym. 2016).

Yhteisöllinen kartoitus vaatii käyttäjiltään eritasoisia kyvykkyyksiä, kuten tietoteknisiä perustaitoja, kartoitettavan ilmiön substanssiosaamista sekä vakioidun keräysprosessin ja paikannustekniikan hallintaa (See ym. 2016). Prosessin edellyttämä tehtäväosaaminen voi puolestaan sisältää luokittelua, digitointia ja eri tietolähteiden yhdistelyä (Albuquerque ym. 2016b).

Yang ym. (2016) tulkitsivat, etteivät aktiivisimmat käyttäjät sovi amatööriprofiiliin, vaan edustavat korkean taitotason asiantuntijamaista ryhmää. Ryhmän jäsenet kartoittavat eniten, sitoutuvat pitkä-aikaisesti ja hyödyntävät monimutkaisia työkaluja. Tämä poikkeaa

varhaisesta käsityksestä käyttäjäkunnasta amatöörien kokemattomana joukkona (Haklay 2010; Flanagin & Metzger 2008).



Kuva 3. Rekisteröityjen OSM-käyttäjien jakauma luokiteltuna kartoitusaktiivisuuden mukaan. Kuva on muokattu alkuperäisestä (Neis & Zipf 2012).

OpenStreetMap-kartoitusta tehdään itsenäisesti toimivien käyttäjien toimesta tai organisoidusti. Yu-Wei Linin (2011) mukaan organisoitua kartoitusta toteuttavat kaupalliset toimijat, hallinto ja kolmassektori, jotka toimivat rinnakkaisesti toisiinsa löyhässä yhteydessä olevien itsenäisesti toimivien yksilökartoittajien kanssa.

Molemmat kartoittajatyypit osallistuvat sisällöntuotantoon vaihtelevin motiivein. Toimintaa ohjaavat kaupalliset, humanitaariset ja henkilökohtaiset intressit, joissa paikallisuus korostuu (Budhathoki & Haythornthwaite 2013). Itsenäisesti toimivien kartoittajien motiiveina ovat esimerkiksi yhteisöllisyys, urahyödyt, paikallinen tieto, aineistojen vapaa saatavuus, osallistuminen yleishyödylliseen toimintaan tai tunnustuksen tavoittelu (See ym. 2016).

Tyypillinen itsenäisesti toimiva kartoittajakäyttäjä on hyvin toimeentuleva koulutettu nuori mies (Haklay 2016; Yang ym. 2016; Gardner ym. 2018). Yleisesti miehet kartoittavat tieverkostoa ja muokkaavat olemassa olevia kohteita, kun taas naiset kartoittavat rakennuksia ja luovat uusia kohteita (Gardner ym. 2018). Myös sisällön temattiset valinnat on pääosin tehty miesvaltaisesti (Stephens 2013). Lisäksi muokkausaktiivisuus on riippuvainen käyttäjien sosio-kulttuurisista taustoista, joissa yksilöllisyyttä ja kollektivismia painotetaan eri tavoin (Quattrone ym. 2014).

Organisoitua kartoitusta järjestävät yritykset, kuten Google, Amazon, Apple, Facebook ja Microsoft (Anderson ym. 2019; Who Uses OpenStreetMap... 2023). Ne ohjaavat kartoittajatiimejä ja tuovat palveluun valmista paikkatietoa massapäivityksin. Kuitenkin alueilla, joilla on yritysten tuottamaa tietoa, 70 % paikkatietokohteista on itsenäisesti toimivien kartoittajien tuottamaa. Tosin tieverkostotieto on pääosin näilläkin alueilla yrityslähtöistä.

Organisoitua kartoitusta tapahtuu myös humanitaarisissa hankkeissa, joita ohjaavat Humanitaarinen OpenStreetMap -tiimi tai kansainväliset organisaatiot (What we do... 2023). Tällöin saatavilla voi olla hyvätasoinen koulutus, korkealaatuinen lähdeaineisto sekä paikannus- ja ilmakuvauslaitteisto (Haklay ym. 2018). Kohteita ovat esimerkiksi luonnonkatasrofialueet tai ne määräytyvät paikallisten yhteisöjen muiden tarpeiden perusteella (Herfort ym. 2021).

2.1.5 Kartoitus

Mooney ja Minghini (2017) kuvaavat kartoituksen toteutustavat seuraavasti. Kohteita digitoidaan avoimista satelliitikuva-aineistoista, mikä mahdollistaa myös etäkartoituksen. Lähempänä kohdetta käytetään GNSS-palvelujen sijaintitietoa paikannukseen sekä reitistön jäljentämiseen laitteiden lokitiedosta. Lisäksi käytössä on analoginen menetelmä, jossa kohdealue kartoitetaan paperille ennen paikkatietokantaan siirtämistä. Sisältöä tuodaan palveluun myös massapäivityksinä muista paikkatietokannoista. Käytäntö herättää kritiikkiä, sillä sen nähdään heikentävän toiminnan keskeisenä ajatuksena olevaa lukuisten käyttäjien suorittamaa validointia, joka perustuu heidän paikalliseen tietoonsa, maantieteelliseen osaamiseen ja kohteissa fyysisesti tehtävään arvioon. Toisaalta Zielstra ym. (2013) ovat havainneet massapäivitysten ja kartoittajayhteisön suorittaman jälkijalostuksen parantavan OSM-tiedon laatua tehokkaimmin.

Yksittäiset käyttäjät voivat rajata kartoitustoimintansa tietyille alueille tai tiettyihin kohteisiin (Bégin ym. 2013). Luonnonkohteet ovat rakennettuja kohteita heikommin kartoitettuja (Arsanjani ym. 2015). Käyttäjillä voi olla paremmin kartoitettuja tuttuja kotialueita sekä ulkoisia alueita, joilla heidän toimintansa on vähäisempää (Zielstra ym. 2014). Osa käyttäjistä kartoittaa erityisen huolellisesti paikkoja, joissa he viettävät paljon aikaa (Napolitano & Mooney 2012). Yleisesti suosiossa olevat alueet kartoitetaan hyvin (Antonioni & Schleider 2018). Kartoitusalueiden koot korreloivat positiivisesti käyttäjän muokkausaktiivisuuden kanssa (Neis & Zipf 2012).

Quattrone ym. (2014) esittävät kartoittajien tunteman kiinnostuksen tiettyjä alueita kohtaan selittävän vaihtelua joukkoistetun aineiston täydellisyydessä. Heidän mukaansa myös alueiden deprivatio vaikuttaa kartoituskohteiden valintaan. Tätä selittää tuttujen alueiden suosituimmuus ja taustaltaan paremmin toimeentuleva tyypillinen käyttäjä.

Truong ym. (2019) ovat havainneet, että käyttäjät saattavat toimia pioneerikartoittajina, jotka lisäävät kartoittamattomalle alueelle kohdehierarkian ylätasoa kohteita. Heitä seuraavat käyttäjät puolestaan tarkentavat karttaa alatasoa kohteilla. Joillain alueilla kartoitus voi pohjautua vain yhden aktiivisen käyttäjän panokseen. Lisäksi osa käyttäjistä keskittyy luomaan uusia kohteita, kun taas toiset saattavat tehdä vain korjauksia.

2.2 Joukkoistetun ja auktoritatiivisen paikkatiedon kytkennät

2.2.1 INSPIRE-paikkatietoinfrastruktuurin ja OpenStreetMapin vertailu

Käyttäjäkunnan vaihtelevien ominaisuuksien lisäksi OSM-aineiston laadulliseen heterogeenisyyteen vaikuttavat sen tuotantoalustan ominaisuudet. Seuraavassa niistä keskeisimpiä verrataan keskitettyyn tuotantomalliin.

Auktoritatiivinen julkinen paikkatietotuotanto on Euroopan Unionissa säännelty INSPIRE-direktiivillä, joka ohjaa jäsenmaita yhdenmukaistamaan paikkatietotuotantonsa eurooppalaiseksi paikkatietoinfrastruktuuriksi (Directive... 2007). INSPIREn tavoitteena on parantaa paikkatiedon jakelua, saatavuutta, hallintaa, dokumentointia ja yhteentoimivuutta jäsenmaiden organisaatioiden välillä. Tämä edellyttää tekniikan, toimintatapojen ja datan vakiointia sekä laadun varmistamista. INSPIREn vaatimustenmukaisuus koskee julkisia toimijoita, mutta ei velvoita kaupallisia toimijoita. Se kuitenkin antaa niille ohjaavan mallin, joka tukee INSPIREn tavoitetta luoda toimintaedellytyksiä eurooppalaiselle datataloudelle (ERT position... 2017).

Mooney ja Minghini (2017) tarkastelevat OpenStreetMap-palvelua joukkoistettuna paikkatietoinfrastruktuurina. Tätä heidän mukaansa perustelee OSM:n ympärille muodostunut monimutkainen palveluiden ja sovellusten ekosysteemi sekä laaja käyttäjäyhteisö, tietokannan hyvä saatavuus ja sen temaattinen monipuolisuus.

Minghinin ym. (2019) INSPIRE-OpenStreetMap-vertailussa tunnistamat erot esitetään oheisessa taulukossa (taulukko 1). Suurin ero on lähestymistavassa. INSPIRE on ylhäältä alas johdettu ja lailla säännelty alueellinen hierarkia, jonka toimintaa valvovat viranomaiset. OSM

sitä vastoin on kaikille avoin, vapaaehtoisen osanoton pohjalta muodostunut joukkoistettu globaali projekti, jota OpenStreetMap-säätiö tukee, muttei ohjaa.

Taulukko 1. OSM:n ja INSPIRE-paikkatietoinfrastruktuurin keskeisimpien piirteiden vertailu. (Minghini ym. 2019).

	INSPIRE	OSM
Lähestymistapa	ylhäältä-alas	alhaalta-ylös
Temaattinen laajuus	34 ympäristöteemaa	mikä tahansa kohde
Tietorakenne	monimutkainen tietomalli	yksinkertainen tietomalli
Koordinaattijärjestelmä	INSPIRE-spesifit järjestelmät	WGS84
Datan saatavuus	OGC-yhteensopivat ohjelmat, geoportaalit	APIt, Planet-tiedosto tai rajatut otteet
Datalisenssi	erilaisia, riippuen datan tuottajista	OdbL

INSPIREN tematiikka kattaa Euroopan laajuisesti 34 asiantuntijoiden ennalta määrittelemää ympäristöön liittyvää paikkatietokohteiden pääluokkaa. OSM:n juuret ovat katutietokannassa, jonka tematiikka laajentui monipuoliseksi vapaamuotoisen tarvelähtöisesti ajan kuluessa.

Tietorakenteissa on huomattavia eroja. INSPIREN teemakohtaiset tietomallit nojaavat sen toimijaverkoston yhteisesti sopimaan näkemykseen. Yhteentoimivuus varmistetaan yksityiskohtaisilla ja moniulotteisilla tietomalleilla. OSM:n yksinkertainen tietomalli perustuu kolmeen elementtityyppiin: solmuihin, teihin ja suhteisiin, joiden yhdistelmillä kuvataan kaikki kohteet. Ominaisuustieto on tallennettu avain-arvo-pareihin eli tageihin, joita käyttäjillä on täysi vapaus luoda tarpeisiinsa.

INSPIRE edellyttää pan-eurooppalaisen koordinaattijärjestelmän käyttöä, kuten ITRS tai ETRS89. Kansallisia järjestelmiä on käytössä useita. OSM-tieto tallennetaan WGS84-järjestelmässä, sillä GPS-laitteet ovat keskeinen kartoitustyökalu. Niiden käyttö takaa automaattisesti yhteensopivuuden koko tietokannassa.

Data on saatavilla INSPIREN palveluorientoidussa arkkitehtuurissa OGC-yhteensopivilla asiakasohjelmilla löytö-, luettelointi-, katselu- visualisointi-, lataus ja validointipalvelujen kautta. Metadataa ylläpidetään erikseen. OSM-datan yksinkertainen tietomalli mahdollistaa

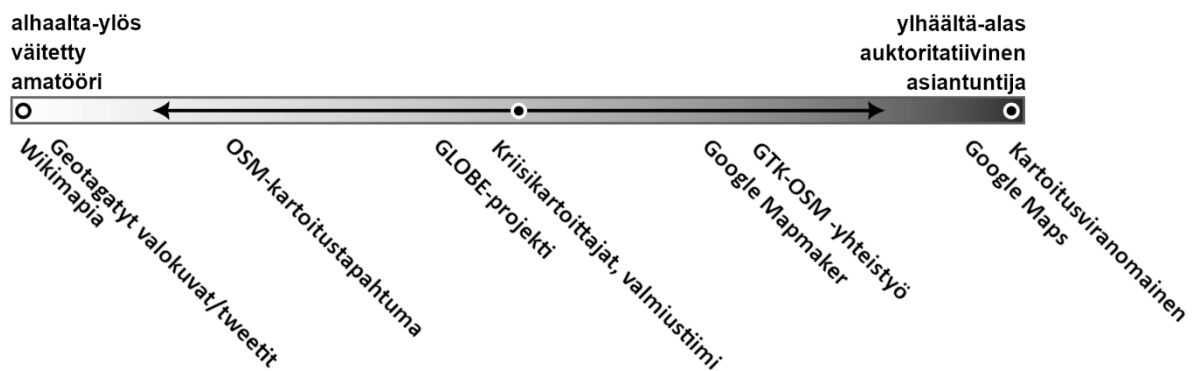
datan saatavuuden useista lähteistä monilla eri formaateilla. Metadataluetteloja ei tarvita, sillä tieto on tallennettu vaihtelevasti elementteihin liitettyihin tageihin.

Datalisenssikäytännöt eroavat INSPIREN ja OSM:n välillä. INSPIRE ei edellytä tiettyä lisenssiä minkä seurauksena niitä, ja erilaisia rajoituksia datan käytölle on monia. OSM-tietokanta on puolestaan kokonaisuudessaan yhden avoimen datan lisenssin alainen (ODbL), joka sallii vapaan käytön ja muokkaamisen, jos tekijäviite mainitaan.

McCullagh ja Jackson (2013) kuvaavat standardeja seuraavaa auktoritatiivista paikkatietoa vakaaksi ja kattavaksi tietolähteeksi. Joukkoistettu paikkatieto on puolestaan osittaista, löyhästi määriteltyä ja repaleista. Se kuitenkin liittyy yhteiskunnan ruohonjuuritasolla tärkeisiin teemoihin.

Tuotantomuotojen eroja ja kehitystä syntetisoi maantieteilijä Jonathan Cinnamonin (2015) näkemys auktoritatiivisen ja joukkoistetun toimintatavan muodostamasta akselista liukumana, jolle sovellustarpeiden sanelemat ratkaisut asemoituvat (kuva 4). Kysymys ei ole joko-tai-asetelmasta, vaan joukkoistetun paikkatiedon hyödyntämismahdollisuudet voidaan nähdä auktoritatiivisessa yhteydessä hybridinäkökulmasta.

PAIKKATIETOTUOTANTO



Kuva 4. Paikkatietotuotanto jatkumona, jonka napoina ovat joukkoistettu ja auktoritatiivinen toimintatapa. Kuva mukailtu alkuperäisestä (Cinnamon 2015).

Datan laatu tunnistetaan sekä hallinnossa että tutkimuksessa olennaiseksi tekijäksi pyrittäessä kehittämään paikkatietoinfrastruktuureja osaksi toiminnallisesti ja liiketaloudellisesti kestävää dataekosysteemiä (Martin ym. 2021; Olteanu-Raimond ym. 2017).

2.2.2 Joukkoistetun paikkatiedon hyödyntäminen

Joukkoistettu paikkatieto on osa digitalisaation synnyttämää datataloutta (EER position... 2017). Se liittyy kehitykseen, jossa julkinen ja kaupallinen paikkatietotuotanto arvioivat roolejaan tuottajina uudelleen ja tutkivat mahdollisuuksia joukkoistuksen hyödyntämiseen.

Datatalouden institutionaaliset toimijat ovat mukana joukkoistetussa paikkatietotuotannossa hyödyntämällä ja hyödyttämällä OpenStreetMap-palvelua (Tietoaineistojen... 2016; Saretta ym. 2023). Joukkoistuksen merkittävimpinä hyötyinä on nähty tiedon vapaa saatavuus, suuri määrä, ajantasaisuus ja potentiaali olemassa olevan paikkatiedon rikastamiseen sekä aineistomuodostuksen kustannustehokkuus. Joukkoistettu paikkatieto on joissain tapauksissa todettu laadultaan jopa auktoritatiivista paikkatietoa paremmaksi. (Olteanu-Raimond ym. 2017). Lisäksi joukkoistuksella tuotettu paikallinen tieto ja kansalaisten osallistaminen tukevat avoimen tiedon saannin sekä päätöksenteon vahvemman demokratisaation yhteiskunnallisia tavoitteita.

Olteanu-Raimond ym. (2017) havaitsivat eurooppalaisen julkisen sektorin kartoitusorganisaatioiden hyödyntävän joukkoistettua paikkatietoa vaihtelevasti. Esimerkkejä ovat kansalaisten viranomaisille tekemät ilmoitukset, muutoshavainnointi, virheen korjaus, kansankielisten paikannimien rekisteröinti ja temaattisesti uudenlaisen sisällön kerääminen.

Kansallisissa kartoitusorganisaatioissa paikkatiedon ylläpitoprosessit seuraavat systemaattisen päivityksen ja revisioinnin toimintatapoja. Olteanu-Raimondin ym. (2017) mukaan erityisesti aineistojen jatkuvan päivityksen prosessi hyötyy joukkoistetun paikkatiedon käyttämisestä lähteenä. Näissä yhteyksissä joukkoistaminen järjestetään pääasiassa viranomaislähtöisesti, mutta myös OSM-aineistoa on hyödynnetty. Hyödyntämisen ongelmoina ovat validoinnin kustannukset, käyttöoikeuskysymykset, tietolähteen pysyvyys sekä datan laatu ja luotettavuus.

Kaupallinen sektori käyttää OSM-dataa mm. pohjakarttoina, navigointi- ja sijaintitietona sekä sisäisesti esimerkiksi logistiseen optimointiin (Major... 2023). Huomattavia hyödyntäjiä ovat Amazon, Apple, Baidu, Facebook, Grab ja Microsoft. (Who... 2023).

Hankkeissaan humanitaarista kartoitusta organisoivat mm. Yhdistyneet kansakunnat, Maailmanpankki, Punainen risti ja Lääkärit ilman rajoja (Herfort ym. 2021). OSM:n rinnalla toimii myös Humanitaarinen OSM-tiimi, jonka tavoitteena on auttaa saavuttamaan kestävän

kehityksen tavoitteet, tukea avustustyötä luonnonkatastrofialueilla ja avustaa riskien hallinnassa (What... 2023).

2.2.3 Hyödyntäminen Suomessa

Joukkoistamisen merkitys paikkatietotuotannossa on tunnistettu suomalaisessa julkisen hallinnon strategisessa suunnittelussa (Karlsson 2018). Eräs suunnittelun painopisteistä on datatalous. Joukkoistetun tiedon hyödyntämisen edellytyksenä pidetään kansallisten tietovarantojen systematisointia, yhteentoimivia tietojärjestelmiä, paikannusteknologioiden käyttötapojen yhdenmukaistamista ja avointa dataa. Tavoitteena on edistää dataa hyödyntävän palvelutalouden toimintaedellytyksiä ja parantaa julkisen hallinnon palveluiden tuottavuutta ja palvelukykyä. Tärkeänä pidetään kykyä tunnistaa miten erilaiset paikkatietovarannot tukevat datatalouden mekanismien syntyä. (Kansallinen... 2021; Karlsson 2018). Liikennetietoa koskevaksi keskeiseksi tavoitteeksi on asetettu datan ja uusien datalähteiden hankintaan tähtäävä kehitystyö, jota halutaan viedä eteenpäin datatalouden paikkatietoekosysteemissä käytävänä toimijoiden välisenä dialogina (Marjamäki 2014).

Liikenneviraston selvityksessä ajantasaisen liikennetiedon arvo nähdään merkittävänä (Laine ym. 2013). Sillä tunnistetaan olevan vaikutus liikkujien päätöksentekoon ja käyttäytymiseen. Tärkeimmät vaikutusmekanismit liittyvät matkustusajankohdan ja reitinvalinnan muuttamiseen sekä liikennekäyttäytymisen muutoksiin. Lisäksi liikennetiedon avulla nähdään mahdolliseksi parantaa liikenneverkon operatiivista hallintaa ja siten lisätä tehokkuutta ja turvallisuutta. Suuren käyttäjämäärän vuoksi liikenneympäristössä joukkoistamalla tuotetun ajantasaisen, mahdollisesti reaaliaikaisen tiedon laatu paranee ja täten käyttöarvo kasvaa.

Liikenneviraston esiselvitys valtakunnallisesta pyöräilyn tiedonhallintamallista sisältää ehdotuksen joukkoistuksen käyttämisestä tiedon keräämisessä (Laitinen et al. 2015). Sitä on suunniteltu hyödynnettävän kansallisen pyöräilyverkoston digitaalisen tietomallin luomisessa, sillä yhtenäistä valtakunnallisesti kattavaa mallia ei vielä ole. Verkoston inventoinnin kustannukset on koettu suuriksi varsinkin pienissä kunnissa, jolloin viranomaisten organisoima joukkoistaminen on nähty varteenotettavana toteutustapana inventoinnille. Tiedetään, että auktoritaaristen lähteiden tietomallit poikkeavat sekä toisistaan että OpenStreetMap-aineistosta. Tällöin aineistojen laatu-tieto tukee niiden yhdistämispyrkimyksiä.

Viranomaiset ovat Suomessa tutkineet auktoritaarisen ja joukkoistetun tietolähteen yhteentoimivuutta mm. joukkoliikennepysäkkätiedon ja pyöräilyverkostotiedon integroimiseksi (Peltonen 2016; Tietoaineistojen... 2016). Jälkimmäisessä tapauksessa hyödyntämiselle on nähty kolme vaihtoehtoa. 1) Paikkatiedon pohjaaminen ensisijaisesti joukkoistettuun lähteeseen, 2) auktoritaarisen ja joukkoistetun lähteen yhteistoiminnallinen malli ja 3) joukkoistetun paikkatiedon käyttäminen lähinnä referenssinä auktoritaarisille lähteille. Pilottiprojektin ehdotuksena oli, että Liikennevirasto perustaisi pyöräteiden paikkatiedot OpenStreetMap-palveluun, jolloin pyöräteiden kartoittamista ohjattaisiin ja tuettaisiin erinäisin hallinnollisin toimenpitein. Tuloksena laadun toivottiin kohenevan molemmissa tietoaineistoissa.

2.2.4 Joukkoistettu paikkatieto ja paikkatietoinfrastruktuurien muutos

Paikkatietoinfrastruktuurit liittyvät rakenneosina laajempiin datatalouden konsepteihin, kuten Digital Earth ja älykaupunki (Guo ym. 2019; Chaturvedi ym. 2019). Siten joukkoistetun paikkatiedon hyödynnettävyys niissä on vahvasti riippuvainen sen integroitavuudesta paikkatietoinfrastruktuuriin.

INSPIRE-paikkatietoinfrastruktuurin kehitys- ja käyttöönottojakson aikana paikkatietotuotanto on kokenut merkittäviä muutoksia. Tärkeimpinä näistä Kotsev ym. (2020) nostavat esille tietolähteiden määrän ja datan volyymin kasvun, tietoteknisen infrastruktuurin arkkitehtoniset muutokset, teknologian ja toimintatapojen standardoinnin tarpeet sekä monipuolistuneet tekniset vaihtoehdot paikkatietoekosysteemissä.

Yksityinen sektori ja kansalaistoiminta ovat ottaneet paikkatiedon tuottajina suuremman roolin, jolloin julkisen paikkatietotuotannon rinnalle on syntynyt uusia paikkatietolähteitä, kuten mobiili-, IOT- ja satelliittidata.

Kotsev ym. (2020) mukaan kehitys ei tee paikkatietoinfrastruktuureista vanhentuneita, sillä paikkatietotuotannon heterogeenisyys, juridiikka, toteutuneet investoinnit ja jo olemassa oleva käyttäjien yhteisö edellyttävät harmonisoivaa ja vakaata tuotantokehystä.

Näistä lähtökohdista Kotsev ym. (2020) näkevät paikkatietoinfrastruktuurin keskeisenä evoluutiossa kohti Euroopan Unionin digitaalisen talouskehityksen tavoitteekseen asettamia "Datatiloja". Ne määritellään "saumattomiksi digitaalisiksi alueiksi mittakaavoissa, jotka mahdollistavat uusien dataan perustuvien tuotteiden ja palvelujen kehittämisen." (Euroopan komissio 3. 2020).

2.3 Paikkatiedon laatu ja luotettavuus

Todellisuuden monimutkaisuutta ei voi täydellisesti kuvata paikkatietoaineistossa, jota koostettaessa on tehtävä sisällöllisiä valintoja esimerkiksi kuvattavista kohteista, tiedonkeruumenetelmistä ja tarkkuudesta. Tämä johtaa paikkatiedon tuottamiseen aina määrättyyn tarkoitukseen. Paikkatiedon odotetaan myös kuvaavan todellisuutta luotettavasti. Kohteiden geometrian, ominaisuuksien, ajantasaisuuden ja keskinäisten suhteiden on täytettävä tuotantotavoitteet vastaten todellisuutta aineiston käyttöyhteyden edellyttämällä tavalla. Näin paikkatiedon laadun ongelmaa tarkastellaan sekä tuottajan että käyttäjän näkökulmista.

Paikkatiedon luotettavuus voidaan määritellä varmuudeksi sen oikeellisuudesta, jonka muodostavat virheen määrä, tarkkuus, looginen eheys, täydellisyys, epävarmuuden määrä ja sopivuus käyttötarkoitukseen (Evans 1997). Näitä tekijöitä käsitellään tarkemmin seuraavassa.

Laatuvaihtelun ytimessä on maantieteellisen tiedon epävarmuus. Se on perusominaisuus, josta nousevia virheitä paikkatiedossa ja siitä tehtävissä tulkinnoissa ei täysin voi poistaa. (Couclelis 2003). Virheitä on kuitenkin pyrittävä minimoimaan, jotta laatu ja sopivuus käyttötarkoitukseen voidaan varmistaa.

Geoinformaatikko Andrew Frankin (2007) mukaan epätäydellinen tieto muodostuu sen keräämisen ja yhdistämisen aikana tehtyjen sitoumusten seurauksena. Näitä ovat esimerkiksi havainto-, yleistys-, ja luokittelusitoumukset, jotka ovat erilaisia käsitteellisiä ja toimintatapoihin liittyviä valintoja. Sitoumukset voivat johtaa virheisiin ja kertaantua tiedontuotantoprosessin eri vaiheissa.

Todellisuuden ja paikkatiedon vastaavuus on riippuvainen paitsi virheistä, myös käsiteavaruudesta, jolla todellisuutta kuvataan (Azouzi 2000; Van Oort 2006). Aalders (2002) jakoi todellisuutta paikkatietoaineistoon mallintavan prosessin kahteen vaiheeseen, jossa ensin rajataan todellisuudesta valittavat kohteet ja kuvataan tavat abstrahoida ne. Toisessa vaiheessa määritellään paikkatiedon tallentamiseen käytetyt menetelmät ja niiden vaatimukset. Paikkatietotutkija Pepijn Van Oortin (2006) mukaan prosessissa muodostetaan paikkatietoaineiston käsitelmä, jossa kuvataan sen ontologinen perusta spatiaalisesta, temporaalisesta ja temaattisesta näkökulmasta. Frank (2007) toteaa, että todellisuuden ja sitä

kuvaavan informaatiomallin linkityksen kautta paikkatiedon laadun voi ymmärtää viimekädessä sopivuutena käyttötarkoitukseen.

Paikkatietoaineisto on luotettavaa, kun se vastaa tuotanto- ja käyttötavoitteiden edellyttämiä käsiteavaruuksia ja epävarmuus suhteessa näihin vaatimukseen on hyväksyttävällä tasolla (Devillers & Jeansoulin 2006: 31–41). Virhelähteiden vaikutus on huomioitava, jotta laadun kontrolli tuotantoprosessissa on mahdollinen ja paikkatietoaineiston rajoitteet ymmärretään. Laatu ja luotettavuus nähdään tässä yhteydessä sekä tuottajan että käyttäjän näkökulmista. Seuraavassa käydään läpi paikkatiedossa vaikuttavat epävarmuuden lähteet, sen laadun osatekijät ja tapa, jolla todellisuutta kuvaavat havainnot muodostetaan paikkatietoinformaatioksi.

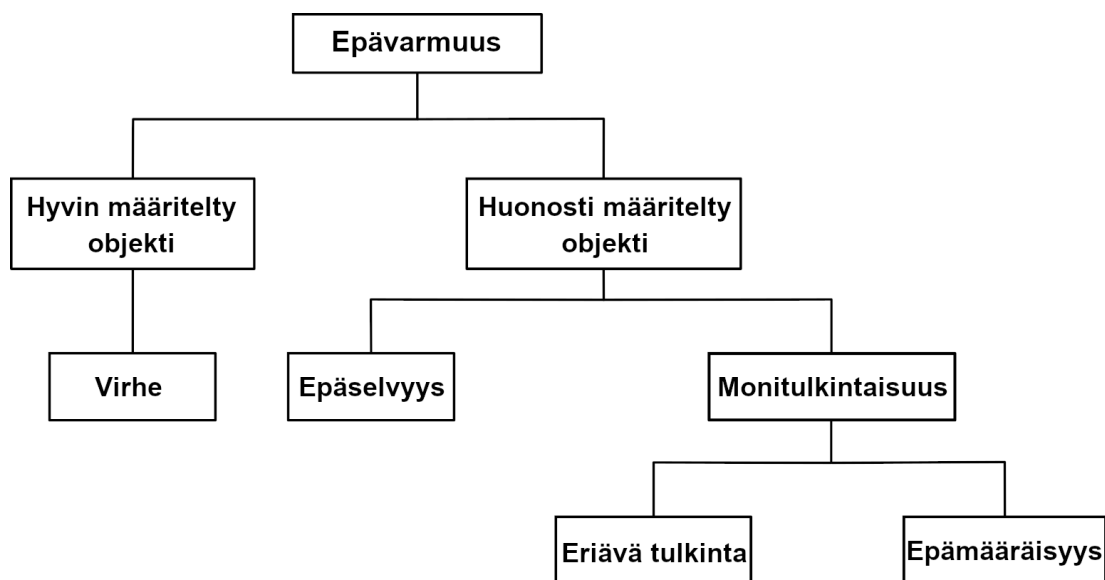
2.3.1 Epävarmuus

Epävarmuuden käsite kuvaa paikkatieton luontaista epätäydellisyyttä. Se nousee neljästä lähteestä: todellisuuden satunnaisuudesta, ihmisen havaintorajoitteista, tiedonkeruun ja mittaustekniikan rajallisuudesta sekä epävarmuuden esiintymisestä ja leviämisestä tiedonkeruun ja analyysin prosesseissa (Shi 2010: 15–26).

Epävarmuuden keskiössä ovat paikkatietokohteiden luokan ja niiden ominaisuuksien määrittelyongelmat. Fisher ym. (2006) jakavat epävarmuuden lähteet kahteen päätyyppiin (kuva 5):

1. *Virheiden aiheuttama* epävarmuus, joka esiintyy *hyvin määriteltyjen* ja objektiivisesti havaittavien kohteiden kohdalla. Tällöin epävarmuus on probabilistista ja sitä hallitaan tilastollisin menetelmin (Tavana ym. 2016). Esimerkiksi pyöräilyverkoston sijaintitiedossa on probabilistista epävarmuutta, kun sen mittaustulokset saadaan mittalaitteen virhemarginaalin vaihteluvälillä tai ne sisältävät inhimillisiä virheitä.
2. *Epäselvä* tai *monitulkintainen* epävarmuus huonosti määriteltyjen kohteiden kohdalla, jota hallitaan ei-tilastollisin menetelmin (Tavana ym. 2016).
 - a. *Epäselvä* epävarmuus: kohteiden tai luokkien huono määrittely, joilla ei ole selkeää luokittelujärjestelmää. Esimerkiksi, jos pyöräilyverkostoon määritellään kuuluvaksi pyöräilyn mahdollistavat reitit, lukeutuvat samaan luokkaan sekä pyörätiet että maantiet.

- b. *Monitulkintainen epävarmuus*: yksittäinen kohde on hyvin määritelty, mutta se voidaan luokitella monin tavoin. Monitulkintainen epävarmuus jakautuu *eriävään tulkintaan* (subjektiivisten luokittelujärjestelmävalintojen seurauksena) ja *epämääräisyyteen* (luokittelujärjestelmien yhteensopimattomuuden aiheuttama ristiriita).



Kuva 5. Paikkatiedon epävarmuuden käsitteellinen malli. Paikkatietokohteen määrittelyn täsmällisyys määrää epävarmuuden luonteen (Fisher 1999).

Fisher (2006) kuvaa hyvin ja huonosti määriteltyjä maantieteellisiä kohteita seuraavasti.

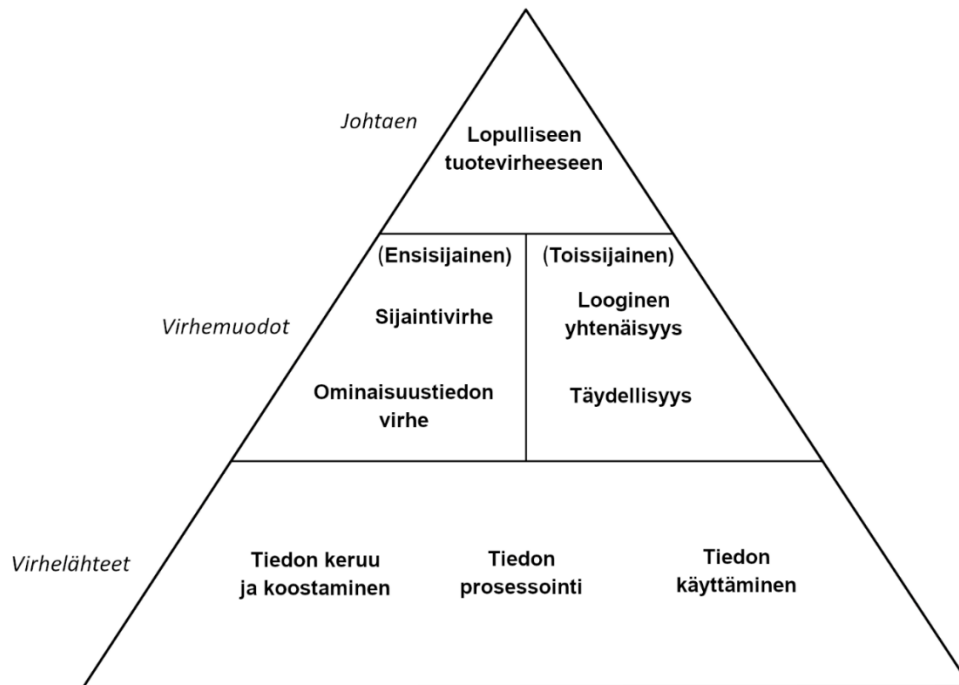
Hyvin määritellyt kohteet, kuten väestönlaskennan tai rakennetun ympäristön osat, perustuvat selkeisiin rajoihin ja ominaisuuksiin, jotka johtuvat vakiintuneista poliittisista ja oikeudellisista järjestelmistä. Huonosti määritellyt kohteet, kuten luonnonilmiöt tai sosiaaliset konstruktiot, ovat haastavampia rajata ja ne perustuvat usein ilmiöiden keskeisiin ominaisuuksiin ja keskimääräisyyteen.

2.3.2 Virheet

Tässä tutkielmassa keskitytään hyvin määriteltyjen kohteiden epävarmuuden arviointiin, jolloin tarkastelun kohteena on virhe ja sen määrä. Virheitä luokitellaan kolmeen tyyppiin: satunnaisiin, karkeisiin ja systemaattisiin (Azouzi 2000; Shi 2010: 5–6).

Satunnaisvirheet noudattavat Gaussin normaalijakaumaa ja niiden esiintymistä on vaikea ennakoida yksittäisissä havainnoissa, mutta niitä voidaan mallintaa tilastollisesti riittävän suuren havaintomäärän avulla. Karkeat virheet ovat selkeästi normaalijakaumasta poikkeavia ääriarvoja tai erehdyksiä, jotka eivät kuvaa todellista populaatiota. Systemaattiset virheet

syntyvät puolestaan esimerkiksi virheellisistä parametreista tai kalibrointivirheistä, ja ne ovat läsnä samanlaisina kaikissa havainnoissa. Nämä voidaan korjata myöhemmissä tiedonkäsittelyn vaiheissa (Azouzi 2000).



Kuva 6. Paikkatiedon virheluokittelu. Virhelähteet vaikuttavat aineistomuodostuksessa laatuparametreittain, kasautuen lopullisen paikkatietotuotteen kokonaisvirheeksi (Hunter & Beard 1992).

Virheitä esiintyy, ja ne leviävät kaikissa paikkatiedon tuotantoprosessin vaiheissa (Hunter & Beard 1992; kuva 6.). Collins ja Smith (1994) tunnistavat virheiden lähteitä esimerkiksi tiedonkeruussa, -syötössä, -tallennuksessa, -käsittelyssä, -tulostuksessa sekä -käytössä. Vaiheet luovat esiintymismahdollisuuden virheille, jotka kumuloituvat ja siirtyvät lopulliseen paikkatietotuotteeseen.

Hunterin ja Beardin (1992) luokittelussa esitetään maantieteellisten havaintojen sisältämiä tilaan, ominaisuuksiin ja keskinäisiin suhteisiin liittyviä virhetyyppejä (kuva 6). Ne esiintyvät kaikissa seuraavassa määriteltävissä paikkatiedon laadun peruskomponenteissa.

2.3.3 Laadun osatekijät

Veregin (2005) määrittelee paikkatiedon laadun peruskomponentit seuraavasti:

Täydellisyys (Completeness): on paikkatietoaineiston kohteiden ja sellaisen abstraktin kohdejoukon suhde, joka sisältää kaikki mahdolliset samantyyppiset kohteet. Abstrakti kohdejoukko muodostetaan paikkatietotuotteen spesifikaation tai käyttötarpeen pohjalta.

Joukkoa rajaavat valintakriteerit ja kohteiden kuvaus on määriteltävä tarkkaan, jotta paikkatietoaineiston ja abstraktin joukon välinen suhde voidaan varmistaa. Vertailuasetelma mahdollistaa arvion paikkatietoaineiston ja abstraktin joukon vastaavuudesta. Täydellisyys käsittää puuttumisen ja sisällymisen alaluokat, jotka ilmaisevat onko aineistossa liikaa vai liian vähän tietoa verrattuna todellisuuteen tai abstraktiin joukkoon.

Tarkkuus (Accuracy): On virheen vastakohta. Virhe on todellisen arvon ja vertailtavan arvon eroavuus. Arvo ei kuvaa todellisuudessa valitsevaa objektiivista totuutta, joka on epävarmuuden vuoksi tulkinnanvarainen, mahdotonta havainnoida, tai epäkäytännöllistä määrittää. Siksi paikkatieto muodostetaan tietokantaan todellisuudesta tehdyn mallinnuksen määräämässä kehysessä. Eroavuus esiintyy todellisuuden ilmiöistä laaditun mallin arvojen ja vertailtavan paikkatiedon arvojen välillä. Tarkkuus on siis suhteellinen mitta, sillä se riippuu paikkatiedolle ennalta asetetusta muodosta ja sisällöstä.

Täsmällisyys (Precision): Täsmällisyys, resoluutio tai granulariteetti on erotettavissa olevien yksityiskohtien määrä. Paikkatieto on täsmällisyydeltään rajoittunutta mittausvälineiden ja menetelmien erotuskyvyn rajallisuuden sekä paikkatiedon läpi käymän yleistyksen vuoksi. Yleistys on välttämätöntä, sillä paikkatietokannat voivat sisältää vain murto-osan todellisuuden kohteista, niiden ominaisuuksista ja kohteiden välisistä suhteista. Täten täsmällisyyden taso mahdollistaa tai rajaa paikkatiedolle soveltuvia käyttötarkoituksia. Matalalla resoluutiolla ei ole samaa negatiivista merkitystä kuin vähäisellä tarkkuudella. Esimerkiksi yleistävät mallit tai aluetason tarkastelu edellyttävät matalampaa täsmällisyyttä. Täsmällisyys on myös tarpeellista huomioida tulkittaessa tarkkuutta. Kahden paikkatietoaineiston tarkkuus voi olla sama, mutta jos ne poikkeavat erotuskyvyltään, on niiden tarkkuuden laatu myös poikkeava.

Eheys (Consistency): merkitsee paikkatiedon ristiriidattomuutta. Usein sillä tarkoitetaan paikkatiedon vaatimustenmukaisuutta kohdemaailmaa kuvaavaan loogiseen malliin nähden. Esimerkiksi topologista vaatimustenmukaisuutta määräävät seuraavan kaltaiset säännöt. Pisteellä voi olla vain yksi sijainti, viivat voivat leikata toisiaan vain solmun kohdalla ja polygonien on oltava suljettuja. Säännöt voivat koskea myös ominaisuustietoa, jolloin Jyväskylä ei voi sijaita Varsinais-Suomen maakunnassa. Eheä paikkatieto ei kuitenkaan takaa tarkkuutta, sillä Pyhäjärviä voi olla useassa kunnassa. Eheys on siis käsitettävä paikkatiedon sisäisenä oikeellisuutena.

Paikkatiedon laadun moniulotteisuus syntyy maantieteellisen tiedon ulottuvuuksien ja laadun perukomponenttien välisistä suhteista (Veregin & Hargitai 1995). Maantieteelliset havainnot kuvaavat todellisuuden ilmiöitä ja niiden välisiä suhteita tilassa, ajassa ja ominaisuuksiltaan. Laadun peruskomponentit tarkkuus, täsmällisyys, eheys ja täydellisyys esiintyvät näissä ulottuvuuksissa.

2.3.4 Paikkatiedon laatustandardi ISO 19157

Geoinfomaatiikan alalla vakiintunut paikkatiedon laadun ontologia on Kansainvälisen standardointijärjestön tuottama standardi ISO 19157:2013 Geographic information — Data quality (ISO 19157 2013). Se määrittelee paikkatiedon laadun osatekijät ja laadunarvioinnin periaatteet. Laatu jaetaan siinä kuuteen osatekijään, laatulementteihin, jotka edelleen jaetaan alaelementteihin (taulukko 2). Standardi antaa ohjeita kvantitatiivisen ja kvalitatiivisen laadun arviointiin, dokumentointiin ja raportointiin. Se korostaa systemaattista laadunarvioinnin prosessia osana laadunhallintajärjestelmää sekä suosittelee metatietojen käyttöä laadun dokumentoimiseksi ja aineistojen vaihdon helpottamiseksi eri järjestelmien välillä. Standardi esittää paikkatiedon laadun osatekijöittäin maantieteellisen tiedon ulottuvuuksissa.

ISO 19157 (2013) -standardi esittää neljä näkökohtaa, jotka selvittämällä paikkatietotuotteen laadun kokonaisuus voidaan hahmottaa suhteessa tuottajan ja käyttäjän tarpeisiin. Näistä 1. Tarkoitus on syy aineiston luomiselle ja ilmaisee sen aiotun käyttötarkoituksen, joka voi olla eri kuin varsinainen käyttösovellus. 2. Käyttö ilmoittaa mihin sovelluksiin aineistoa on aiemmin käytetty tuottajan tai muiden käyttäjien toimesta. 3. Provenanssi kertoo aineiston elinkaaren sen nykymuotoon saakka keräämisestä koostamisen kautta mahdollisiin osa-aineistojen erotuksiin. Myös tieto suoritetuista laadunvarmistustoimenpiteistä voidaan ilmoittaa. 4. Kvantitatiivinen mittaus tuotespesifikaatiota ja laatuvaatimuksen kynnysarvoja vastaan, jotka yhdessä muodostavat kohdemaailman.

Standardissa Kohdemaailman ja spesifikaation käsitteet kuvaavat tuottajan ja käyttäjän tarpeita (ISO 19157 2013). Ne käsittävät todellisesta maailmasta paikkatietoon sisältyvät kohteet, joiden spatiaalinen, temporaalinen ja temaattinen valinta mukaan kohdemaailman malliin riippuu tuotanto- käyttötarpeiden asettamasta rajauksesta. Kohdemaailmaan määritellään esimerkiksi, millaisia objekteja paikkatietotuotteen tulee sisältää, millaisella geometrialla ne mallinnetaan, mitä temaattista tietoa niihin tulee liittyä ja kuinka ajankohtaista tiedon on oltava. (David & Fasquel 1997). Lisäksi siinä ilmaistaan näiden vaatimusten kynnysarvot, jotka täytettyään paikkatieto on vaatimusten mukaista.

Taulukko 2. Paikkatiedon laadun ontologia ISO 19157:2013-standardin mukaan. Tason yksi laatulementit sisältävät tason kaksi alaelementit. Suomalainen terminologia perustuu Geoinformatiikan sanastoon (2018).

Luokitus (Taso 1)	Alaluokitus (Taso 2)	Kuvaus
Täydellisyys	Puuttuminen Sisältyminen	Kohteiden, niiden ominaisuuksien ja suhteiden läsnäolo tai poissaolo aineistossa. Aineistosta puuttuva tieto. Aineistossa oleva ylimääräinen tieto.
Temaattinen tarkkuus	Kuvailevan ominaisuustiedon oikeellisuus Kvantitatiivisen ominaisuustiedon tarkkuus Luokittelun oikeellisuus	Kvantitatiivisten ominaisuuksien tarkkuus, kuvailevien ominaisuuksien, kohteiden luokittelun ja niiden suhteiden oikeellisuus. Kuvailevan ominaisuustiedon oikeellisuus. Kvantitatiivisen ominaisuuden arvon läheisyys todellisen tai todelliseksi hyväksytyyn arvon kanssa. Kohteitten ja niiden ominaisuuksien luokittelun läheisyys kohdemaailman vastaavan luokittelun kanssa.
Sijaintitarkkuus	Absoluuttinen / Ulkoinen sijaintitarkkuus Suhteellinen / sisäinen sijaintitarkkuus Hilatiedon sijaintitarkkuus	Ilmoitettujen koordinaattiarvojen läheisyys todellisten tai sellaisiksi hyväksytyjen koordinaattiarvojen kanssa. Kohteiden ilmoitetun suhteellisen sijainnin tarkkuus verrattuna kohteiden todelliseen tai sellaiseksi hyväksytyyn suhteelliseen sijaintiin. Hila-alkioiden sijainnin läheisyys alkioiden kuvaamien kohteiden todellisen tai sellaiseksi hyväksytyyn sijainnin kanssa.
Looginen eheys	Käsitteellinen eheys Arvojoukkoeheys Formaattieheys Topologinen eheys	Tietoaineiston yhdenmukaisuus tietorakenteen, ominaisuuksien ja suhteiden loogisten sääntöjen kanssa. Säännönmukaisuus käsitemallin kanssa. Säännönmukaisuus käsitemallin sääntöjen kanssa. Säännönmukaisuus arvojoukkosääntöjen kanssa. Tietoaineiston yhtäläisyys fyysisen tietomallin kanssa. Tietoaineiston vastaavuus topologisen mallin kanssa.
Ajallinen laatu	Ajan mittauksen tarkkuus Ajallinen eheys Aikavastaavuus	Kohteiden ajallisten ominaisuuksien ja ajallisten suhteiden laatu. Aikamittausten läheisyys todellisen tai sellaiseksi hyväksytyjen aikamittausten kanssa. Tapahtumien järjestyksen oikeellisuus. Tietoaineiston tietojen vastaavuus ilmoitettuun ajanhetkeen.
Käytettävyys		Perustuu käyttövaatimukseen. Laatumittareiden aggregaatti käytettävyysarviota varten.

2.3.5 Sisäinen ja ulkoinen laatu

ISO 19157 (2013) -standardi määrittelee paikkatiedon laadun tasoksi, jolla kohteen ominaiset piirteet täyttävät vaatimukset. Määritelmä pohjaa paikkatietotutkimuksessa tunnistetuihin lähestymistapoihin ymmärtää laatu sekä tuottajan että käyttäjän tarpeista käsin (Devillers & Jeansoulin 2006: 31–41). Tuottajan ja käyttäjän tarpeet liitetään käsitteisiin ‘datan laatu’ ja ‘sopivuus tarkoitukseen’. Molemmat näkökulmat ovat oleellisia pyrittäessä ymmärtämään paikkatietotuotteen kokonaislaatua (Mocnik ym. 2018).

Devillers ja Jeansoulin (2006) kutsuvat tuottajan ja käyttäjän näkökulmia sisäiseksi ja ulkoiseksi laaduksi. Sisäinen laatu on paikkatietotuotteen ominaisuuksien yhtäläisyyttä tuottajan sille asettamiin vaatimuksiin. Ulkoinen laatu ilmaisee paikkatietotuotteen sopivuutta käyttötarkoitukseen.

Sisäisen laadun ulottuvuudet ovat vakiintuneet geoinfomatiikan alalla käytettyihin laatustandardardeihin (Devillers & Jeansoulin 2006: 35–39). Standardien kehityskaaressa viimeisin ja yleisimmin käytetty ISO 19157 (2013) esittää ne kuutena laatulementtinä (taulukko 2). Täydellisyys, temaattinen tarkkuus, sijaintitarkkuus, looginen eheys ja ajallinen laatu lukeutuvat kvantitatiivisesti mitattavaan sisäiseen laatuun. Käytettävyys kuvaa ulkoista laatua.

ISO 19157 (2013) -standardin mukaan sisäiset laatuvaatimukset esitetään tuotespesifikaatiossa, ulkoiset laatuvaatimukset käyttötarpeena. Nämä edustavat tuottajan ja käyttäjän kohdemaailmoja. Auktoritatiivisessa paikkatietotuotannossa sisäisen laadunarvioinnin fyysinen vertauskohde on tuotespesifikaation vaatimusten mukainen ulkoinen paikkatietoaineisto tai todellisessa maailmassa suoritettavat kenttämittaukset.

Geoinfomatiikan tutkija Franz-Benjamin Mocnik ym. (2018) linkittävät sisäisen ja ulkoisen laadun seuraavasti. Koska datan tavoiteltu kvantitatiivinen laatutaso asetetaan ilman yhteyttä mihinkään tiettyyn käyttötarkoitukseen, sen monikäyttöisyyden maksimointi ei välttämättä takaa parasta mahdollista käytettävyyttä. Tällöin datan laatu on eroavuus optimaalisen datan käytettävyuden ja todellisen datan käytettävyuden välillä, ottaen huomioon kaikki mahdolliset käyttötarkoitukset.

Mocnik ym. (2018) määrittelevät käytettävyuden datan affordanssiksi, kun data tulkitaan ja käytetään asiayhteydessä, joka mahdollistaa tietyn käyttötavan. He perustavat käytettävyuden määritelmänsä psykologi James Gibsonin (1977) affordanssiteoriaan, josta johtuen datan

affordanssit ovat käyttäjän havaitsemia käyttömahdollisuuksia, joissa soveltuva asiayhteys antaa datalle tulkinnan mahdollisuuden. Muodostettaessa datasta informaatiota, siitä tehdyt tulkinnat johtavat erilaisiin lopputuloksiin. Kun data tulkitaan sen affordanssiin soveltuvassa asiayhteydessä, informaatiota voidaan käyttää käsillä olevan ongelman ratkaisuun.

Käytettävyys riippuu siis sekä datasta että sen tulkinnasta, jolloin saman tuotteen laatu ei ole absoluuttista, vaan ilmenee eritasoisena eri käyttäjille. Rolls-Roycen korkeatasoinen valmistusprosessi tarjoaa todennäköisesti paremman sisäisen laadun, mutta Fordin hankintahinta ja alhaisempi kysyntä luvattomien käyttäjien keskuudessa vastaa paremmin keskivertokuluttajan tarpeisiin. Samoin esimerkiksi reitityssovelluksissa samaa verkostoa hyödyntävien autoilijan ja pyöräilijän matka-aika-arvioissa etäisyyskustannuksella on heille eri merkitys.

Sisäinen laatu kytkeytyy siis ulkoiseen laatuun niin, että se muodostaa välttämättömän, mutta ei yksinään riittävän osan arvioida aineiston sopivuutta käyttötarkoitukseen. Tarkoitukseen sopivan aineiston on edustettava sisäisen laadun perusominaisuuksiltaan riittävää laatutasoa ollakseen käytettävää. Tämän lisäksi käytettävyys vaatii laatuparametrien relevanssia aineiston käyttöyhteyhteyden ja -vaatimusten kannalta.

2.3.6 Paikkatiedon laadunhallinta

Paikkatietotuotannossa yllä käsitellyt määrittelyt, virhelähteet ja aineistomuodostuksen vaiheet ovat osa laadunhallintaa. Suomessa auktoritatiivisen paikkatiedon tuotanto ja laadunhallinta noudattavat lainsäädäntöä ja standardeja, jotka määräytyvät sekä Euroopan unionin että kansallisen lainsäädännön kautta. Käytössä on vaiheittainen ja kontrolloitu prosessi.

Euroopan unionin INSPIRE-direktiivi asettaa lähtökohdat kansalliselle paikkatietotuotantoa ohjaavalle lainsäädännölle (INSPIRE... 2014). Sen tavoitteena on varmistaa paikkatiedon löydettävyys, saatavuus, käytettävyys ja yhteentoimivuus jäsenvaltioissa. Suomen kansallisella tasolla laki paikkatietoinfrastruktuurista velvoittaa Maanmittauslaitoksen hoitamaan INSPIRE-direktiivin mukaista paikkatiedon hallintaa (421/2009... 2009).

Asetus paikkatietoinfrastruktuurista määrittää, mitkä paikkatietoaineistot kuuluvat lain soveltamisalaan (725/2009... 2009). INSPIRE-velvoitettuja aineistojen ylläpitäjiä ohjeistetaan noudattamaan direktiiviä paikkatietotuotannossaan. Kansalliseen aineistoluetteloon on kirjattu

tärkeimmät paikkatietoaineistot ja niitä hallinnoivat viranomaiset (INSPIRE velvoitetulle 2023).

Liikenneverkostoa kuvaavasta paikkatiedosta vastaavat Maanmittauslaitos, Väylävirasto ja kunnat. Maanmittauslaitos ylläpitää verkostojen geometriatietoja (Perus... 2023), kun taas Väylävirasto huolehtii ominaisuustiedosta ja aineiston koostamisesta ja jakelusta (Tietoa...2023). Kunnat ylläpitävät katujen, omistamiensa kevyen liikenteen väylien ja yksityisteiden ominaisuustietoja (Digiroad-aineiston... 2023).

Laki julkisen hallinnon tiedonhallinnasta edellyttää viranomaisia tuottamaan ja hallinnoimaan aineistonsa asianmukaisesti (906/2019... 2019). Laki korvasi JHS-suosituksia seuranneen käytännön (JHS 160... 2009; JHS-suositukset... 2023). Uusi tiedonhallintalaki velvoittaa julkiset organisaatiot ylläpitämään tiedonhallintamallia, joka kertoo, kuinka vaatimustenmukaisuus toteutuu organisaatiossa (Annala 2020).

Paikkatiedon laadunhallinta ulottuu koko tuotantoketjuun aina keräämisestä käytöstä poistamiseen saakka. Laadunhallinta seuraa etukäteen määriteltyjä prosesseja ja standardeja, kuten ISO- ja OGC-standardeja, jotka kuvavat paikkatietojen laadunarvioinnin prosessin ja vaatimukset (ISO 19157 2013; Paikkatietoalan... 2023)

Aikaisemmin viitekehyksenä käytettiin JHS 160 -suositusta (JHS 160... 2009), mutta se perustui vanhentuneisiin ISO-standardeihin, jotka on nyt päivitetty ISO 19157 (2013) -standardiksi. Tämä muutos on johtanut tarpeeseen päivittää myös JHS 160 -suositusta (Jacobsson 2015).

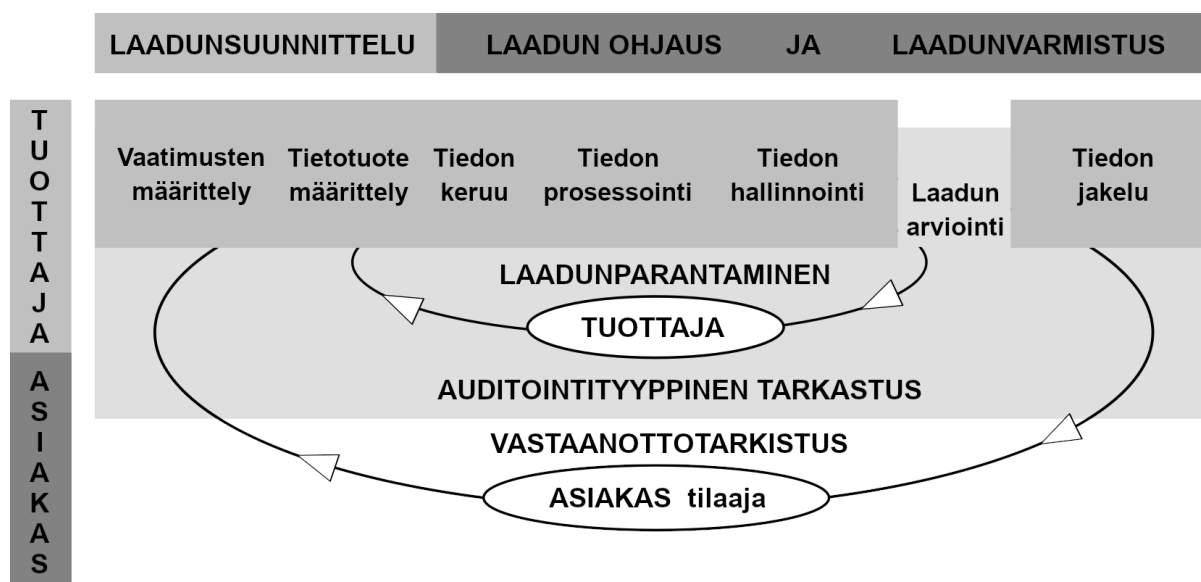
Suosituksessa JHS 160 ohjeistetaan tiedontuottajia suorittamaan vaatimustenmukainen paikkatiedon laadunarviointi ja raportointi (JHS 160... 2009). Se kuvailee paikkatiedon laadunhallinnan prosessin ja kertoo, miten sisäinen ja ulkoinen laatu asemoituvat tuotantotoiminnassa (kuva 7).

Laadunhallinta keskittyy neljään osa-alueeseen: laadun suunnitteluun, laadunohjaukseen, laadunvarmistukseen ja laadun parantamiseen (kuva 7). Suosituksessa määritellään myös kolme erilaista lähtökohtaa laadulle: asiakaskeinen, suunnittelukeskeinen ja systeemikeskeinen laatu.

Maanmittauslaitoksen tiedonhallintamalli täyttää tiedonhallintalain vaatimukset (906/2019). Tiedonhallintamallissa on määritelty, kuinka tiedon elinkaarta hallitaan ja kuinka tietoja käytetään, jaetaan ja säilytetään organisaatiossa.

Tuotetasolla laadunhallintaan liittyvät yksityiskohdat on kuvattu tuotekohtaisissa dokumenteissa, kuten Kansallisen Maastotietokannan laatukäsikirjat ja tietomallit (KMTK... 2020). Nämä ohjeistavat tuottajia seuraamaan tiukasti laadunhallinnan prosesseja paikkatietojen tuotannon aikana.

Auktoritatiivinen paikkatietotuotanto on Suomessa tarkoin säännelty prosessi, jossa laadunhallinta priorisoidaan. Osa laadunhallintaa on laadunarviointi, jota seuraavassa tarkastellaan joukkoistetun paikkatiedon laatuanalyysiin liittyen.



Kuva 7. Paikkatiedon laadun suunnittelu, ohjaus, varmistus sekä parantaminen, johon laadun arviointi tähtää, muodostavat laadunhallinnan kokonaisuuden (Jacobsson 2015).

2.4 Laadun arviointi

Alaluvussa 2.2.5 Sisäinen ja ulkoinen laatu kuvattiin paikkatiedon kokonaislaadun komponentit, jotka on huomattava erottaa alla kuvattavista ulkoisesta ja sisäisestä laatuarviosta. Sisäisen laadun todettiin olevan paikkatiedon vastaavuutta tuottajan sille asettamien vaatimusten kanssa ja ulkoisen laadun sen sopivuutta käyttötarkoitukseen käyttäjän näkökulmasta. Koska sisäisen ja ulkoisen laadun näkökulmat valaisevat samaa laadun kokonaisuutta (Devillers & Jeansoulin 2006: 31–41), voidaan käytettävyyttä arvioida

sisäisen laadun mittauksiin perustuen (Chrisman 1984). ISO 19157 (2013) -standardissa ulkoista laatua eli sopivuutta tarkoitukseen edustaa käytettävyyden laatulementti.

Paikkatiedon laatua arvioidaan ISO 19157 (2013) -standardin mukaan neljän näkökohdan perusteella. Näitä ovat 1) *tarkoitus*, jonka vuoksi aineisto luotiin, 2) aineistolla aiemmin toteutetut *käyttösovellukset*, 3) aineiston *provenanssi* ja 4) laadun osatekijöiden *kvantitatiiviset mittaukset*. Kolme ensimmäistä antavat tietoa aineiston kontekstista, jolloin arvio on *epäsuora*. Kvantitatiiviset mittaukset arvioivat aineistoa itseään ja ovat *suora* arviointitapa. Standardissa mittauksin määrällistettävät laadun osa-alueet käsitteellistetään laatulementeiksi sijaintitarkkuus, temaattinen tarkkuus, täydellisyys, ajallinen laatu, looginen eheys ja käytettävyys. Käytettävyttä lukuun ottamatta elementit jaetaan edelleen alaelementteihin (Taulukko 2).

Suora laadun arviointi voidaan toteuttaa joko *ulkoisena* tai *sisäisenä* arviointina (ISO 19157 2013). Ulkoinen arvio vertaa aineistoa toiseen, lähtökohtaisesti parempilaatuisena pidettyyn aineistoon, joka joko täyttää ennalta asetetun laatuvaatimuksen tai jonka laatutasoa pidetään suunniteltuun käyttötarkoitukseen hyväksyttävänä (Devillers & Jeansoulin 2006: 39–41).

ISO 19157 (2013) -standardin suora kvantitatiivinen arvio mittaa aineistojen välisiä eroja laatulementtikohtaisesti menetelmin, joiden laatu tuloksia voivat olla esimerkiksi lukumäärät, kokonaispituudet, pinta-alat, suhdeluvut tai keskiarvot. Loogisen eheyden vertailukohteena ei tavallisesti ole toinen aineisto, vaan arvioitavan aineiston oma looginen skeema.

Käytettävyyden laatulementin sisältö muodostetaan aineiston käyttötarkoituksesta johdettujen kriteerien mukaisesti muiden mitattujen laatulementtien aggregaattina painotuksilla tai ilman.

Joukkoistetun paikkatiedon suora vertailuun perustuva ulkoinen laatu arvio ei aina ole mahdollinen johtuen vertailuaineiston saatavuudesta, hankintakustannuksista tai lisenssirajoituksista (Fonte ym. 2017). Usein vertailua hankaloittavat puuttuvat spesifikaatiot. On myös mahdollista, että joukkoistettu aineisto on laadultaan vertailuaineistoa parempaa.

Tällöin laatua on mahdollista arvioida sisäisesti ilman vertailuaineistoa, jolloin mitataan aineiston laatua välillisesti ilmaisevia ominaispiirteitä. Näitä sisäisiä mittauskohteita ovat esimerkiksi kohteiden muokkaushistoria, maantieteelliset kontekstisidonnaiset tekijät tai kartoittajakäyttäytyminen (Antoniou ja Skopeliti 2015). Ne eivät ilmaise suhteellista laatua erona vertailuaineistoon, vaan antavat viitteen aineiston luotettavuudesta (Barron ym. 2014).

Paikkatietotutkijat Vyron Antoniou ja Andriani Skopeliti (2015) kutsuvat ISO 19157 (2013) -standardin laatulementtejä määrällistäviä menetelmiä laatumittareiksi. Tutkielmassa mitatuille laatuparametreille niitä esitetään alla täydellisyydelle, sijaintitarkkuudelle ja loogiselle eheydelle (taulukot 5, 7 ja 8).

Sisäisessä laatuarviossa analysoitavat muuttujat ovat Antoniou ja Skopelitin (2015) erittelyssä puolestaan laatuindikaattoreita, jotka jaetaan neljään luokkaan. 1)

Dataindikaattoreita ovat paikkatietokohteisiin liittyvät muuttujat, kuten pituus ja pistetiheys, versioiden lukumäärä, vakaus muutoksia vastaan, korjaukset, ja palautukset. 2) Demografisia indikaattoreita ovat väestötiheys, ikäjakauma ja kaupunkirakenne, joiden perusteella voidaan arvioida tiedon luotettavuutta. 3) Sosio-ekonomisia indikaattoreita ovat mm. deprivatio, tulotaso ja sosio-ekonominen todellisuus. Demografiset ja sosio-ekonomiset indikaattorit korreloivat positiivisesti laatuparametrien kanssa ikäjakaumaa lukuun ottamatta. 4)

Käyttäjäindikaattoreihin luetaan mm. motivaatio, koulutus, kokemus, yhteisön antama tunnustus, maine ja paikallinen tietous. Ne osoittavat paikkatiedontuottajien luotettavuutta.

Antoniou ja Skopelitin (2015) luokittelu esittää yhteenvedon kirjallisuudessa ehdotetuista laadunarvioinnin lähestymistavoista joukkoistetulle paikkatiedolle (Fonte ym. 2017). Se sisältää esimerkiksi laadunvarmistuksen tuotannon aikaiset tarkistusmekanismit 1) joukkoistetun monen tarkistajan prosessin 2) sosiaaliset menetelmät, kuten asiantuntijatason portinvartijat ja 3) maantieteelliseen kontekstiin nojaavan päättelyn (Goodchild & Li 2012). Näistä laadunvarmistuksen menettelytavoista voidaan johtaa sisäisen laadunarvioinnin lähestymistavat.

Myös geoinformatiikan tutkijat Senaratne ym. (2017) jakavat laadunarvioinnin menetelmät mittareihin ja indikaattoreihin. Heidän luokittelussaan laatumittareita ovat ISO 19157 (2013) -standardin mukaista ulkoista vertailevaa laadunarviointia seuraavat menetelmät, jotka määrällistävät vastaavuutta ja tarkkuutta. Indikaattorit puolestaan ilmaisevat kvalitatiivisia laatuparametreja, kuten paikallista tietoutta, kokemusta ja mainetta. Senaratne ym. (2017) näkevät laadun varmistuksen pääpiirteissään samoin kuin Goodchild ja Li (2012), mutta lisäävät uutena menetelmätyyppinä tiedon louhinnan.

Mocnik ym. (2018) tarkentavat laatumittarien epistemologiaan luokittelemalla mittarit niiden käyttämän informaatiolähteen mukaan (taulukko 3). Laatumittarien ankkuroidussa ontologiassaan he perustavat laatuarvion validiteetin arvioitavan aineiston, ja sitä mittaavan

laatumittarin kytkentöihin toisistaan riippumattomiin informaatiolähteisiin. Lisäksi he huomioivat vahvuuden, jolla informaatiolähteet kytkeytyvät ympäristöön.

Taulukko 3. Laatumittarien ankkurointiin perustuva ontologia. Mittarien validiteetin vahvuus perustuu niiden kytkentään paikkatiedon mallintamaa todellisuutta kuvaavaan informaatiolähteeseen. (Mocnik ym. 2018)

Kytkeä	Tyyppi	Ankkurointi
havaintoihin	ulkoinen	Havaintoperusteinen ankkurointi
dataan samasta aineistosta	sisäinen	Dataperusteinen ankkurointi
toiseen aineistoon	ulkoinen	Dataperusteinen ankkurointi
prosesoituun dataan samasta aineistosta	sisäinen	Ankkurointi prosessoituun dataan
muuhun prosesoituun dataan	ulkoinen	Ankkurointi prosessoituun dataan
sääntöihin / ... jotka saatu datasta	sisäinen	Ankkurointi sääntöihin / malleihin / tietoon
sääntöihin / ... jotka saatu havainnoista	ulkoinen	Ankkurointi sääntöihin / malleihin / tietoon

Laatumittarit ja -indikaattorit vastaavat kysymykseen mitä laadun osatekijää arvioidaan. Jako ulkoiseen ja sisäiseen laadunarviointiin kertoo mitä aineistoa arviointiin käytetään. Mocnikin ym. (2018) ontologia puolestaan vastaa kysymykseen mistä laadun arviointiin käytettävien aineistojen informaatio on peräisin.

Mocnikin ym. (2018) mukaan tietolähteiden ympäristökytkentöjen vahvuus muodostuu aineistoa luotaessa havainnoimalla ympäristöä, yhdistämällä havainnot olemassa olevaan tietoon ja esittämällä informaatio formalisoidussa muodossa. Tuotantoprosessin epätäydellisyydet heikentävät laatua, jonka ymmärtämisen olennainen edellytys on tunnistaa prosessin piirteet. Jos vertailtavien aineistojen kytkennät informaatiolähteisiin muistuttavat liian läheisesti toisiaan, jää laadunarvioinnin validiteetti heikoksi, sillä aineistot eivät tuo merkitsevää tietoa vertailuun. Mocnik ym. (2018) näkevät kaksi mahdollisuutta vahvaan laatu-arvioon. Arvio joko kytketään korkealaatuiseksi tunnettuun riippumattomaan vertailuaineistoon, tai arviointiin käytetään useita toisistaan riippumattomia sisäisiä kytkentöjä.

Valinta ulkoisen ja sisäisen laatu-arvion sekä laatumittareiden ja -indikaattorien välillä riippuu arvioinnin tavoitteista ja käytettävissä olevista resursseista. Joukkoistetun aineiston soveltuvuutta ulkoiseen tai sisäiseen laatu-arvioon voidaan arvioida geoinformaatiikan tutkija Cidália Fonten ym. (2017) esittämän kriteeristön ohjaamana (Taulukko 4). Jos laatu-elementin edellyttämät vaatimukset täyttyvät, on ISO 19157 (2013) -standardin kuvaama arviointitapa laatu-elementille mahdollinen.

Taulukko 4. ISO 19157:2013-standardin laatulementit, niiden mittaamisen esivaatimukset ja huomioitavat näkökohdat sovellettaessa joukkoistettuun paikkatietoon. (Fonte ym. 2017)

ISO-laatulementit		Vaatimukset	Haasteet VGI-sovelluksessa
Sisäinen laatu	Sijaintitarkkuus	<ul style="list-style-type: none"> • Data spesifikaatio • Vertailuaineiston olemassaolo vastaavilla ominaisuuksilla 	<ul style="list-style-type: none"> • Tiedon dynaaminen luonne • Spesifikaatioiden puute • Ei vastaavaa vertailuaineistoa • Spatiaalinen ja temaattinen heterogeenisuus
	Temaattinen tarkkuus		
	Täydellisyys		
	Ajallinen laatu		
	Looginen eheys	<ul style="list-style-type: none"> • Toinen aineisto samasta lähteestä tai riippumaton aineisto 	<ul style="list-style-type: none"> • Sovellettavissa joukkoistettuun tietoon • Mahdollistaa automaattisen validoinnin
Ulkoinen laatu	Käytettävyys	<ul style="list-style-type: none"> • Käyttövaatimukset 	<ul style="list-style-type: none"> • Voidaan arvioida yhdistämällä laatumittareita ja -indikaattoreita

Sekä ulkoinen että sisäinen arvio voidaan toteuttaa täydellisenä arviona tai otostaen (ISO 19157 2013). Täydellisessä arviossa testataan arvioitavan aineiston jokainen paikkatietokohde, otostuksessa puolestaan vain soveltuvalla tilastollisella menetelmällä valittu osa aineistosta. Pieniin populaatioihin, suorituskykymittaukseen tai automatisoitavissa oleviin testeihin ISO-standardi suosittelee sovellettavaksi täydellistä arviota. Otostamiseen perustuen voidaan puolestaan luoda todennäköisyysmalleja.

Geoinformatiikan tutkijat Longley ym. (2015: 42–43) esittävät seuraavat varaukset otostamisen soveltamiseen joukkoistetun paikkatiedon laatuarviossa. Joukkoistettu paikkatieto on usein jonkun muun keräämää ja se on koostettu jotain toista käyttötarkoitusta varten. Myös kerääjät voivat olla vaihtelevin motiivein itsevalikoitunut harrastelijoiden joukko. Aineiston muodostamisprosessit, sisältö ja kattavuus ovat usein puutteellisia, mikä kasvattaa painotusten mahdollisuutta. Vaikka painotukset havaittaisiin, otosteoria suhtautuu korjaavaan painottamiseen riskialttiina toimenpiteenä, jota käytettäessä tulosten yleistettävyys kyseenalaistuu.

ISO 19157 (2013) -standardi ohjaa määrittämään myös laatuarvion soveltamisalan. Riippuen arvion tavoitteista, se voi käsittää koko aineiston kaikki paikkatietokohteet ja kaikki mahdolliset laatuparametrit tai vain osan näistä.

2.4.1 Menetelmien esivaatimukset

Tutkielmassa arvioitujen laatuparametrien mittaamiseen käytetyt menetelmät asettavat käytettäville laatumittareille esivaatimuksia. Menetelmien valintaperusteet esitetään luvussa 3 Aineistot ja menetelmät.

Täydellisyyttä ja sijaintitarkkuutta mittaavien menetelmien täsmällinen tulos edellyttää aineistojen välistä vastaavuutta (Forghani & Delavar 2014; Tveite & Langaas 1999). Täydellisyyden mittaustulos on epätarkka, jos aineistoon sisältyvä ylimääräinen osuus ja siitä puuttuva osuus kumoavat toisensa. Samoin sijaintitarkkuuden tulokseen sisältyy virhettä, jos aineistot poikkevat täydellisyydeltään huomattavasti. Harhaanjohtavien tulosten välttämiseksi tarvitaan ylimääräisen ja puuttuvan aineiston kontrollia.

Aineistojen vastaavuus voidaan saavuttaa monivaiheisin algoritmisin tai manuaalisin menetelmin, jotka perustuvat toisiaan vastaavien pisteiden yhteensovittamiselle (Girres & Touya 2010; Seo ym. 2009; Ludwig ym. 2011). Vaihtoehtona näille geoinformatiikan tutkijat Changyong Liu ym. (2015) ehdottivat puskurointiin perustuvaa Kasvavan puskurin menetelmää, jolla kattavuuden erot aineistoissa saadaan esiin. Tapaustutkimuksessaan he katsoivat OSM-aineiston muodostamisessa käytettyjen suhteellisen tarkkojen digitointimenetelmien ja vertailuaineiston mittakaavan 1:10 000 (viiden metrin tarkkuus) sallivan menetelmän soveltamisen. Täten he olettivat vertailtavien aineistojen olevan sijaintitarkkuudeltaan lähtökohtaisesti lähellä toisiaan.

Liu ym. (2015) menetelmä on kaksivaiheinen. Ensimmäisessä vaiheessa määritetään tieverkoston leveyden vaihteluvälin minimi- ja maksimi vertailuaineistossa. Toisessa vaiheessa vertailuaineiston ympärille luodaan vaihteluvälillä asteittain jatkuvasti kasvavia puskureita, joihin sisältyvät arvioitavan aineiston kohteet lasketaan. Niiden lukumäärän ja pituuden suhteellisen osuuden kasvua koko aineistoon verrattuna seurataan mittaamalla kasvun muutosnopeutta. Kun puskuriin sisältyvien kohteiden osuus ei enää kasva, se sisältää kaikki vertailuaineiston tuntemat kohteet. Tuloksena saadun kohdejoukon perusteella ylimääräiset ja puuttuvat kohteet voidaan erottaa luotettavasti molemmista aineistoista leikkausoperaatioilla.

Geoinformatiikan tutkijoiden Hävard Tveiten ja Sindre Langaasin (1999) Päällekkäisten puskurien tunnusluvut -menetelmän esivaatimuksena on spatiaalisesta painotuksesta vapaa aineisto. Painotus voi esiintyä kohteiden koordinaateissa aineiston laajuisena systemaattisena

virheenä tai vain osassa aineistoa alueellisesti, heikentäen sijaintitarkkuuden arviota. Painotus voidaan havaita kohteiden yhdenmukaisena siirtymänä tiettyyn suuntaan. Jos siirtymä havaitaan, on koko aineistoa tai sen osaa siirrettävä virhettä vastaava etäisyys vastakkaiseen suuntaan.

Tveite ja Langaas (1999) ehdottavat painottuneisuuden mittaamiseen Oskillaation-tunnuslukua. Forghani ja Delavar (2014) puolestaan määrittivät aineistojen alueellisen jakauman keskipisteen mittaamalla niiden mediaanikeskustojen erotuksen, mitä voidaan myös käyttää painottuneisuuden havaitsemiseen.

2.4.2 Täydellisyys

Täydellisyys on paikkatietoaineistosta puuttuvien tai siihen ylimääräisinä sisältyvien kohteiden mitta. Vertailukohteena on aineiston käsitelmällin abstrakti kohdemaaailma. Toisin sanoen täydellisyys kertoo kuinka hyvin aineisto kattaa valitun teeman mukaiset todellisuuden kohteet. (Van Oort 2006: 2–6).

ISO 19157 (2013) -standardissa täydellisyyden laatulementin alakategoriat puuttuminen ja sisältyminen kuvaavat kattavuuden virheitä. Niiden läsnäoloa aineistossa mitataan datan tai tietomallin täydellisyytenä. Datan täydellisyys on paikatietokohteiden ja niiden ominaisuustiedon kattavuutta. Tietomallin täydellisyys puolestaan kuvaa aineiston tietosisällön ulottuvuuksien kykyä vastata halutun käyttötarkoituksen vaatimuksiin. (Van Oort 2006: 11–18).

ISO 19157 (2013) -standardi toteaa täydellisyyden ja tarkkuuden välillä vallitsevan voimakkaan yhteyden. Se vaikuttaa täydellisyyden arviointiin, sillä heikko temaattinen tarkkuus heikentää myös kattavuutta. Kohdemaaailmaa määriteltäessä paikkatietokohteiden luokittelun oikeellisuus on tämän vuoksi tärkeää.

Täydellisyyttä määrällistäviä laatumittareita on luokiteltu kohde- ja yksikköpohjaisiin menetelmiin (Zhou 2018). Kohteita arvioivissa menetelmissä etsitään aineistojen paikkatietokohteiden vastinpareja, kuten tielinkin segmenttejä. Lähestymistavassa sovelletaan monimutkaisia menetelmiä (Koukoletsos ym. 2012), joiden tarkkuus on osin riippuvainen digitoinnissa syntyneistä rakenteellisista eroista aineistossa (Van Niel & McVicar 2002). Esimerkiksi eri tavoin digitoitujen aineistojen segmentit saattavat samalla tieosuudella erota toisistaan lukumääräisesti tai olla eri pituisia. Yksikköpohjaiset menetelmät mittaavat tiesegmenttien pituutta (Girres & Touya 2010; Haklay 2010), eivätkä ole yhtä herkkiä

joukkoistetun paikkatiedon tuotannossa esiintyvälle digitointikäytänteiden heterogeenisyydelle kuin segmenttipohjaiset algoritmit (Van Niel & McVicar 2002).

Kaikissa alla käsitellyissä tutkimuksissa on käytetty julkisen hallinnon tuottamaa vertailuaineistoa ja moottoriliikenneteiden verkostoa lukuun ottamatta Hohmair ym. (2014) tutkimusta (Taulukko 5).

Taulukko 5. Tutkimuskirjallisuudesta tunnistetut menetelmät täydellisyyden mittaukseen.

Laadun ulottuvuus	Mittarin nimi	Mittarin kuvaus	Esimerkit
Täydellisyys	Verkoston kokonaispituus	Kohdejoukkojen segmenttien kokonaispituuksien erotus.	(Haklay 2010), (Koukoletsos et al. 2012), (Forghani & Delavar 2014), (Girres & Touya 2010), (Wu et al. 2020), (Elias et al. 2020), (Hochmair et al. 2014), (Moradi et al. 2021)
	Visuaalinen arvio	Erojen toteaminen silmämääräisesti.	(Haklay 2010), (Ciepluch 2010), (Hochmair et al. 2014)
	Vastaavien kohteiden vertailu	Kohteille vastinparit segmenttien tai solmujen piirteisiin perustuen.	(Koukoletsos et al. 2012), (Abdolmajidi et al. 2015), (Ludwig et al. 2011)
	Lukumäärä	Kohdejoukkojen lukumäärien erotus.	(Girres & Touya 2010)
	Mediaanikeskustojen etäisyys	Vertailtavien kohdejoukkojen yhteisten minimietäisyyden sijaintien erotus.	(Forghani & Delavar 2014), (Wu et al. 2020)
	Rajaavan suorakaiteen minimi	Kohteiden suorakaidekonveksien pinta-alojen summien vertailu.	(Forghani & Delavar 2014)
	Suuntauksen jakauma	Kohteiden ellipsikonveksien suuntien vertailu.	(Forghani & Delavar 2014)
	Minkovskin ulottuvuus	Verkostojen fraktaalirakenteiden eroavuus.	(Wu et al. 2020)

Vakiintunut menetelmä arvioida täydellisyyttä on verrata testi- ja vertailuaineiston tieverkostojen kokonaispituuksia (Haklay 2010; Koukoletsos ym. 2012; Forghani & Delavar 2014; Girres & Touya 2010; Wu ym. 2020; Elias ym. 2020; Hochmair ym. 2014; Moradi ym. 2021). Pituuksien erotus on yksikköpohjainen menetelmä, joka ei vaadi segmenttien yhteensovittamista eikä näin ollen ole herkkä digitoinnin eroavuuksille. Forghani ja Delavar (2014) toteavat kokonaispituuden olevan epätarkka tilanteessa, jossa sisältyvä ja puuttuva aineisto kumoavat toistensa tuloksen.

Moradi ym. (2021) tutkivat OSM-tieverkoston täydellisyyttä kokonaispituuksien erotuksella. Quebecin provinssissa täydellisyys oli aineistotasolla 32,91 %, kaupunkialueilla valtaosin 91–95 % ja maaseudulla enimmäkseen 0–15 %.

Hohmair ym. (2014) tutkivat OSM-pyöräilyverkoston täydellisyyttä kahdessa Yhdysvaltain kaupungissa vertailemalla aineistojen kokonaispituuksia. Lisäksi he käyttivät pyöräilyväylien tiheyttä arviota tukevana mittarina. Verkosto luokiteltiin katujen ulkopuolella kulkeviin pyöräilypolkuihin sekä kaduilla kulkeviin pyöräilykaistoihin (taulukko 6). Sisällyttämisen virheet johtuivat pääasiassa vääristä tageista. Huomattavaa oli, että toinen arvioitu aineisto, vastaava Google Maps -pyöräilyverkostotieto, oli OSM-aineistoa jonkin verran heikkolaatuisempaa.

Taulukko 6. OSM-pyöräilyverkoston täydellisyys Portlandissa (OR) ja Miamissa (FL).

Täydellisyys	Puuttuminen, %	Sisältyminen, %
Portland		
Polut	86,4	22
Kaistat	95,3	2,8
Miami		
Polut	22,8	1,4
Kaistat	99,2	0,1

Osassa täydellisyysarvioita on käytetty silmämääräistä analyysiä aineistojen erojen tarkasteluun (Haklay 2010; Ciepluch 2010; Hochmair ym. 2014). Tarkkaa tulosta tavoiteltaessa menetelmän heikkoutena on aikaa vievä toteutus. Visuaalinen arvio voi kuitenkin soveltua tukemaan muita menetelmiä esimerkiksi epätyypillisempien kohteiden analyysissä.

Lineaarikohteiden vastinparien vertailuun perustuvia kohdepohjaisia menetelmiä on täydellisyyden mittaamiseen kehitetty useita. Koukoletsos ym. (2012) käyttivät vastaavuuden havaitsemiseen kohteiden suuntauksia ja ominaisuustietoa. Abdolmajidi ym. (2015) sovittivat kohteita toisiinsa vertailemalla Skånen OSM-tieverkoston solmuja. He havaitsivat täydellisyyden (puuttuminen) olevan 87 % kaupunkialueilla ja 69 % maaseudulla.

Ludwig ym. (2011) ehdottivat monivaiheista vastinparien yhteensovittamista geometrisen vastaavuuden, kohdeluokan ja ominaisuustiedon perusteella. Menetelmällä mitattiin OSM-tieverkoston täydellisyyttä Saksassa. Yleisesti korkean väestötiheyden alueilla täydellisyys (puuttuminen) oli 79,8 % ja matalan tiheyden alueilla 50,8 %. Täydellisyyden (puuttuminen)

keskiarvo 15 suuressa kaupungissa oli 90,5 %. Myös muissa tutkimuksissa matalan väestötiheyden on havaittu korreloivan puutteellisen täydellisyyden kanssa (Girres & Touya 2010; Haklay 2010; Wu ym. 2020).

Girres ja Touya (2010) käyttivät täydellisyyden arvioon kohteiden lukumääriä. Menetelmää voi pitää erotuskyvyltään heikkona, sillä tielinkkien segmenttien pituudet vaihtelevat aineistojen välillä. Tämän seurauksena täydellisyydeltään toisiaan täysin vastaavien aineistojen poikkeavat segmenttien lukumäärät tulkittaisiin menetelmällä täydellisyyden puutteeksi.

Forghani ja Delavar (2014) testasivat täydellisyyden mittaamiseen usean menetelmän yhdistelmää. Teheranin OSM-tieverkoston spatiaalista epävarmuutta tutkittiin kokonaisuudella, johon kuuluivat Rajaavan suorakaiteen minimi, kohteiden suuntauksen jakauma ja mediaanikeskustojen etäisyys. Menetelmät mittasivat vertailtavien aineistojen yleistä samankaltaisuutta paikallisesti. Tutkimuksen tavoitteena oli havaita aineistojen eroavuutta, jota vakiintuneiden menetelmien erotuskyky ei tuo esiin. Myös verkostojen kokonaispituus mitattiin.

Johtopäätöksensä Forghani ja Delavar (2014) totesivat, etteivät menetelmät kerro yksinään riittävästi aineiston täydellisyydestä tulosten ollessa suuntaa antavia. Ruudukkoon jaetun tutkimusalueen soluista 80 % oli täydellisyydeltään (puuttuminen) "aika hyviä", eroten vertailuaineistosta korkeintaan 1,3 km.

Wu ym. (2020) mittasivat OSM- ja auktoritaarisen aineistojen keskimääräisiä piirteitä. Myös he käyttivät mediaanikeskustojen etäisyyttä menetelmäkokonaisuuden osana osoittamaan aineistojen yleistä samankaltaisuutta. Lisäksi he sovelsivat fraktaaligeometriaan perustuvaa Minkovskin ulottuvuutta, jossa tieverkostojen karkearakenteisuutta arvioitiin mittaamalla niiden kykyä täyttää tila.

2.4.3 Sijaintitarkkuus

ISO 19157 (2013) -standardi määrittelee sijaintitarkkuuden paikkatietokohteiden sijainnin tarkkuudeksi koordinaattijärjestelmässä (Taulukko 2). Koordinaattien sijainteja arvioidaan suhteessa kolmen tyyppiin oikeiksi hyväksytyihin koordinaatteihin. Ulkoinen sijaintitarkkuus vertailee niitä ulkoiseen vertailuaineistoon. Suhteellisen sijaintitarkkuuden vertailukohteina ovat arvioitavan aineiston muiden kohteiden koordinaatit. Hilatiedon sijaintitarkkuus on pikseliaineiston sijaintien oikeellisuutta.

Kirjallisuudessa ulkoisen sijaintitarkkuuden mittaamiseen sovelletut menetelmät jakautuvat kahteen tyyppiin (Taulukko 7). Osa menetelmistä edellyttää toisiaan vastaavien kohteiden löytämistä testi- ja vertailuaineistoista. Osa puolestaan nojaa aineistojen yleiseen samankaltaisuuteen, jolloin tarkkuutta mitataan esimerkiksi kohteiden ympärille luotujen puskureiden suhteilla tai kohteiden spatiaalisen jakauman yhteisillä keskiluvuilla.

Taulukko 7. Tutkimuskirjallisuudesta tunnistetut menetelmät ulkoisen sijaintitarkkuuden mittaukseen.

Laadun ulottuvuus	Mittarin nimi	Mittarin kuvaus	Esimerkit
Ulkoinen sijaintitarkkuus	Pisteiden etäisyys	Vastaavat pisteet etsitään algoritmilla tai manuaalisesti ja tarkistetaan manuaalisesti.	(Seo et al. 2009), (Girres & Touya 2010)
	Kohteiden vastinparit	Kohteiden vastinparit etsitään monivaiheisella algoritmilla. Tunnuksluvut lasketaan trigonometriaan perustuen, lukumääräisesti tai yksikköpohjaisesti.	(Seo et al. 2009), (Ludwig et al. 2011)
	Hausdorfin etäisyys	Kahden moniviivan pisteiden maksimietäisyys.	(Girres & Touya 2010)
	Keskimääräinen etäisyys	Murtoviivoja erottavan pinnan ja keskimääräisen pituuden suhde.	(Girres & Touya 2010)
	Linkkitarkkuus / pintaetäisyys	Kohteiden puskureiden toisiaan leikkaavan ja yhdistetyn pinta-alan suhde.	(Wu et al. 2020), (Girres & Touya 2010)
	Yksinkertainen puskuri	Mitattavan viivan pituus vertailukohteen puskurin sisällä.	(Goodchild & Hunter 1997), (Haklay 2010) ja (Jacobs 2018), (Zhang 2018)
	Mediaanikeskustojen etäisyys	Vertailtavien kohdejoukkojen yhteisten minimietäisyyden sijaintien erotus.	(Forghani & Delavar 2014)
	Päällekkäisten puskureiden tunnusluvut	Vertailtavien kohteiden keskimääräinen eroavuus niiden puskureiden pinta-alalajien suhteina.	(Elias et al. 2020), (França et al. 2018), (Tveite & Langaas 1999), (Moradi et al. 2021)
	Yhtäläisen suorakaiteen menetelmä	Kohteet vertaillaan niiden välisen pinta-alan ja yhteisten ääriviivojensa suhteellisena suorakaiteena.	(França et al. 2018)

Seo ym. (2009) vertailivat kohteiden vastienparien löytämiseen perustuvia piste- ja viivapohjaisia menetelmiä. Vastinparit havaittiin erottamalla tielinkeistä solmupisteet tarkoitukseen ohjelmoidulla automatiikalla, jonka jälkeen ne tarkastettiin visuaalisesti. Girres ja Touya (2010) käyttivät vastaavaa pistepohjaista menetelmää.

Seon ym. (2009) menetelmän viivapohjaisella versiolla oli mahdollista tutkia geometrista yhdenmukaisuutta. Tunnistetuille tielinkeille tehtiin vektori-rasteri-muunnos, jolla vältettiin

vastienparien löytämisen edellyttämä topologisen oikeellisuuden vaatimus. Vastinparit etsittiin trigonometrisillä funktioilla. Linkkien sijaintien eroavuus voitiin laskea lopuksi tehtävän rasteri-vektori-muunnoksen jälkeen. He havaitsivat viivapohjaisen menetelmän tarkkuudeltaan ja monipuolisuudeltaan pistepohjaista paremmaksi.

Hausdorffin etäisyys on pistejoukon kauimmaisen pisteen etäisyys tätä lähinnä olevaan pisteeseen toisessa pistejoukossa. Lineaarisia kohteita kohteita voi käsitellä pistejoukkoina, joten menetelmä soveltuu myös tielinkkien sijaintien välisten erojen mittaamiseen. Menetelmä kuvaa keskiarvon tai mediaanin sijaan maksimiarvoa, mikä lisää sen herkkyyttä ääriarvoille. (Girres & Touya 2010). Hausdorffin etäisyys ottaa paremmin huomioon geometrisen muodon ja sijainnin, kuin käytettäessä pistejoukkojen lähimpien pisteiden etäisyyttä (Min ym. 2007).

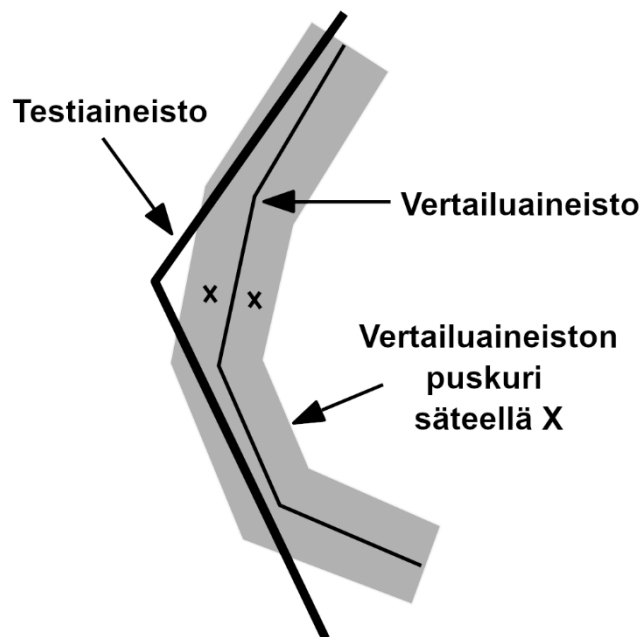
Girres & Touya (2010) vertailivat Hausdorffin etäisyyden ja keskimääräisen etäisyyden suorituskykyä sijaintitarkkuusmittauksessa. Keskimääräinen etäisyys lasketaan kahden moniviivan välille muodostuvan pinnan ja molempien viivojen pituuksien summan suhteena. Vertailu osoitti keskimääräisen etäisyyden tuottaman tuloksen vastaavan paremmin vertailuaineistoa.

Goodchildin ja Hunterin (1997) Yksinkertainen puskuri on varhainen ja vakiintunut sijaintitarkkuuden laatumittari (Haklay 2010; Jacobs 2018; Zhang 2018; kuva 8). Vertailuaineiston tiesegmentin ympärille määritetään virheen epävarmuusaluetta edustavalla säteellä puskuri, jonka sisään jäävä prosenttiosuus arvioitavan aineiston tiesegmentistä vastaa puskurisäteen osoittamaa tarkkuutta. Mittaus toistetaan kasvattamalla puskurin sädettä jokaisella mittauskerralla, jolloin saadaan sijaintitarkkuuden jakauma mittausvälillä. Menetelmä on nopeasti sovellettavissa.

Tveiten ja Langaasin (1999) menetelmässä puskureiden toisiaan leikkaavat sekä toistensa ulkopuolelle rajautuvat pinnat ilmaisevat viivojen keskimääräistä painotettua geometrista eroavuutta (kuva 9). Mitä suurempi molempien puskureiden sisällä oleva yhteinen alue on, sitä läheisemmin aineistot vastaavat toisiaan. Menetelmä ei ole herkkä ääriarvoille. Täsmällisen tuloksen edellytyksenä ovat täydellisyydeltään toisiaan vastaavat aineistot. Arvioitavassa aineistossa ei myöskään saa olla spatiaalista painotusta.

Päällekkäisten puskureiden tunnuslukuihin sisältyvät myös oskillaation, täydellisyyden ja virhekoodausten mittarit. Oskillaatio ilmaisee kuinka usein viivat leikkaavat toisiaan

puskurisäteen osoittamalla etäisyydellä. Tunnusluku antaa viitteen myös aineiston spatiaalisesta painottuneisuudesta. Täydellisyys ilmaisee kattavuutta ja Virhekoodaukset arvioitavaa aineistoa, jolle ei löydy vastaavuutta vertailuaineistosta. Täydellisyyden ja virhekoodausten tunnuslukujen käyttämä puskurointitapa muistuttaa tekniikaltaan läheisesti Goodchildin ja Hunterin (1997) menetelmää (kuva 8).

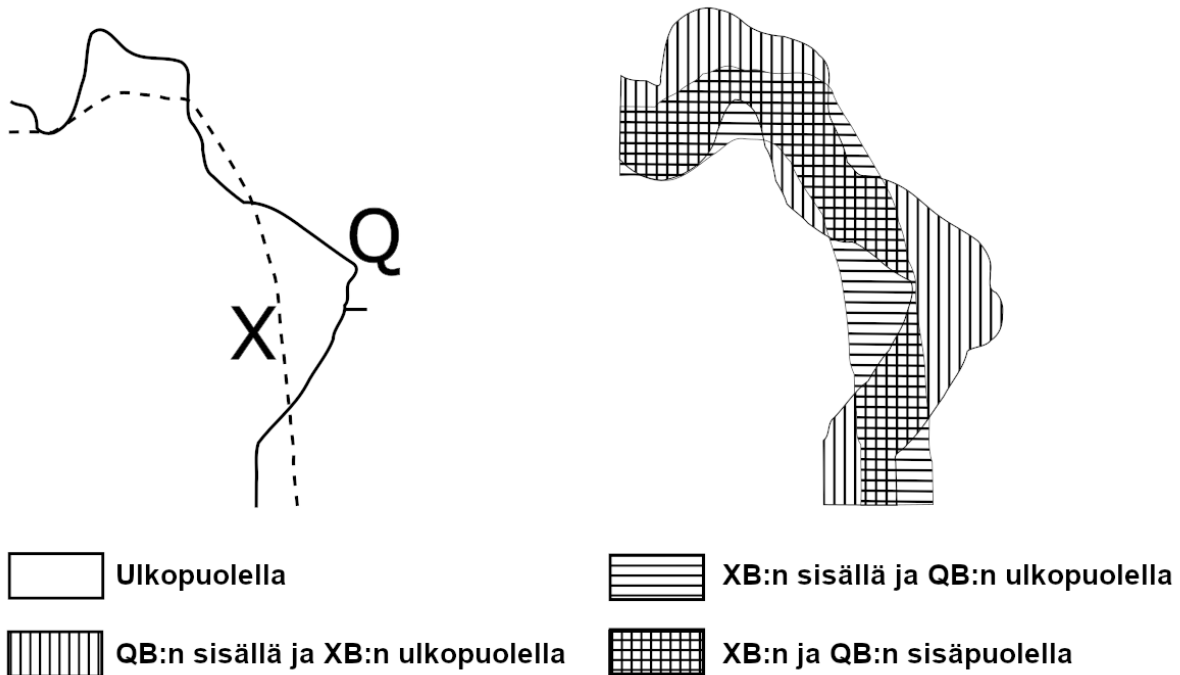


Kuva 8. Yksinkertaisen puskurin menetelmän toimintaperiaate ulkoisen sijaintitarkkuuden mittaukseen. Vertailuaineiston tiesegmentin ympärille luodaan puskurilla säteellä X. Puskurin sisältämä osa arvioitavasta testiaineistosta sijoittuu enintään säteen etäisyydelle vertailuaineistosta. Kuva muokattu alkuperäisestä (Goodchild & Hunter 1997).

Wu ym. (2020) sekä Girres ja Touya (2010) käyttivät tieverkostojen yhdenmukaisuuden mittarina Linkkitarkkuus / Pintaetäisyys -menetelmää. Yhdenmukaisuus määritetään päällekkäisten puskuritujen verkostojen yhdistetyn alan ja leikkaavan alan suhteella. Menetelmä on samankaltainen Tveiten ja Langaasin (1999) Geometrisen eroavuuden -tunnusluvun kanssa. Molemmat kuvaavat puskuripolygonien päällekkäisyyden astetta. Tveiten ja Langaasin menetelmä tarjoaa kokonaisuudessaan Linkkitarkkuutta monipuolisemmat työkalut eroavuuden mittaamiseen, sillä puskurien päällekkäisyyden pohjalta voidaan laskea myös niiden keskimääräinen eroavuus.

Forghani ja Delavar (2015) ja Wu ym. (2020) vertasivat OSM- ja vertailuaineistojen keskimääräistä eroavuutta mittaamalla mediaanikeskustojen etäisyyttä. Mediaanikeskusta on aineiston kohteiden spatiaalisen jakauman keskipiste, joka edustaa niiden yhteistä sijaintia.

Tieverkoston sovellettaessa menetelmässä lasketaan pistesijainti, joka minimoi tiesegmenttien euklidisen kokonaisuutensa kaikkiin tiesegmentteihin. Pisteellä on taipumus sijoittua alueelle, jossa on eniten kohteita. Tuloksena menetelmä antaa aineiston yleisen trendin. Yksityiskohtaisempi käsitys geometrioiden suhteista ei kuitenkaan erotu selkeästi.



Kuva 9. Ulkoisen sijaintitarkkuuden mittauksessa sovelletun Pöytäkartan menetelmän toimintaperiaate. Vertailtavat aineistot puskuroidaan. Puskureiden toisiaan leikkaavien ja toistensa ulkopuolelle rajautuvien polygonien pinta-alojen suhteet ilmaisevat sijaintitarkkuuden, huomioiden myös viivojen väliset geometriset erot. Kuva muokattu alkuperäisestä (Tveite & Langaas 1999).

Ferreiran ja Cintran (1999) mukaan pistepohjaisten sijaintitarkkuusmenetelmien haasteena on toisiaan vastaavien pisteiden valinta, joka on valintaa suorittavan asiantuntijan subjektiivinen näkemys. Vaikka otostus suoritettaisiin säännöllisin intervallein, tulokset voivat vaihdella riippuen lähtöpisteestä ja intervallien tiheydestä.

Van Niel ja McVicar (2002) toteavat puskurimenetelmien soveltuvan piste menetelmiä paremmin lineaaristen kohteiden sijaintitarkkuuden arviointiin, sillä lineaaristen segmenttien pisteitä ei päätepisteitä lukuun ottamatta useinkaan ole määritelty paikkatietokantaan riittäväällä tarkkuudella. Lisäksi he toteavat solmuihin pohjaavien menetelmien virhejakauman olevan painottunut segmenttien päätepisteisiin, mikä lisää epävarmuutta koko segmentin sijainnista. Heidän mukaansa vertailuaineistojen kohdemaailmat eivät myöskään kuvaa todellisuutta absoluuttisen tarkasti, vaan ilmoittavat sijainnin epävarmuusmarginaalin sisällä.

Siksi yksinkertaisemmat puskurimenetelmät tuottavat yleisesti ottaen luotettavana pidetyn kokonaistuloksen lineaaristen kohteiden sijaintitarkkuutta arvioitaessa.

Santosin ym. (2015) sijaintitarkkuuden mittausmenetelmien vertailuun sisältyivät mm. Yksinkertainen puskuri, Päällekkäisten puskurien tunnusluvut ja Hausdorffin etäisyys. He suosittelivat käytettäväksi Päällekkäisten puskurien tunnuslukuja sijaintitarkkuuden mittaamiseen, sillä menetelmä huomioi ääriarvot ja trendit aineistossa. Lisäksi se Santosin ym. (2015) mukaan tuottaa vastaavia tuloksia pistepareihin perustuvien menetelmien kanssa silloin, kun niissä käytetty otostus on vapaa painotuksista. Myös Mozas-Calvache (2021) suosittelee puskurimenetelmien valitsemista pisteiden ja segmenttien vastinparien vertailuun perustuvien menetelmien sijaan silloin, kun halutaan selvittää tarkkuuden tiettyä epävarmuustasoa edustavien lineaaristen kohteiden osuus.

França ym. (2018) vertailivat Yhtäläisen suorakaiteen ja Päällekkäisten puskureiden tunnuslukujen menetelmien suorituskykyä. Yhtäläisen suorakaiteen menetelmässä vertailtavien viivojen välinen ala muunnetaan suorakaiteeksi, jonka ala ja äärioviiva vastaavat alkuperäisten viivojen välistä alaa ja yhteistä äärioviivaa. Aineistojen eroavuus mitataan suorakaiteita vertaamalla. França ym. (2018) totesivat Päällekkäisten puskurien tunnuslukujen menetelmän olevan Yhtäläisen suorakaiteen menetelmää luotettavampi.

Jacobs (2018) arvioi OSM-tieverkoston sijaintitarkkuutta Ottawa-Gatineau alueella Kanadassa Yksinkertaisella puskurilla ja havaitsi 83 % aineistosta sijoittuvan alle 5 metrin säteelle vertailuaineistosta. Zhang (2018) toteutti vastaavan arvion Ontariossa Kanadassa havaiten 77 % aineistosta sijoittuvan alle 5 metrin säteelle, ja 90 % alle 30 metrin säteelle vertailuaineistosta. Edelleen Moradi ym. (2021) täyttivät Kanadan OSM-tieverkoston laatututkimuksen sijaintitarkkuusmittausten puutteita Yksinkertaisella puskurilla. He havaitsivat noin 73 % Quebecin verkostosta sijaitsevan alle 5 metrin säteellä vertailuaineistosta ja noin 60 % sijaitsevan alle 3 metrin säteellä. Kaupunkien lähialueilla sijaintitarkkuus oli muuta provinssia parempi. He havaitsivat myös väestötiheyden ja tulotason laatuindikaattorien korreloivan positiivisesti sijaintitarkkuuden kanssa.

Ludwig ym. (2011) käyttivät monivaiheista kohteiden vastinparien löytämiseen perustuvaa menetelmää OSM-tieverkostojen ja -pyöräilyverkostojen sijaintitarkkuuden arviointiin Saksassa. Tarkkuuden kokonaistuloksessa OSM-verkoston osuudet olivat 5 metrin säteellä 73 %, 5–10 metrin välillä 21 % ja 10–30 metrin välillä 6 %. Pyöräilyverkoston osuudet olivat 5 metrin säteellä 61,5 % ja 10–30 metrin säteellä 38,5 % vertailuaineistosta. Sijaintitarkkuus

oli parempi korkean väestötiheyden alueilla, mikä vastasi aiempia havaintoja (Girres & Touya; 2010; Haklayn 2010).

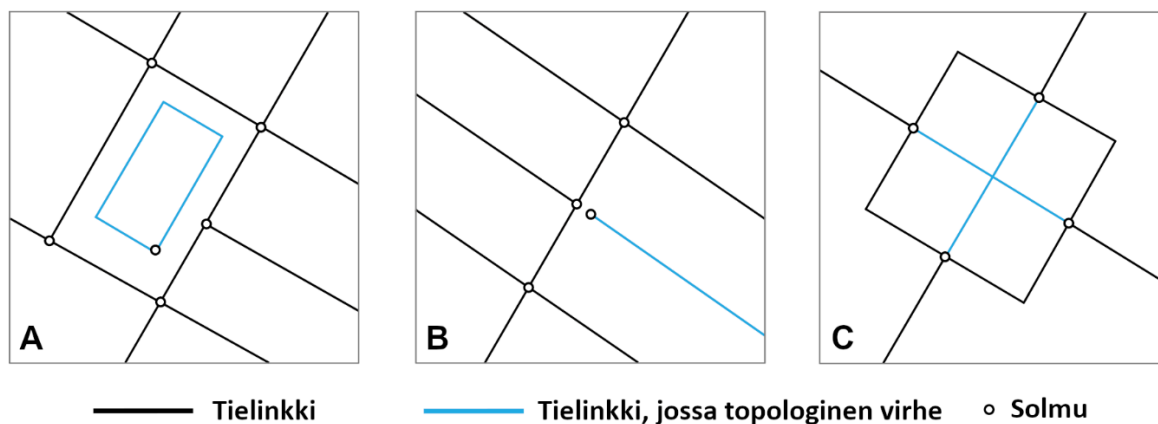
2.4.4 Looginen eheys

Looginen eheys määritellään paikkatietoaineiston kohteiden uskollisuudeksi tietorakennetta, ominaisuustietoa ja keskinäisiä suhteita kuvaaville loogisille säännöille (ISO 19157 2013; Veregin 2005). Sen neljä alaluokkaa Käsitteellinen eheys, Arvojoukkoeheys, Formaattieheys ja Topologinen eheys kuvattiin edellä esitetyssä taulukossa (Taulukko 2). Kolme ensimmäistä eivät täysin sovellu OSM-tiedon laatuarvioon (Hashemi & Abbaspour 2015), mutta topologinen eheys sen sijaan on aineiston keskeinen laatutekijä. Se korreloi positiivisesti voimakkaasti sijaintitarkkuuden ja täydellisyyden kanssa, joiden laatu tyypillisesti vaihtelee OSM-aineistossa (Barrington-Leigh & Millard-Ball 2017; Basiri ym. 2016). Täydellisyyden ja topologisten virheiden ero voi olla häilyvä, sillä puuttumisen virheet vaikuttavat topologiseen eheyteen. Brassel ym. (1995) toteavat virheen olevan topologinen, jos se ei ole johdonmukainen aineiston kokonaisuudessa. Muuten se luetaan täydellisyyden virheeksi.

Topologinen eheys arvioidaan yleensä kohteiden yhdenmukaisuutena tietomallin eheysääntöihin nähden. Tieverkoston segmenttejä koskevia sääntöjä ovat esimerkiksi “segmentit eivät saa leikata toisiaan” ja “segmentit saavat koskettaa toisiaan vain päistään”. Säännöstö muodostaa aineiston topologisen skeeman (Servigne ym. 2000). Eheyttä voidaan arvioida teeman sisäisesti tai teemojen välillä, esimerkiksi tieverkoston kohteiden keskinäisiä suhteita tai tieverkoston ja hallinnollisten rajojen keskinäisiä suhteita (Girres & Touya 2010). OSM käyttää käsitteellistä skeemaa (OSM XML 2023), mutta sen laadunvarmistus ei seuraa määriteltyä prosessia eikä vaatimustenmukaisuutta pakoteta, jolloin topologian oikeellisuus on kartoittajien huolellisuuden varassa (Sehra ym. 2020).

Tieverkosta kuvaavan OSM-aineiston topologisia virheitä on olemassa lukuisia (OSM virhetyypit 2023). Niiden yleisinä piirteinä on kirjallisuudessa tunnistettu 1) kokonaan verkostosta irralliset segmentit, 2) segmenttien vajavuus tai liiallinen pituus niiden liittymäkohdissa sekä 3) leikkauskohdassaan solmuttomat segmentit. Wu ym. (2020) tyypittelivät nämä virheluokkiin 1) kelluva saari, 2) melkein liittymä ja 3) risteys ilman liittymää (kuva 10). Kukulska-Kozieł ym. (2018) käyttivät melkein liittymistä nimitystä riipukkeet (dangles). He tunnistivat myös duplikaatteja, nollapituuksia ja mikrosegmenttejä. Sehra ym. (2020) käyttivät edeltäviä vastanneita virhetyyppejä ja jakoivat Wu ym. (2020) virheitä vastanneet poikkeamat topologisiin virheisiin ja Kukulska-Kozieł ym. (2018)

tunnistamat geometrisiin virheisiin. Niiden havaitsemiseen kirjallisuudesta tunnistetut menetelmät esitetään oheisessa taulukossa (taulukko 10)



Kuva 10. Yleisimmät topologiset virheet OSM-aineistossa: A) Kelluva saari, B) Melkein liittymä, C) Risteys ilman liittymää. Kuva on muokattu alkuperäisestä (Wu ym. 2020).

Virheiden manuaalinen korjaaminen on aikaa vievä prosessi. OSM-yhteisö rohkaiseekin topologian ohjelmalliseen laadunvarmistukseen. OSM Quality Assurance Page (Quality... 2023) listaa työkaluja topologian validointiin. Näistä monipuolisin esimerkki on Osmose, joka on aktiivisesti ylläpidetty, tunnistaa monia virhetyyppejä ja tarjoaa ohjelmointirajapinnan sekä dokumentaation. Myös tunnetuimmat paikkatieto-ohjelmistot sisältävät algoritmeja topologian tarkistamiseen, kuten QGIS ja GRASS liitännäisineen (Sehra ym. 2020) sekä ArcGIS (Kukulska-Koziel ym. 2018).

ArcGIS-menetelmät ovat vaihtoehtoihin verraten hyvin integroitu kokonaisuus. Topologian luominen, validointi, virheiden korjaaminen ja visualisointi on vakioitu ja dokumentoitu kattavasti (Creating a topology 2023). Virheiden tarkastelu -työkalusta virheet voidaan tuoda sijaintitietoineen käsiteltäväksi paikkatietoanalyysiin (Find topology errors... 2023). QGIS ei sisällä tätä ominaisuutta (Export... 2023).

Girres ja Touya (2010) validoivat ArcGIS-algoritmeilla Ranskan OSM-tieverkoston topologiaa. Teeman sisäinen yhdistävyys oli noin 95 %. Muita virheitä ei määrellistetty. Testissä aineisto todettiin reititykseen soveltumattomaksi.

Wu ym. (2020) tutkivat faktorianalyysillä täydellisyyden, sijaintitarkkuuden ja topologisen eheyden selitysvoimaa niiden yhdistetyn kokonaislaadun vaihtelussa. Elementit selittivät 71,8 % varianssista, josta elementtien osuudet olivat vastaavasti 37,2 %, 15 % ja 19,7 %.

Tutkimuskohteena oli Pittsburgh (USA) ja sitä ympäröivän alueen tieverkosto. Topologinen

eheys mitattiin ArcGIS-algoritmein. He havaitsivat topologisen eheyden olevan korkea kaupunkialueilla ja vähenevän tasaisesti kohti ympäröiviä alueita. Sen laatu oli kahta muuta elementtiä selkeästi parempi. Laadun alueellinen heterogeenisuus oli huomattava.

Taulukko 10. Tutkimuskirjallisuudesta tunnistetut menetelmät topologisen eheyden mittaukseen.

Laadun ulottuvuus	Mittarin nimi	Mittarin kuvaus	Esimerkit
Topologinen eheys	Topologiset virheet	ArcGIS-algoritmi, virheet: Erillinen saari, Melkein liittymä, Risteys ilman liittymää; menetelmät: (Servigne et al. 2000)	(Girres & Touya 2010), (Wu et al. 2020)
	Topologiset ja geometriset virheet	QGIS- ja GRASS-algoritmit, virheet: Riipukkeet, Duplikaatit, Nollapituus, Mikrosegmentit, suljetut polygonit ja solmujen järjestys	(Sehra et al. 2020), (Kukulaska-Koziet et al. 2018)
	Spatiaalinen samankaltaisuus	Kohteiden versioiden vertailu historiatiedossa.	(Hashemi & Abbaspour 2015)
	Topologiset virheet	ArcGIS-algoritmien säännöt havaitsevat segmenttien päällekkäisyydet ja puutteet solmuissa.	(Zacharopoulou et al. 2021)

Hashemi ja Abbaspour (2015) ehdottivat OSM-muokkaushistoriatietoon pohjaavaa monivaiheista menetelmää topologisen eheyden arviointiin. Mittauskohteena olivat muutokset paikkatietokohteiden suuntauksissa sekä topologisissa ja etäisyysuhteissa verrattuna muutosta edeltävään tilanteeseen. Menetelmä perustui topologisen skeeman ja vertailuaineiston sijaan aineiston sisäiseen arvioon. Menetelmädemostratio ei tuottanut tutkielman kannalta relevantteja tieverkoston topologisia laatutuloksia.

Zacharopoulou ym. (2021) laativat ArcGIS-työkalun, joka havaitsee leikkaavat viivat ja solmujen puutteet. Testi keskittyi analysoimaan ainoastaan päällekkäisyyttä. Teeman sisäisen topologian testaus antoi tulokseksi 0,01 % virheitä moottoriliikennöityjen teiden analyysissä, joka kattoi Ateenan, Berliinin, Pariisin, Utrechtin, Vienin ja Zurichin.

2.4.5 Käytettävyys

Käytettävyttä arvioidaan ISO 19157 (2013) -standardissa Käytettävyyden laatu-elementillä. Standardi kuvaa sille sekä tuottaja- että käyttäjälähtöisen sovellustavan. Tuottajan näkökulmasta Käytettävyyden elementillä voidaan ilmaista kvantitatiivisilla laatumittareilla

tuotettujen arvojen aggregaattia. Se voi olla monen yhtä laadun osatekijää mittaavan menetelmän hdistelmä tai kaikkien tutkittujen osatekijöiden yhteinen, yleistä laatutasoa edustava arvo. Mitattua laatutasoa verrataan paikkatietotuotteen tuotespesifikaatioon.

Toinen sovellustapa katsoo laatua käyttäjän näkökulmasta. Käytettävyyden laatulementillä voidaan kuvata aineiston soveltuvuutta tiettyyn käyttötarkoitukseen, jolloin kvantitatiivisten laatumittarien tuloksia painotetaan aggregaatissa kertoimin käyttötarpeeseen perustuvilla laatuvaatimuksilla. Standardi sallii liikkumavaraa Käytettävyyden laatulementin sisältövalinnassa, jolloin sen merkitys muuttuu tapauskohtaisesti.

Aggregaatti sulauttaa laadun ulottuvuuksia toisiinsa siinä määrin, että Käytettävyyden laatulementillä ilmaistu tulos on usein vertailukelpoinen vain aineiston sisäisesti. Laatuparametrien tulokset eivät erikseen erikseen indikoi soveltuvuutta käyttötarkoitukseen. Niiden yhteisvaikutuksen peilaaminen käytettävyyden kriteereihin voi kuitenkin antaa siitä käsityksen. Aggregaatin arvon vaihtelu kuvaa aineiston tasalaatuisuutta käyttötarkoitukseen.

Jotta tulos olisi ymmärrettävä, ohjeistaa ISO 19157 -standardi (2013: 144) avaamaan sen muodostamisen perusteet kattavalla erillisellä raportilla, jossa myös kvantitatiiviset laatusuoritukset ovat erillisinä arvioitavissa. Tutkielma täyttää tämän vaatimuksen kattavasti. Seikkaperäinen kuvaus parantaa myös mahdollisuuksia tehdä tarvittaessa ulkoisia vertailuja muihin samoilla perusteilla arvioituihin aineistoihin.

Sopivuutta käyttötarkoitukseen on arvioitu esimerkiksi tutkimalla OSM-aineiston soveltuvuutta reitykseen, jossa aineiston keskeisiä piirteitä ovat olleet yhdistävyys, kattavuus, (Zielsra & Willem 2014; Neis 2015). Kartoituskäyttöä varten tutkimuskohteena on ollut OSM-aineiston sijaintitarkkuus (El-Ashmawy 2016).

3 Aineistot ja menetelmät

3.1 Tutkimusalue ja paikkatietoaineistojen valinta

Tutkimusalueeksi valittiin Turun ydinkaupunkiseudun kuntayhtymän alue, johon kuuluvat Turku ja sen ympäryskunnat Naantali, Raisio, Rusko, Lieto ja Kaarina (Turun... 2022). ETRS-TM35FIN-koordinaatistossa alue rajataan koordinaatein I: 223819, P: 6702705; I: 254460, P: 6722683 (kuva 11). Valintaa perustelevat seuraavat syyt.

Turun ydinkaupunkiseudun kuntayhtymä muodostaa valtakunnallisesti suuren ja tiiviin työssäkäynti- ja asiointialueen, jolla pyöräily on huomattava liikennemuoto. Sitä tukeva infrastruktuuri on kehittynyttä ja suunnitelmallisten kehitystoimenpiteiden kohteena Maapinta-alaltaan 1 185,24 km² kokoisella alueella asuu 305 465 asukasta (Suomen... 2023; Tunnuslukuja... 2023).

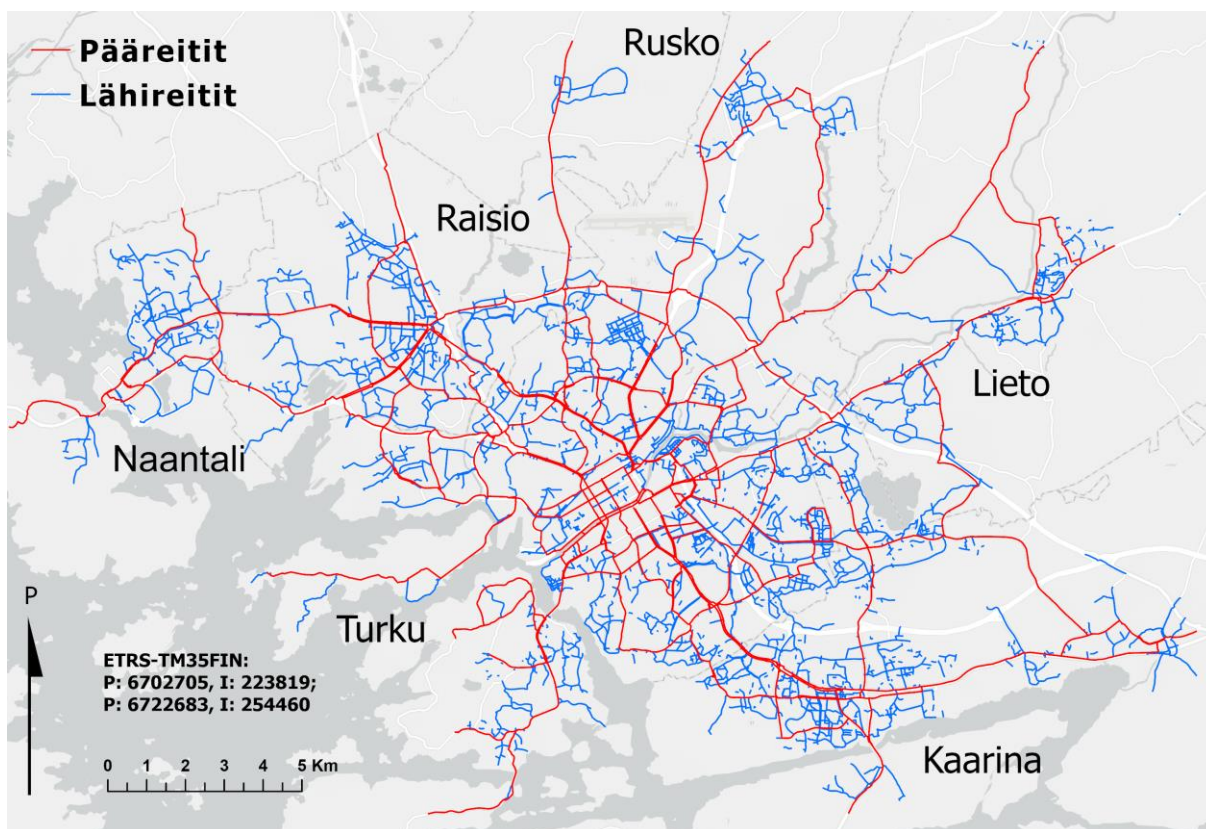
Kuntayhtymän pyöräilyverkosto on yhtenäinen toiminnallinen kokonaisuus (kuva 11), josta on saatavilla kattava digitaalinen verkostomalli. Verkoston kaksitasoisessa hierarkiassa korkeamman palvelutason pääreitit yhdistävät lähireittien erilliset osuudet toisiinsa. Liikennemerkein merkityt reitit muodostuvat erillisistä pyöräiteistä, autokaistojen oheen rakennetuista pyöräilykaistoista ja kevyen palvelutason reiteistä, esimerkiksi puistokäytävistä. Verkoston kokonaispituus on 981 km.

Pyöräily on kartoitustutkimuksen teemana perusteltu, sillä pyöräliikennöintiä pyritään valtakunnallisesti lisäämään ja Turussa sitä edistetään suunnitelmallisilla toimenpiteillä (Jääskeläinen 2018; Turun seudun... 2021; Turun pyörä... 2021). Pyöräilyn osuus Turun vuosittaisista liikennemääristä on noin neljännes, jolla se vertautuu edustavana tutkimuskohteena Tampereen ja Helsingin kaltaisten keskusten rinnalle (Valtakunnallinen... 2018).

Pyöräily on paikallinen liikennemuoto, jota käytetään lähialueella tapahtuviin työ- ja vapaa-ajanmatkoihin. Paikallinen mittakaava huomioi OpenStreetMap-palvelun perusidean paikallisen tiedon kartoituksesta (Haklay & Weber 2008). Tutkimusalueen pyöräilyverkostoa on kartoitettu OpenStreetMap-palveluun kattavasti. Lisäksi Turku ilmoittaa hyödyntävänsä kansalaispalautetta ja paikallistuntemusta kaupunkiseudun verkostomallin tuotannossa (Pärssinen 2022). Laatututkimuksen testiaineistona on OpenStreetMap-karttapalvelun pyöräilyverkostotieto (taulukko 11; kuva 12). Vertailuaineisto muodostuu kahden

auktoritatiivisen julkisen toimijan, Turun kaupungin ja Väyläviraston tuottamista paikkatietoaineistoista (taulukko 11; kuva 11).

OpenStreetMap-karttapalvelu valittiin laatututkimuksen testiaineistoksi sen pitkäaikaisen aktiivisen toiminnan, vakiintuneen aseman, tunnettuuden ja tutkimuskohteena nauttiman suosion perusteella (Yan ym. 2020; Mooney 2015; Jokar Arsanjani ym. 2015). Sitä myös hyödynnetään huomattavasti hallinnossa ja kaupallisesti (Tietoaineistojen... 2016; Who... 2023).

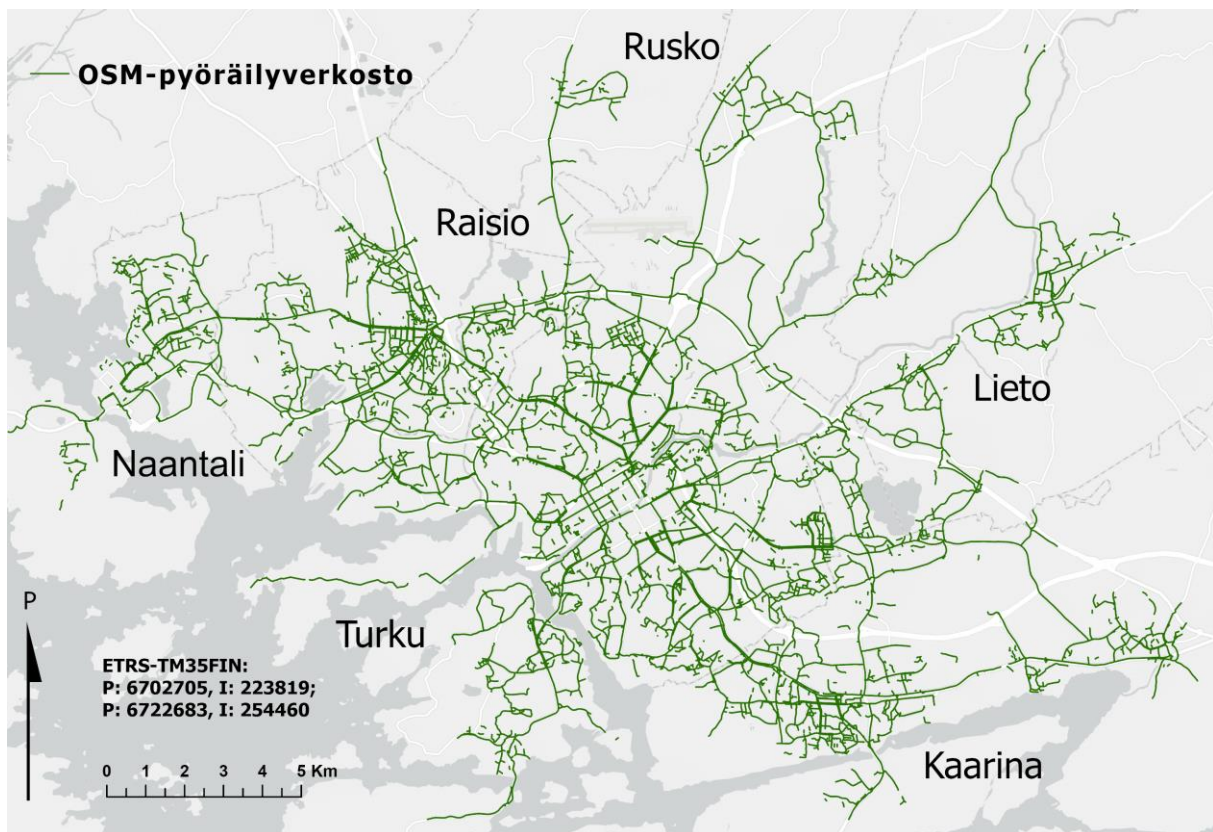


Kuva 11. Koordinaatein rajattu tutkimusalue ja vertailuaineistona oleva Turun ydinkaupunkiseudun pyöräilyverkosto. Tutkimuskohteena on pyöräilyverkoston keskeisin osa. Punainen pääverkosto yhdistää sinisen lähiverkoston osat toisiinsa. Taustakartta: Lantmäterie, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Testiaineistoa ylläpitää joukkoistetusti OpenStreetMap-yhteisö. Se on saatavilla Geofabrik GmbH:n palvelusta, jonka latausvaihtoehdot ovat koko tietokannan käsittävä tiedosto tai valtioittain rajatut osiot (Ramm 2021; Shapefiles 2022; Download... 2022). OSM-pyöräilyverkostotieto sisältää erilliset pyörätiet, katujen yhteyteen liitetyt pyöräkaistat ja pyöräilylle liikennemerkein osoitetut muut väylät.

Vertailuaineistoksi valittiin Turun kaupunkiseudun kuntayhtymän yhteinen pyöräilyverkostotieto, jonka ylläpitoa koordinoi Turun kaupunki (taulukko 11; kuva 11). Aineistoon sisältyvät kunnat ovat Turku, Naantali, Raisio, Rusko, Lieto, Kaarina, Parainen, Sauvo, Paimio, Masku, Aura, Nousiainen ja Mynämäki. Se on kattavin ja ajantasaisin saatavilla oleva auktoritatiivinen mallinnus tutkimusalueen pyöräilyverkostosta.

Turun Liikennesuunnittelutoimistos ilmoittaa, että vertailuaineiston pohjana on Digiroad-paikkatietokanta, jota kunnat ylläpitävät pyöräilyväylien osalta (Pärssinen 2022). Kunnat lisäävät Turun kaupunkiseudun pyöräilyverkostoaineistoon oman verkostonsa mukaisia muutoksia, jotka poikkeavat Digiroad-aineistosta. Turun Liikennesuunnittelu ilmoittaa, ettei aineisto kuvaa vain pyöräteitä ja -kaistoja, vaan malliin kuuluu vaihtelevasti muita kohteita, esimerkiksi pihakatuja, kävelykatuja, tonttikatuja, puistokäytäviä, huoltoteitä ja jopa ”metsäteitä”. Pyöräilyväylät kuvataan keskilinjageometrialla.



Kuva 12. Testiaineistona on Turun ydinkaupunkiseudun tutkimusalueelle rajattu OpenStreetMap-pyöräilyverkosto. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Aineisto on saatavilla Avoindata.fi-palvelusta kahtena erillisenä lineaarikohdejoukkona ArcGIS Shapefile -tiedostomuodossa, jotka kuvaavat verkoston pää- ja lähireittejä (Turun...

2021a; Turun... 2021b; Turun pyöräilyn... 2018: 15). Paikkatietokohteen tunnistetta lukuun ottamatta Turun pyöräilyverkoston geometriassa ei ole ominaisuustietoa.

Vertailuaineistoon valittiin myös Väyläviraston ylläpitämä valtakunnallinen Digiroad-tiestöaineisto (taulukko 11). Sen sisältämällä ominaisuustiedolla pyöräilyverkoston yli- ja alikuluista täydennetään Turun aineistoa, josta kyseinen tieto puuttuu. Aineisto on saatavissa kansallisella ja maakuntatasoilla. Se on ladattavissa Väyläviraston WFS-rajapinnalta (2022).

Paikkatietoanalyysiin ja karttojen laatimiseen käytettiin työssä ArcGIS Pro 3.1.0-ohjelmaa. Pohjakarttana käytetään Esrin World light gray base (2012) -karttaa. Esri edellyttää kuvateksteihin liitetyn lähdetiedon lisäksi seuraavan maininnan lisäämistä julkaisuihin (What is the correct... 2023):

Kartat tässä julkaisussa on luotu käyttämällä Esrin ArcGIS® -ohjelmistoa. Esrillä on immateriaalioikeus ArcGIS®- ja ArcMap™ -tuotteisiin ja niitä käytetään julkaisussa lisenssin alaisesti. Tekijänoikeus © Esri. Kaikki oikeudet pidätetään. Lisätietoja Esri® -ohjelmistosta löydät osoitteesta www.esri.com.

Taulukko 11. Tutkimusaineistojen metatiedot.

Aineisto	Turun kaupunkiseudun pyöräilyverkosto	Digiroad-ylikulktieto	OSM-pyöräilyverkosto
Ylläpitäjä	Turun kaupunki	Väylävirasto, (pyörätiet: kunnat)	OpenStreetMap-yhteisö
Formaatti	Viivavektori, ShapeFile	Viivavektori, WFS/XML	Viivavektori, ShapeFile
Koordinaatisto	ETRS89-TM35FIN	ETRS89-TM35FIN	WGS84
Tuotantomuoto	Auktoritatiivinen, keskitetty	Auktoritatiivinen, keskitetty	Joukkoistettu, hajautettu
Väyläluokat	Pyörätiet, -kaistat, muut merkityt reitit	Pyörätiet	Pyörätiet, -kaistat, muut merkityt reitit
Saatavuus	Avoindata.fi	Väylävirasto WFS-rajapinta	Geofabrik GmbH
Päivitys	Tarpeen mukaan, päivitetty 7.1.2022	Jatkuva, aineisto irrotettu 25.3.2022	Jatkuva, aineisto irrotettu 25.3.2022
Käyttötarkoitus	Vertailuaineisto	Vertailuaineisto	Testiaineisto

3.2 Tutkimuksen lähestymistapa

Tutkimus eteni kuudessa vaiheessa (kuva 13). Laatuarvion kehyksenä käytettiin ISO 19157 (2013) -standardia, jota vertailuaineiston tuottajat soveltavat laadunhallinnassaan (JHS 160... 2009; KMTK... 2020). Tavoitteena oli selvittää testiaineiston suorituskyky laatuvaatimusta vastaan, johon suora ulkoinen kokonaisvertailu ja kvantitatiivinen mittaus soveltuvat parhaiten (Devillers & Jeansoulin 2006: 39–41; ISO 19157 2013, Fonte ym. 2017; Mosaz-Calvache 2021). Sisäistä laatuarviota ei sovellettu, sillä se kuvaa viimekädessä kartoituspanosten uskottavuutta ja ilmaisee luottamusta yksittäisiä kartoittajia kohtaan, eikä suhdetta tiettyyn laatutasoon (Spielman 2014; Antoniou & Skopeliti 2015). Valinnoilla varmistettiin laatuarvion validiteetti (Mocnik ym. 2018).

Testiaineistosta mitattavat laatuparametrit valittiin vertailuaineiston sisällön mahdollistamalla tavalla. Laatuparametrien mittausmenetelmät valittiin kirjallisuuskatsauksen perusteella. Menetelmien sovellustapana oli tutkia testiaineiston kohdejoukkoa kokonaisuudessaan otostamisen sijaan. Tätä perusteli joukkoistetulle paikkatiedolle ominainen tuotantoprosessin heterogeenisyys, joka lisää aineiston painottuneisuuden riskiä (Longley ym. 2015: 42–43). Kokonaistutkimus mahdollisti alueellisen laatuvariaation selvittämisen.

Aineistot esikäsiteltiin ja testiaineiston laatuarvio suoritettiin jokaiselle valitulle laatuparametrille sekä aineistotasolla että alueellisesti 1 km² ruudukossa. Alueellinen jako mahdollisti laadun heterogeenisyyden paikallisen mittauksen. Paikalliset laskelmat suoritettiin ruutukohtaisesti alla kuvatuilla laatumittareilla (taulukko 10). Lisäksi laadun kasautuminen selvitettiin etsimällä ruudukon tuloksista spatiaalisesti autokorreloivia arvoja.

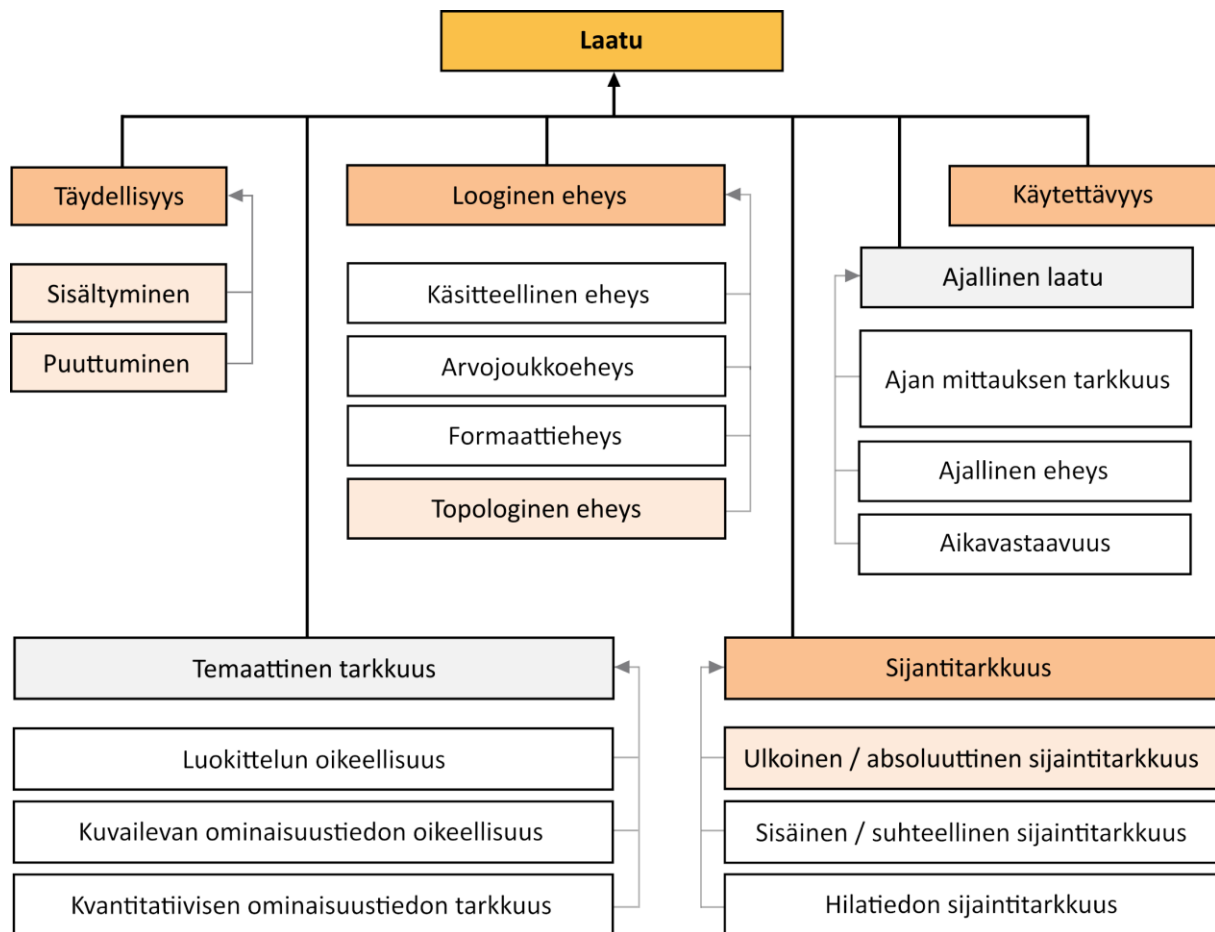
Laatumittauksen tuloksia verrattiin Maanmittauslaitoksen ylläpitämän maastotietokannan laatuvaatimukseen (Maastotietojen... 2023; Maastotietokanta... 2023). Lisäksi aineiston sopivuus käyttötarkoitukseen ilmaistiin numeerisesti painottamalla käyttötarkoituksen aineistolta edellyttämiä keskeisiä laatuparametrejä. Käytettävyys arvioitiin laatumittauksen tuloksiin, niiden vastaavuuteen laatuvaatimukseen ja käytettävyyden alueellisen vaihtelu mittaukseen perustuen. Laatuparametrien ja käytettävyysarvion tasot sekä niiden alueellinen vaihtelu, erityisesti vaihtelun keskijajonta, olivat luotettavuusarvion pohjana.



Kuva 13. Tutkimuksen kuusi vaihetta osatehtävineen. Tutkimus etenee valituille laatuparametreille soveltuvien mittausmenetelmien tunnistamisen kautta niiden esivaatimusten täyttämiseen ja edelleen varsinaiseen mittaukseen. Käytettävyysmittaus perustuu laatuarvioon. Kokonaisarvio perustuu molempiin. Vrt. taulukko 12.

3.2.1 Laatuparametrien valinta

Mittaukseen soveltuvat laatuparametrit valittiin vertailuaineiston spatiaalisen, temporaalisen ja temaattisen sisällön perusteella (Sundaram ym. 2022). Valintaa ohjasi Fonten ym. (2017) esittämä kriteeristö (taulukko 4). Turun kaupunkiseudun aineisto mahdollisti kolmen kuvassa 2. oranssilla merkityn ISO 19157 (2013: 9–10) -standardin määrittelemän laatu-elementin valinnan (kuva 14).



Kuva 14. Valitut laatuparametrit muodostavat kokonaislaadun. ISO 19157 (2013: 9) -standardin kehyksessä valitut laatu-elementit alaelementteineen on korostettu oranssilla. Täydellisyys, looginen eheys ja sijaintitarkkuus muodostavat sisäisen laatu-arvion, käytettävyysarvio kuvaa ulkoista laatua. Mukailtu alkuperäisestä kuvasta (Möisja ym. 2018).

Mitattaviksi soveltuvat tason yksi elementit ovat Täydellisyys, Looginen eheys, Sijaintitarkkuus ja Käytettävyys. Nämä koostuvat Käytettyvyyden laadullista elementtiä lukuun ottamatta tason kaksi kvantitatiivisista elementeistä, joista tutkittavaksi soveltuivat vastaavasti Sisältyminen ja Puuttuminen, Topologinen eheys sekä Ulkoinen sijaintitarkkuus. Topologinen eheys on ainut loogisen eheyden alaelementti, joka soveltuu täysin joukkoistetun paikkatiedon arviointiin, tietomallin avoimuuden vuoksi (Hashemi & Abbaspour 2015).

Ulkoisen sijaintitarkkuus puolestaan on sijaintitarkkuutta arvioitaessa luonnollinen riippumaton valinta suorituskykymittaukseen (Mocnik ym. 2018).

Laadun temaattisia ja ajallisia ulottuvuuksia ei tutkittu, sillä Turun kaupunkiseudun kuntayhtymän aineistossa ei ole ominaisuustietoa, eikä ISO-standardin mukaista ajallista informaatiota. Kuitenkin joitain temaattisia huomioita voitiin tehdä, aineistojen edustaessa OSM-yhteisön ja viranomaistahon näkemyksiä pyöräilyverkostosta.

Menetelmäkokonaisuuden lähestymistapana oli ulkoinen arvio. Siksi sisällyttämisen alaelementtin, eli vertailuaineaineistosta poikkeavan testiaineiston laatua arvioitiin vain topologian osalta.

Lisäksi menetelmäkokonaisuuteen valittiin Käytettävyyden geneerinen laatulementti, jonka soveltamistapa on valinnainen. Tutkielmassa se kuvaa mitattujen laatulementtien yhteisvaikutusta aineiston soveltuvuuteen eri käyttötarkoituksiin. Paikkatiedon laatua määrittävät sekä tuottajan että käyttäjän näkökulmat kvantitatiiviseen laatuun ja tarkoitukseen sopivuuteen (Devillers & Jeansoulin 2006: 31–41; Mocnik ym. 2018).

3.2.2 Laatumittarien valinta

Arvioinnille mahdollisten laatuparametrien mittaamiseen käytettävät menetelmät valittiin kartoittavaan, ei-systemaattiseen kirjallisuuskatsaukseen perustuen. Aiemmasta tutkimuksesta pyrittiin tunnistamaan vakiintuneimmat menetelmät. Valinta rajattiin lineaarisia kohteita ulkoisesti ISO 19157 (2013) -standardin pohjalta vertailleisiin tutkimuksiin, jotka tukivat validin suorituskykymittauksen tavoitetta (Mozas-Calvache 2021; Mocnik ym. 2018).

Artikkeleita haettiin Turun Yliopiston Volter-järjestelmästä ja Google Scholar -hakukoneella. Hakutermeinä käytettiin eri yhdistelmin spatial, data, quality, OpenStreetMap, topological consistency, topology, logical consistency, completeness, accuracy, VGI, crowdsourced, fitness-for-use, assessment, ja evaluation. Lisäksi valintoja tehtiin artikkelien lähdeluetteloiden perusteella. Löydettyistä artikkeleista 21 kpl käsitteli lineaaristen kohteiden ulkoiseen laatuarvioon soveltuvia menetelmiä. Katsauksessa käsitelty, menetelmävalintoja tukeva relevantti tutkimuskirjallisuus käytiin läpi teorialuvussa 2.3 Laadun arviointi (Taulukot 5, 7 ja 8). Tutkielman menetelmävalinnat on koottu oheiseen taulukkoon ja ne perustellaan alla (taulukko 12).

Lopullinen valinta rajattiin menetelmiin, jotka eivät vaatineet kohteiden vastinparien yhteensovittamista aineistojen välillä, vaan riittävä vastaavuus oli mahdollista saavuttaa

aineistotasolla. Tätä puolsivat vertailtavien aineistojen tuotantomallien ja digitointikäytänteiden ja kartografisen yleistyksen erot ja subjektiivisen pistevalinnan välttäminen (Antoniou & Skopeliti 2015; Van Niel & McVicar 2002; Ferreira & Cintra 1999). Vastaavuus varmistettiin valitsemalla temaattisesti samankaltaiset kohdejoukot aineistoja muodostettaessa (Sundaram ym. 2022).

Taulukko 12. Tutkimuksessa sovelletut menetelmät. Laatumittareilla määrällistetään laatuparametrejä. Mittausmenetelmien esivaatimukset täytetään mittausta tukevilla menetelmillä.

Mittauskohde	Menetelmän nimi	Menetelmän kuvaus	Esimerkit
Laatumittarit			
Täydellisyys	Verkoston kokonaispituus	Kohdejoukkojen segmenttien kokonaispituuksien erotus.	(Haklay 2010), (Koukoletsos ym. 2012), (Forghani & Delavar 2014), (Girres & Touya 2010), (Wu ym. 2020), (Elias ym. 2020), (Hochmair ym. 2014), (Moradi ym. 2021)
Sijaintitarkkuus	Päällekkäisten puskureiden tunnusluvut	Vertailtavien kohteiden keskimääräinen eroavuus niiden puskureiden pinta-alojen suhteina.	(Elias ym. 2020), (França ym. 2018), (Tveite & Langaas 1999), (Moradi ym. 2021)
Topologinen eheys	ArcGIS-algoritmit	Topologiset virheet: Erillinen saari, Melkein liittymä, Risteys ilman liittymää; menetelmät: (Servigne ym. 2000)	(Girres & Touya 2010), (Wu ym. 2020)
Käytettävyys	Soveltuvuusmittaus	Sopivuus käyttötarkoitukseen. Laatuparametreittäin painotettu aggregaatti.	(ISO 19157 2013)
Mittausta tukevat menetelmät			
Spatiaalinen painottuneisuus	Mediaanikeskustojen etäisyys	Vertailtavien kohdejoukkojen yhteisten minimietäisyyden sijaintien erotus.	(Forghani & Delavar 2014), (Wu ym. 2020)
Alikulkutieto, melkein liittymät	Visuaalinen arvio	Laatumittarin tulosten varmistus silmämääräisesti.	(Haklay 2010), (Ciepluch 2010), (Hochmair ym. 2014)
Puuttuvan ja sisältyvän aineiston ero	Kasvava puskuri	Optimaalinen puskurisäde täydellisyyden komponenttien erotteluun.	(Liu ym. 2015)

Menetelmävalintoihin sisältyi sekä kohde- että mittayksikköpohjaisia menetelmiä (Zhou 2018). Sijaintitarkkuuden arvioon valittujen, keskimääräiseen eroon nojaavien puskurimenetelmien etuna oli pieni herkkyys ääriarvoille eivätkä ne vaatineet kohteiden yhteen sovittamista (Tveite & Langaas 1999).

Täydellisyyden arvioon valittu verkostojen kokonaispituuksia aineistojen välillä vertaava menetelmä ei sekään edellyttänyt kohteiden vastaavuutta (Haklay 2010). Se saavutettiin Kasvavan puskurin menetelmällä, jonka tuloksen perusteella puuttumisen ja sisällyttämisen osuudet kyettiin täydellisyyttä arvioitaessa erottamaan luotettavasti (Liu ym. 2015).

Topologisen eheyden automatisoitu mittaus, jota voidaan osin tukea manuaalisesti, on tutkimuskirjallisuudessa vakiintunut ja käytännön sanelema menetelmävalinta (Girres & Touya 2010; Kukulska-Kozieł ym. 2018; Wu ym. 2020). Mittaus suoritettiin ArcGIS-paikkatieto-ohjelman algoritmeilla (Creating... 2023).

Aineiston kartoitus- ja navigointikäytön soveltuvuusmittaus edellyttää spesifiä määrittelyä, joka rajattiin tutkielmasta pois mittauskohteiden runsauden vuoksi. Käytettävyyden mittaus toteutettiin yleisellä tasolla ISO 19157 (2013) -standardin kuvaamalla tavalla, painottaen keskeisiä laatuparametrejä. Mittauksen eksploraatiivinen tavoite oli tutkia laatuparametrien yhteisvaikutusta käyttöyhteyksissä yhdistäen tuottajan ja käyttäjän laatuäkökulmat.

3.3 Aineistojen esikäsittely

Turun kaupunkiseudun aineistosta rajattiin tutkittavaksi Turun ydinkaupunkiseudun kuntayhtymän alueelle sijoittuva pyöräilyverkosto. Rajauksen ulkopuolelle jäänyt osuus oli selvästi erillään verkoston ytimeistä ja sen verkostorakenne vähäinen.

Laatumittauksen tarkkuus varmistettiin lähtöaineiston esikäsittelyllä. Aineistojen luotettava vertailu edellytti temaattisesti samankaltaisten paikkatietokohteiden erottamista aineistoista. Topologian arviota varten Turun kaupunkiseudun aineistoa oli täydennettävä puuttuvalla ylikulktiedolla. Vertailuaineiston pää- ja lähireittien toisistaan erillisinä mallinnetut verkostot olivat mahdollinen virhelähde, joka oli poistettava ne yhdistämällä. Tutkielmassa yhdistettyjen aineistojen katsottiin edustavan auktoritatiivista vertailukohdetta, sillä viranomaiset yhdistävät lähtöaineistoja omassa tuotantoprosessissaan. Lisäksi yhdistämistoimenpiteet ovat minimivaatimus aineiston käytettävyydelle ja johdonmukaiselle laatuarviolle.

Puuttuvan ja ylimääräisen verkosto-osuuden tarkka erottaminen tuki täydellisyyden ja sijaintitarkkuuden mittausta. Sijaintitarkkuusarvion edellytyksenä olivat toisiaan täydellisyydeltään vastaavat aineistot (Tveite & Langaas 1999). Vastaavuus saavutettiin erottamalla Testiaineistosta verkoston osat, jotka poikkesivat vertailuaineistosta ylimääräisenä ja toisaalta puuttuvina osuuksina. Erottelu mahdollisti myös tarkan täydellisyyden arvioinnin (Forghani & Delavar 2014; Liu ym. 2015). Lisäksi vertailuaineistolle tuntemattoman osuuden erottaminen Testiaineistosta oli edellytys Testiaineiston arvioinnille auktoritatiivisen aineistotuotannon tietolähteenä. Sijaintitarkkuuden arvioinnin toinen ennakkoehto oli spatiaalisen painottuneisuuden poistaminen arvioitavasta aineistosta (Tveite & Langaas 1999).

3.3.1 Vertailtavien paikkatietokohteiden valinta

Digiroad-aineisto soveltuu vertailuaineistoksi Turun kaupunkiseudun aineistoa heikommin. Valittaessa "pyörätiet" ja "yhdistetyt pyörätiet ja jalkakäytävät", Digiroad-malli sisältää enemmän kohteita verkoston reuna-alueilla kuin OSM- tai Turun kaupunkiseudun aineistot. Lisäksi lähin Testiaineiston "muut merkityt pyöräilyreitit" -tyypin vastaavuus on Digiroadissa "ajopolut", joka sisältää suuren määrän merkitsemättömiä kohteita mm. metsissä. Lisäksi ajopolut eivät muuallakaan vastaa kuntien Turun kaupunkiseudun aineistoon tekemiä lisäyksiä. Ydinkeskustassa Digiroad-aineiston täydellisyys on Turun kaupungin aineistoon nähden selkeästi puutteellinen. Kaupunkiseudun aineisto kuvaa siis virallista pyöräilyinfrastruktuuria temaattisesti Digiroad-aineistoa tarkemmin.

Testiaineistossa OSM-pyöräilyverkostoa kuvataan kahdella tunnistella (tag). Pyörätiet ja pyöräkaistat merkitään tunnisteella "highway=cycleway" (Bicycle... 2022). Pyöräilylle liikennemerkein osoitettuja muita reittejä merkitään tunnisteyhdistelmällä "Highway=path" ja "bicycle=designated" (Bicycle... 2022). Edelliset vastaavat Digiroadin ja Turun kaupunkiseudun aineiston pyöräteitä ja yhdistettyjä pyöräteitä ja jalkakäytäviä. Jälkimmäiset vastaavat Turun kaupunkiseudun aineiston sisältämiä Digiroadista poikkeavia merkittyjä kohteita, kuten tonttikatuja, puistokäytäviä ja metsäteitä. Geofabrikin ArcGIS shapefile -muotoon tuottaman Testiaineiston tietomallissa nämä kolme tunnistetta on yhdistetty tiestötason ominaisuuden "fclass" arvoon "cycleway". Vertailuun valitut kohteet erotetaan Testiaineistosta tällä arvolla (Geofabrikin 2021).

3.3.2 Digiroad-ylikulikutiedon valmistelu

Vertailuaineistoa täydennettiin liittämällä Digiroad-aineiston yli- ja alikulutieto Turun kaupunkiseudun kuntayhtymän aineistoon. Topologisen risteys ilman liittymää-virheen mittaaminen edellytti tietoa verkoston yli- ja alikuluista. Turun ylläpitämä pyöräilyverkostomalli ei sisältänyt tätä tietoa, joten se hankittiin Digiroadista. Aineistosta valittiin "pyörätiet" ja "yhdistetyt pyörätiet ja jalkakäytävät", jotka erotettiin omaksi tasokseen.

3.3.3 Pää- ja lähiverkostojen yhdistäminen vertailuaineistossa

Vertailuaineiston geometria koostui pää- ja lähiverkostojen erillisistä kohdejoukoista. Laatumittaus edellytti verkostojen yhdistämistä, sillä tasojen irralliset yhtymäkohdat ovat täydellisyyden ja sijaintitarkkuuden arvioissa sekä topologian visuaalisessa tarkastelussa mahdollinen virhelähde. Ne yhdistettiin ArcGIS Pron Planarisoivi viivat -työkalulla (Planarize Lines).

3.3.4 Vertailuaineistosta poikkeavan testiaineiston erottaminen

Kirjallisuudessa yleisimmin käytetty lineaariaineistojen täydellisyyden mittari, kokonaispituuksien erotus tuotti luotettavan tuloksen, kun sisältyvä ja puuttuva osuus erotettiin toisistaan Kasvavan puskurin menetelmällä (Haklay 2010; Moradi ym 2021; Forghani & Delavar 2015; Liu ym. 2015). Tällä varmistettiin myös sijaintitarkkuuden mittauksen edellyttämä verkostojen vastaavuus (Tveite & Langaas 1999). Kasvavan puskurin menetelmän katsottiin soveltuvan vertailuasetelmaan, koska vertailuaineiston mittakaava vastasi kirjallisuudessa kuvattua tapausta 1:10 000 mittakaavasta (Liu ym. 2015).

Kasvavan puskurin menetelmällä puskuroidulla vertailuaineistolla erotettiin leikkausoperaatiolla (Clip) testiaineistosta vertailuaineistoa vastaava osuus (Liu ym. 2015). Puskurin ulkopuolelle jäänyt ylimääräinen osuus lukeutui sisältymisen luokkaan. Optimaalinen leikkausleveys selvitettiin kasvattamalla puskurin sädettä, kunnes segmenteille ei enää löytynyt vastaavuutta testiaineistossa. Tämän ilmaisi puskurin sisälle sijoittuvien segmenttien muutosnopeuden vakiintuminen. Se mitattiin segmenttien kappalemäärän ja kokonaispituuden yhteisellä osuudella testiaineistosta puskurisäteen funktiona. Optimisäteellään vertailuaineiston puskurin tavoitti kaikki vastaavat segmentit testiaineistosta, muttei ulottunut vielä vertailuaineistosta puuttuvien päälle.

Menetelmän ensimmäisessä vaiheessa selvitettiin pyöräilyverkoston minimi- ja maksimileveys. Valtaosa reiteistä on erillisiä pyöräilyreittejä omalla keskilinjallaan, kuten kapeat puistokäytävät. Niiden leveyden minimi on 2 m. Maksimissaan vaihteluväli on autoliikennöidyn tien keskilinjaan merkityillä reitillä, jolloin arvioksi soveltuivat Väyläviraston leveyssuosituksat pyöräliikenteelle maantiellä (Lehtonen ym. 2021: 20; Vaarala ym. 2020: 58–90). Täten puskurin leveyden vaihteluväli oli kahden autokaistan, kahden maantiepientareen ja vertailuaineiston tarkkuusmarginaalin yhteenlaskettu leveys $2 \times 3,75 \text{ m} + 2 \times 1,5 \text{ m} \pm 3 \text{ m}$ eli välillä $7,5 \text{ m} - 13,5 \text{ m} \approx 7-14$. Valittu vaihteluväli kattoi kaikki mahdolliset reitin leveydet kohdealueella.

Toisessa vaiheessa pyöräilyverkostomalli purettiin risteyksistä erillisiksi moniviivoiksi ja muunnettiin kaksiulotteiseksi tasoksi (Planarize all lines), jolloin yli- ja alikulut saivat risteyskohtaansa solmun. Risteysten käyttäminen jakoperusteena yksittäisten segmenttien sijasta paransi menetelmän erotuskykyä korosten muutosnopeuden vaihtelua. Segmenttejä käytettäessä muutosnopeus ei vakiintunut, sillä lyhyempiä viivoja sisältyi puskuriin tasaisesti säteen kasvaessa. Risteyksistä purkaessa menetelmä myös tunnisti haarautuvat tiet oikein, jolloin pelkästään testiaineistossa olevaa verkostoa ei sisällynyt virheenä puskuriin.

Vertailuaineistolle luotiin 16 puskuria leventäen jokaisen puskurin sädettä edeltävään verrattuna metrillä. Verkoston leveyden vaihteluväliin lisättiin kaksi metriä muutosnopeuden tarkkailun helpottamiseksi. Testiaineistosta valittiin jokaiseen puskurisäteeseen sisältyvät tiesegmentit operaatiolla Valinta sijainnin mukaan (Select by Location) ehdolla ”On valintaan käytettävän kohteen sisällä” (Are within the source layer feature), jonka käänteinen valinta sisälsi vertailuaineistolle tuntemattomat reitit. Tuloksesta taulukoitiin sädekohtainen tieosuuksien lukumäärä $n(r_i)$ ja kokonaispituus $l(r_i)$. Nämä jaettiin testiaineiston tieosuuksien T_n ja pituuden T_l kokonaismäärillä ja summattiin yhteiseksi suhdeluvuksi $P(r_i)$ (yhtälö 1).

$$P(r_i) = \frac{n(r_i)}{T_n} + \frac{l(r_i)}{T_l} \quad (1)$$

Optimaalisin säde erotettiin jakamalla suhdeluvut osin päällekkäisiin $N - M + 1$ ryhmiin, jossa ryhmän koko M on neljä. Ryhmän viimeinen säde kuului aina myös seuraavaan ryhmään. Jokaisessa ryhmässä laskettiin säteiden välillä maksimimuutos (yhtälö 2).

$$\Delta P(r_i) = p(r_{i+1}) - p(r_i) \quad i = 1, 2, \dots (N - 1) \quad (2)$$

Ryhmästä, johon sijoittui kaikkien ryhmien maksimiarvoihin verrattuna pienin maksimiarvo, laskettiin optimaalisin puskurisäde ryhmän säteiden keskiarvona yhtälöllä,

$$r_o = \frac{1}{M} \sum_{j=k}^{k+M-1} r_j \quad (3)$$

jossa k on valitun ryhmän ensimmäinen puskurisäde ja M ryhmän koko. Puskurisäteiden järjestysnumerot juoksevat välillä r_1-r_{14} (yhtälö 3).

3.3.5 Spatiaalisen painottuneisuuden analyysi

Sijaintitarkkuuden mittaamisen toinen ennakkoehto oli spatiaalisesta painottuneisuudesta vapaa testiaineisto (Tveite & Langaas 1999). Painotusta mitattiin aineistojen mediaanikeskustojen etäisyydellä (Forghani & Delawar 2014; Wu ym. 2020). Mittaus suoritettiin aineistotasolla ja paikallisesti kilometriruuduittain.

Molempien aineistojen mediaanikeskustojen välinen etäisyys laskettiin yhtälöllä, jossa X_{OSM_MC} , Y_{OSM_MC} , X_{REF_MC} ja Y_{REF_MC} merkitsevät OSM- ja vertailuaineistojen mediaanikeskustojen koordinaatteja (yhtälö 4). Pieni tulos viittaa aineistojen spatiaalisten jakaumien olevan lähellä toisiaan.

$$MCD = \sqrt{(X_{OSM_MC} - X_{REF_MC})^2 + (Y_{OSM_MC} - Y_{REF_MC})^2} \quad (4)$$

Aineistotason mediaanikeskustojen etäisyysarvo antoi aiheen tutkia painottuneisuutta paikallisesti. Paikallinen vaihtelu selvitettiin toistamalla laskenta aineistojen välillä kilometriruuduittain. Kilometriruutukohtaisten mediaanikeskustojen etäisyys- ja atsimuuttiarvojen kasautumista ja hajontaa analysoitiin globaalilla Moranin I -testillä.

Ennen laskentaa molemmat tieverkostot leikattiin (intersection) kilometriruudukolla. Leikatut tiesegmentit saivat ruutukohtaisen tunnisteiden, jolla ne yhdistettiin (dissolve) moniviivaksi.

Etäisyyden lisäksi spatiaalisesta painotuksesta kertoo aineistojen mediaanikeskustojen suuntaus toisiinsa nähden. Tätä tutkittiin luomalla läheisyystaulukko (Generate Near Table), johon lisättiin tieto testiaineiston pisteen sijoittumisesta suhteessa vertailuaineiston lähimpään pisteeseen. Atsimuuttikulman asteina saatu suuntaustieto liitettiin ominaisuustiedon liitoksella läheisyystaulukosta moniviivan tunnisteiden mukaan kilometriruudukkoon.

Pisteiden sijainnit tarkasteltiin visuaalisesti ja varmistettiin tilastollisesti. Sekä MCD- että atsimuuttiarvojen spatiaalinen autokorrelaatio testattiin kilometriruudukossa. Globaali Moranin I -testitulokset ei antanut aihetta tutkia paikallista arvojen kasautumista Anselinin Moranin I -testillä.

Painottuneisuudesta saatua käsitystä oli vielä mahdollista vahvistaa sijaintitarkkuuden mittaamisen jälkeen Päällekkäisten puskurien tunnuslukujen menetelmään sisältyvällä oskillaatioarvolla (Tweite & Langaas 1999). Menetelmä kuvataan alla alaluvussa 4.3.4 Ulkoinen sijaintitarkkuus.

3.4 Täydellisyyden mittaus

Tutkielmassa täydellisyyttä mitattiin yksikköpohjaisesti OSM- ja vertailuaineiston tieverkostojen kokonaispituuksien erotuksella (Zhou 2018; Haklay 2010; Moradi 2021). Tällä vältettiin verkostojen digitoitikäytänteiden erojen tuottama epätarkkuus (Van Neil & McVicar 2002).

Mittaus suoritettiin ArcGIS Pro -ohjelman toiminnoilla hyödyntäen vakiopaikkatieto-operaatioita ja kohdejoukkotasojen ominaisuustietotauluja sekä laskutoimitustyökalua (Calculate Field).

Täydellisyyden laatulementin kaksi alaluokkaa, *sisältyminen* (omission) ja *puuttuminen* (comission) voivat havainnoissa kumota toisensa (Forghani & Delavar 2014), joten kumpaankin lukeutuville kohteille suoritettiin erilliset mittaukset Kasvavan puskurin menetelmällä erotetuilla kohdejoukkoilla (Liu ym. 2015). Menetelmällä tuotettiin kohdejoukko, joka sisälsi testiaineiston segmentit, jotka eivät vastanneet vertailuaineistoa. Nämä kohteet luettiin sisältymisen alaluokkaan.

Puuttumisen alaluokkaan lukeutuvat kohteet erotettiin puskuroimalla testiaineisto aiemmin määritellyllä 12,5 m optimaalisella puskurilla. Aineistosta oli tätä ennen erotettu vertailuaineistolle tuntematon osuus. Vertailuaineiston tieverkostoa leikattiin (Clip) puskuroidulla testiaineistolla. Tällöin vertailuaineistosta jäljelle jäi kohdejoukko, jolle ei löydy vastaavuutta testiaineistosta.

Kohteiden yhteispituus laskettiin metreissä sekä sisältyvälle että puuttuvalle aineistolle. Molempien osuus vertailuaineistosta ilmoitettiin prosenttilukuna aineistotasolla ja kilometriruudukossa. Suhdeluku laskettiin sekä sisältymisen että puuttumisen tasoille

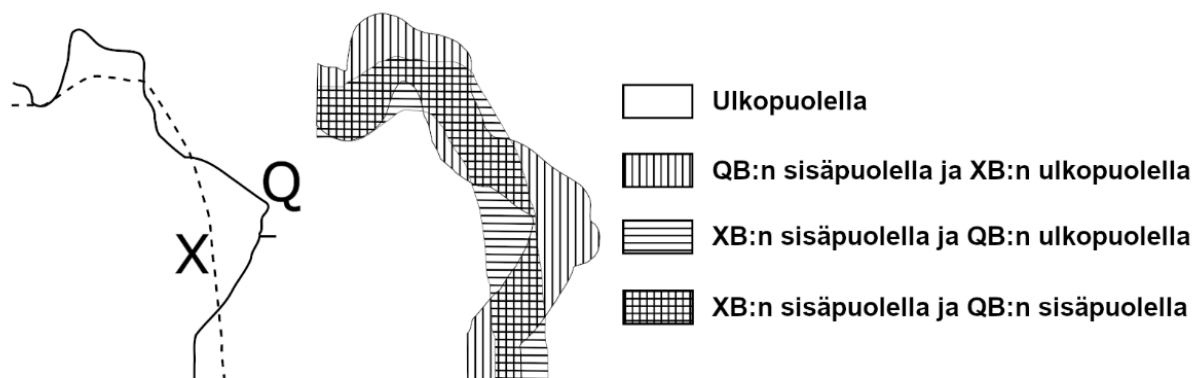
yhtälöllä, jossa RL_{OSM} on testiaineiston verkoston yhteispituus ja RL_{REF} vertailuaineiston verkoston yhteispituus (yhtälö 5).

$$Sis\lytyminen, puuttuminen = \sum RL_{OSM} / \sum RL_{REF} \quad (5)$$

Täydellisyyden tulosten spatiaalinen autokorrelaatio testattiin Anselinin paikallisella Moranin I -testillä.

3.5 Ulkoisen sijaintitarkkuuden mittaaminen

Sijaintitarkkuuden arvioon valittiin Tveiten & Langaasin (1999) Päällekkäisten puskurien tunnusluvut -menetelmäkokonaisuus seuraavin perustein. Menetelmää käyttämällä huomioitiin viivojen geometria sekä hallittiin ääriarvojen ja segmenttien solmukohtiin painottuneiden virheiden tuottamaa epätarkkuutta (Tveite & Langaas 1999; Goodchild & Hunter 1997; Van Neil & McVicar 2002). Lisäksi se soveltui tarkkuuden epävarmuustasojen selvittämiseen vertailevassa kokonaistutkimuksessa (Mozas-Calvache 2021). Menetelmän on myös osoitettu vastaavan tuloksiltaan pistepohjaisia menetelmiä (Franca 2018). Lisäksi puskurimenetelmällä vältetään tuloksiin vaikuttava toisiaan vastaavien pisteiden subjektiivinen valinta (Ferreira & Cintra 1999).



Kuva 15. Ulkoisen sijaintitarkkuuden mittauksessa sovelletun Päällekkäisten puskurien tunnuslukujen menetelmän toimintaperiaate. Vertailtavat aineistot puskuroidaan. Puskureiden toisiaan leikkaavien ja toistensa ulkopuolelle rajautuvien polygo polygonien pinta-alojen suhteet ilmaisevat sijaintitarkkuuden, huomioiden myös viivojen väliset geometriset erot. (Tveite & Langaas 1999)

Tveite & Langaas (1999) käyttivät menetelmässään epävarmuusalueen puskuria sekä vertailu-että testiviivan ympärillä. Puskureiden toisiaan leikkaavat sekä toistensa ulkopuolelle rajautuvat pinnat kuvaavat viivojen keskimääräistä eroavuutta geometrisella painotuksella. Mitä suurempi molempien puskureiden sisällä oleva yhteinen ala XB:n ja QB:n sisäpuolella

on, sitä läheisemmin aineistot vastaavat toisiaan (kuva 16). Vertailuaineiston laatuvaatimuksia vastaava 95 % todennäköisyys testiaineiston sijainnista 3 metrin ja 12,5 metrin epävarmuusalueilla arvioitiin kasvattamalla puskurin sädettä asteittain. Jos testiaineiston keskimääräinen poikkeama sijoittui vertailuaineiston epävarmuusalueelle, laatuvaatimus täytettiin. Mittaukset suoritettiin aineistotasolla ja paikallisesti kilometriruudukossa. Sijaintitarkkuuden kasautuminen mitattiin testaamalla Anselinin paikallinen Moranin I.

Tveiten & Langaasin (1999) menetelmäkokonaisuuteen sisältyy Täydellisyyden tunnusluku, jonka mittaaminen suoritettiin esitoimenpiteenä varsinaiselle sijaintitarkkuuden mittaukselle. Täydellisyyden tunnusluku vastaa läheisesti Yksinkertaisen puskurin menetelmää (Goodchildin & Hunterin 1997; kuva 8). Se myös poikkeaa toteutukseltaan huomattavasti lineaaristen kohteiden kokonaispituuteen perustuvasta täydellisyyden mittauksesta (Haklay 2010).

Menetelmässä käytettävät puskuripolygonitasot tuotettiin kolmessa vaiheessa. Metreittäin kasvavalle puskurisäteelle $pk_i, i \in \{1, 2, 3, \dots, 19\}$ suoritettiin seuraavat toimenpiteet:

Ensimmäiseksi OSM- ja vertailuaineisto, X ja Q , puskuroitiin säteellä pk_i , joka tuotti puskuroidut polygonitasot XB_i ja QB_i .

Toisessa vaiheessa tasoilla XB_i ja QB_i suoritettiin kolme päällekkäisoperaatiota, joilla tuotettiin laskennassa käytetyt polygonitasot (taulukko 13).

Taulukko 13. Päällekkäisillä puskureilla suoritetaan kolme leikkausoperaatiota menetelmän toisessa vaiheessa tulostasojen polygonien pinta-alojen laskentaa varten.

Leikkaava taso	Leikattava taso	Operaatio	Tulostaso
XB_i	QB_i	Intersection	$Ala(XB_i \cap QB_i)$
QB_i	XB_i	Erase	$Ala(XB_i \cap \overline{QB_i})$
XB_i	QB_i	Erase	$Ala(\overline{XB_i} \cap QB_i)$

Kolmanneksi laskettiin toisen vaiheen tulostasojen polygoneille pinta-alat ja lukumäärä seuraavissa tapauksissa:

- Alueet XB_i :n sisäpuolella ja QB_i :n sisäpuolella: $Ala(XB_i \cap QB_i)$
- Alueet XB_i :n sisäpuolella ja QB_i :n ulkopuolella: $Ala(XB_i \cap \overline{QB_i})$
- Alueet XB_i :n ulkopuolella ja QB_i :n sisäpuolella: $Ala(\overline{XB_i} \cap QB_i)$

Prosessi automatisoitiin Mallin rakentaja -työkalussa (Model Builder) ja laskenta suoritettiin ensin aineistojen vastaavuusehtoa täyttämättömälle testiaineistolle Täydellisyyden ja Virhekoodausten tunnuslukujen tuottamiseksi. Täydellisyyden tunnusluku laskettiin jakamalla puskuroidun testiaineiston sisäpuolelle jäävän vertailuaineiston kokonaispituus ($Pituus(Q \cap XB_i)$) testiaineiston kokonaispituudella ($Pituus(X)$) (yhtälö 6).

$$Täydellisyys(X)_i = \frac{Pituus(Q \cap XB_i)}{Pituus(X)} \quad (6)$$

Täydellisyys ilmaisi testiaineiston kattavuuden (puuttuminen) puskurisäteen leveydellä (yhtälö 6). Vastaavuusehtoa täyttämättömälle aineistolle laskettuna se antoi tuloksen, jossa oli mukana puuttumisen ja sisällymisen toisiinsa vaikuttava epätarkkuus. Täydellisyyden kuvaajan perusteella selvitettiin sijaintitarkkuuden mittaamiseen käytetyn puskurisäteen maksimi, jota tarvittiin toistettaessa laskenta täydellisyysehdon täyttävälle aineistolle.

Täydellisyyden laskennassa tarvittavat kohdejoukot tuotettiin leikkaamalla (intersect) puskuroidulla testiaineistolla (XB) puskuroimatonta vertailuaineiston tieverkostoa (X). Tulokseksi saatiin puskurin sisäpuolella oleva vertailuaineisto, jonka kokonaispituus laskettiin puskurisäteittäin.

Virhekoodausten tunnusluku ilmaisee osuuden testiaineistosta, joka ei vastaa vertailuaineistoa (sisällyminen). Tunnusluvulla voitiin varmentaa Täydellisyyden laatulementin mittauksen Sisällymisen alaelementin tulos. Virhekoodausten tunnusluku laskettiin jakamalla vertailuaineiston puekurin sisäpuolelle jäävän vertailuaineiston kokonaispituus testiaineiston kokonaispituudella (yhtälö 7).

Virhekoodausten laskennassa tarvittavat kohdejoukot tuotettiin leikkaamalla (Erase) puskuroidulla vertailuaineistolla (QB) puskuroimatonta testiaineistoa (X) puskurisäteittäin. Tulokseksi saatiin puskurin ulkopuolelle jäävä testiaineisto, jonka kokonaispituudet laskettiin puskurisäteittäin.

$$Virhekoodaukset(X)_i = \frac{Q: n\ pituus\ QB_i: n\ ulkopuolella}{Pituus(X)} \quad (7)$$

Molempien tunnuslukujen puskurisäteittäiset arvot tuotiin ArcGIS Pro:sta Excel-taulukkoon ja niistä luotiin viivakaaviot. Täydellisyyden kuvaajan perusteella valittiin puskurisäde sijaintitarkkuuden arviointita varten. Sopiva puskurisäde oli piste, jossa kuvaaja alkoi

tasaantua (Tveite & Langaas 1999; Liu ym. 2015). Testiaineiston kohdalla tasaantuminen alkoi 5 metrin lähetyvillä, mutta nousi hitaalla muutosnopeudella 24 metriin saakka. Valinta tehtiin 13 metrin kohdalla käytännöllisenä kompromissina kirjallisuuden ja tutkielman laatumittauksen tavoitteiden välillä.

Tämän jälkeen puskuri- ja päällekkäisoperaatioiden prosessi toistettiin täydellisysehdon täyttävällä aineistolla sijaintitarkkuuden selvittämiseksi aineistotasolla. Toistot tehtiin ensin välillä 0–19 m, jotta puskurisäteittäisten tulostasojen pinta-alojen vaihtelua voitiin tarkastella puskurisäteen funktiona kuvaajassa laajalla vaihteluvälillä. Pistettä, jossa $Ala(XB_i \cap QB_i)$ ei enää osoittanut merkittävää kasvua (13 m) ja saavutti 96,5 % osuuden normalisointiin käytetystä XB_i :n alasta, käytettiin toisen laskennan ylärajana puolitetulla havaintovälillä pk_i , $i \in \{0,5; 1; 1,5; \dots; 13\}$.

Verkostojen keskimääräinen eroavuus KE_i laskettiin vaihteluvälillä 0–13 m puskurisäteittäin koko aineistolle yhtälöllä, jossa pk_i on puskurisäde, $Ala(\overline{XB_i} \cap QB_i)$ alueet XB_i :n ulkopuolella ja QB_i :n sisäpuolella, $Ala(XB_i)$ testiaineiston puskuri ja i vuorossa oleva iteraatio (yhtälö 8). Tulokset tuotiin ArcGIS Pro:sta Excel-taulukkoon ja niistä luotiin viivakaaviot.

$$KE_i = \pi \cdot pk_i \cdot \frac{Ala(\overline{XB_i} \cap QB_i)}{Ala(XB_i)} \quad (8)$$

Koko aineiston kattavan analyysin jälkeen oli mahdollista selvittää tieverkoston viivojen välinen oskillaatio O_i puskurisäteittäin aineistotasolla yhtälöllä, jossa $\#A(\overline{XB_i} \cap QB_i)$ on XB_i :n ulkopuolella ja QB_i :n sisäpuolella olevien polygonien kappalemäärä ja $Pituus(X)$ testiaineiston verkoston pituus (yhtälö 9). Havainnoilla voitiin vahvistaa mediaanikeskustojen etäisyysmittauksella saatua spatiaalisen painottuneisuuden tulosta. Koko aineiston kattavaa oskillaatiota visualisoitiin viivakaaviolla.

$$O_i = \frac{\#A(\overline{XB_i} \cap QB_i)}{Pituus(X)} \quad (9)$$

Sijaintitarkkuuden paikallista arviota varten molempien aineistojen polygonitasot leikattiin (intersect) kilometriruudukolla. Näin tuotetut polygonit yhdistettiin ruutukohtaisesti moniosaisiksi kohteiksi (dissolve) ja laskenta suoritettiin ruudukkotasoon

ominaisuustietotaulukkoon. Sijaintitarkkuuden paikallisen vaihtelun mittarina käytettiin arvioitavan aineiston etäisyyden keskimääräistä eroa KE_i suhteessa vertailuaineistoon, joka laskettiin kilometriruuduittain (yhtälö 8).

Sijaintitarkkuuden tulokset varmennettiin QGIS-liitännäisellä (Tveite 2020). Liitännäisellä ei onnistuttu laskemaan koko OSM- ja vertailuaineistot käsittävää kuvaajaa polygonitasojen suhteista eikä muita muuttujia. Laskenta keskeytettiin, kun se oli kestänyt noin kaksi vuorokautta. Liitännäinen ei jostain syystä toiminut työasemassa normaalisti. Vianselvitys ohitettiin ja nopeammaksi menettelytavaksi valittiin laskentaprosessin rakentaminen manuaalisesti ja sen automatisointi ArcGIS Pro:n Mallin rakentajassa.

Hitaudestaan huolimatta QGIS-liitännäinen oli hyödyllinen työkalu varmennettaessa ArcGIS Pro:ssa laskettua keskimääräistä eroavuutta. Keskimääräinen eroavuus laskettiin kilometriruudukkoon 0–13 m vaihteluvälillä havaintopisteissä 1,5 m, 3 m, 6 m ja 12,5 m. Kilometriruudukosta valittiin neljä ruutua, joiden alueelle sijoittuneet OSM- ja vertailuaineiston tieverkostot tuotiin erillisinä ShapeFile-tiedostoina QGIS-liitännäiseen.

Laskenta suoriutui neljästä neliökilometrin kokoisesta näytteestä nopeasti. Vertailu osoitti ArcGIS- ja QGIS-laskennan tulosten olevan yhdenmukaisia kaikissa neljässä vaihteluvälillä havaintopisteessä. Pisteiden arvot vastasivat toisiaan aineistojen välillä kolmen desimaalin tarkkuudella.

3.6 Topologisen eheyden mittaus

Topologista eheyttä arvioitiin mittaamalla kolmen yleisimmän topologisen virheen esiintymistä testiaineistossa. Näitä olivat *risteykset ilman liittymää*, *melkein liittymät* ja *erilliset saaret* (Wu ym. 2020; kuva 10). Virheet kuvaavat puutteita verkoston solmuttumisessa ja katkoksia segmenttien liittymäkohdissa. Havaintojen lukumäärät, virheprosentit, sijainnit ja mahdollinen spatiaalinen autokorrelaatio selvitettiin aineistotasolla ja paikallisesti kilometriruudukossa. Virheiden kasautuminen testattiin tilastollisesti. Vertailuaineistolla tuettiin topologisten virheiden visuaalista tarkastelua. Tulokset visualisoitiin teemakartoille.

Testiaineiston topologia validoitiin työkalujen monipuolisinta automatisoitua tarjontaa edustavilla ArcGIS Pro -algoritmeilla (Creating... 2023; Find... 2023). Ohjelman paikkatietokantaan luotiin topologinen skeema, joka sisälsi virheitä erottamaan kykenevät säännöt. Ne määriteltiin OpenStreetMap-yhteisön verkkosivuilla esitettyyn kartoitus- ja

digitointiohjeistukseen perustuen (Normit... 2023; Suomi... 2023; Editing... 2023). Lisäksi sääntövalinnat pohjattiin Maastotietokannan laatuvaatimukseen (Maasto... 1995; Maastotietokanta... 2023).

Soveltuvien sääntöjen tunnistamiseksi mitattujen topologisten virheiden, Maastotietokannan laatuvaatimusten ja ArcGIS Pro -algoritmien vastaavuudet arvioitiin ja valittiin skeemaan kirjallisuuteen perustuen (Wu ym. 2020; Kukulska-Kozielec ym. 2018; taulukko 14).

Tutkittavaksi valittiin Wu ym. (2020) kuvaamat topologiset virheet, koska ne selkeästi heikentävät paikkatietotuotteen laatua valituista käytettävyyšnäkökulmista katsottuna.

Mainitut geometriset virheet eivät tuota huomattavaa haittaa kartoitus- ja uuden tiedon käyttötarkoituksissa. Ne eivät myöskään oleellisella tavalla heikennä käytettävyyttä reititys- ja navigointisovelluksissa, joissa topologisesti tärkeintä on verkoston yhdistävyys.

Taulukko 14. Topologisia ja geometrisia virheitä kuvaavien käsitteistöjen vastaavuus eri lähteissä. Mittaamiseen käytetyt algoritmit tunnistettiin laatuvaatimuksen ja tutkimuskirjallisuuden perusteella.

ArcGIS Pro -algoritmi	Maastotietokannan topologinen / geometrinen laatuvaatimus	OSM:n yleisimmat topologiset / geometriset virheet
Find Disconnected Features	Verkoston jatkuvuus	Erillinen saari (Wu ym. 2020)
Duplicate Geometry Check	Duplikaatit	Geometriset virheet (Kukulska-Kozielec ym. 2018)
Must Not Have Dangles	Liian pitkät tai lyhyet segmentit	Melkein liittymä (Wu ym. 2020)
Must Not Intersect	Viivojen solmuttuminen	Risteys ilman liittymää (Wu ym. 2020)
Must Not Have Dangles	Verkoston jatkuvuus	Riipukkeet (Kukulska-Kozielec ym. 2018)
Check Geometry	Ei geometrisiä virheitä	Geometriset virheet (Kukulska-Kozielec ym. 2018)

Paikkatietokannalle luotiin topologinen skeema (Create Topology), jota vasten Testiaineistoa arvioitiin (Validate). Tulos tutkittiin Virheiden tarkastelu -työkalussa (Error Inspector). Virrehavainnot tuotiin työkalusta sijaintitietoineen kohdejoukoksi ja tuotiin spatiaalisella liitoksella (Spatial join) kilometriruudukkoon, mikä mahdollisti spatiaalisen jakauman tutkimisen.

Erillisten saarten mittaus edellytti testiaineiston kohdejoukon muuntamista yhtenäiseksi verkstorakenteeksi ArcGIS Pron Luo geometrinen verkosto -toiminnolla (Create Geometric

Network). Katkokset verkoston yhdistävyydessä tunnistettiin työkalulla Find Disconnected Features in a Geometric Network, joka tuotti löydöksistä viivatason.

Erillisten saarten tuloksen vertailukelpoisuus varmistettiin vähentämällä testiaineiston Erillisistä saarista vertailuaineiston irralliset segmentit. Vertailuaineiston 12,5 m puskurilla leikattiin testiaineiston Erillisten saarten tasoa. Tuloksena saatiin vain testiaineistossa olevat Erilliset saaret.

Risteyksiä ilman liittymää löydettiin ”Must Not Intersect” -säännöllä 261 kpl. Tulos tuotiin kohdejoukkotasoksi Tuo topologiset virheet -työkalulla (Export Topology Errors) ja varmistettiin vertaamalla sitä visuaalisesti Digiroad-aineistosta erotettuun yli- ja alikulkutietoon sekä paikoin Google Maps -palvelussa. Todelliset risteykset ilman liittymää merkittiin virheiksi.

Kokonaismäärä ja virheprosentti Risteyksille ilman liittymää selvitettiin suorittamalla leikkausoperaatio (intersect) testiaineistolle ilman toista leikkaavaa tasoa. Tuloksena saatiin kaikkien risteävien viivojen risteyskohdat pistetasona (9476 kpl). Taso sisälsi pisteinä myös solmuttomat risteykset. Risteyskohdista vähennettiin aidot yli- ja alikulut ($261 - 172 = 89$ kpl), jolloin saatiin planaaristen risteysten kokonaismäärä (9387 kpl). Virheiden ja kokonaismäärän suhde ilmaisi aitojen solmuttomien risteysten tuloksen.

Melkein liittymä -virhetyypä validoiveen ”Must not have dangles”-säännön poikkeamiin sisältyvät virheiden lisäksi myös verkstorakenteeseen luonnollisina osina kuuluvat päätepisteet. ArcGIS Pro ja QGIS eivät tarjonneet niiden erottamiseen automatiikkaa.

Säännönvastaisia poikkeamia löydettiin 2467 kpl, joista murto-osan voitiin olettaa olevan todellisia, lähellä liitoskohtiaan sijaitsevia virheitä. Visuaaliseen tarkasteluun määrä oli epäkäytännöllisen suuri, joten tarkastelu rajattiin alle 10 metrin säteelle Melkein liittymä -poikkeamista. Tämän ylittävät etäisyydet luettiin kirjallisuudessa esitetyn näkemyksen mukaisesti täydellisyyden virheiksi (Brassel ym. 1995).

Säteen kattamalle etäisyydelle sijoittuvien virheiden sijainnit selvitettiin Virheiden tarkastelu -työkalun Snap-toiminnolla. Työkalussa asetetaan toleranssisäde, jonka sisäpuolella olevat virheet korjataan. Liitoskohdistaan vajaaksi jääneet (undershoots) ja liitoskohdat ilman solmua ylittävät (overshoots) segmentit jatketaan tai lyhennetään lähimpään toleranssisäteeseen sisäpuolella olevaan kiinnityskohtaan. Tuo topologiset virheet -työkalulla (Export Topology Errors) virheistä luotiin pistetaso sijaintitiedolla. Korjatun ja korjausta edeltäneen pistetason

spatiaalinen liitos lisäsi ominaisuustietoon käsitellyille virheille etäisyyden lähimpään korjattaessa muuttuneeseen pisteeseen. Alueellinen jakauma kyettiin esittämään erottamalla korjaukset etäisyystiedon perusteella. Havaintojen lukumäärät tuotiin Spatiaalisella liitoksella kilometriruudukkoon.

Toleranssisädettä kasvatettiin 0,5 metrin välein 10 metriin saakka ja korjattujen poikkeamien lukumäärät taulukoitiin säteittäin. Tarkasteluvälillä havaituista poikkeamista erotettiin visuaalisella tarkastelulla aidot Melkein liittymät vertailuaineistoa ja Google Maps -palvelua apuna käyttäen. Topologisten virheiden kasautumista tutkittiin testaamalla tulosten spatiaalista autokorrelaatiota Anselinin paikallisella Moranin I -testillä.

3.7 Laadun ja käytettävyyden arviointi

Laatuarviossa testiaineistosta mitattuja laatutuloksia verrattiin vertailuaineiston laatuvaatimuksen kynnsarvoihin. Käytettävyydsarviossa käyttötarkoituksille keskeisiä laatutuloksia painotettiin kertoimin. Jokaiselle käyttötarkoitukselle muodostettiin painotetuista laatutuloksista yhteinen käytettävyyttä kuvaava arvo. Laadun ja käytettävyyden alueellista vaihtelua arvioitiin visuaalisessa tarkastelussa sekä keskiarvoon ja keskihajontaan perustuen.

Tuottajanäkökulmasta hyväksyttävän laatutason kriteereiksi asetettiin Maanmittauslaitoksen Maastotietokannan kynnsarvot (taulukko 15). Valintaa perusteli Turun kaupunkiseudun aineiston pohjautuminen Digiroad-tietokantaan, joka puolestaan saa geometriatietonsa Maastotietokannasta (Torkkeli ym. 2019).

Digiroad ei ole määritellyt laatumallissaan aineiston tuottajien laatuvaatimuksia, joten Maastotietokannan laatumallin vaatimukset pätevät Maastotietokannan kynnsarvojen mukaisesti (Maastotietojen laatumalli 2023; taulukko 15). Digiroad täydentää laatumallissaan Maastotietokannan topologisen eheyden vaatimuksia seuraavalla maininnalla: Ei geometrisiä tai topologisia virheitä (duplikaatit, liian pitkät tai lyhyet segmentit) (Digiroad... 2012).

ISO 19157 (2013) -standardissa määritellyistä laatuarvion tunnusluvuista tutkielmassa käytettiin virhetason prosenttilukua (error rate) täydellisyydelle ja sijaintitarkkuudelle. Topologisen eheyden tunnuslukuna käytettiin virheiden lukumäärää (error count) (taulukko 15).

Taulukko 15. Laatuparametrien mittaustapa ja laadun arviointiin käytetyt Maastotietokannan laatuvaatimukset lineaarisille kohteille.

Laatuelementti	Mittauskohde	Tunnusluku	Laatuvaatimus
Täydellisyys	Puuttuminen	Pituuden vastaavuus	$\geq 85 \%$
Täydellisyys	Sisältyminen	Ylimääräisen aineiston osuus	$\leq 15 \%$
Sijaintitarkkuus	Laatuluokka A	Keskimääräinen poikkeama	$\leq 3 \text{ m}$
Sijaintitarkkuus	Laatuluokka B	Keskimääräinen poikkeama	$\leq 12,5 \text{ m}$
Topologinen eheys	Melkein liittymä	Lukumäärä	0 Kpl
Topologinen eheys	Risteys ilman liittymää	Lukumäärä	0 Kpl
Topologinen eheys	Erillinen saari	Lukumäärä	0 Kpl

Koko aineiston laatutasoa koskevat tulokset esitettiin virhetason tunnuslukuna. Paikalliset tulokset esitettiin koropleettikarttoina, joissa virhetason jakauma luokiteltiin kilometriruudukossa. Kartat laadittiin sekä luokitelluille jatkuville muuttujille että spatiaaliselle autokorrelaatiolle. Laatuelementtien tunnuslukujen jatkuvat muuttujat ilmaistaan desimaalina. Tulokset luokiteltiin niin, että ensimmäinen luokka vastasi laatuvaatimuksen täyttävää aineistoa. Loput luokkarajoista asetettiin käsin. Pyrkimyksenä oli selkeyttää alueellisen laatuvaihtelun eroja. Käytettävyyden suhteellista arviota varten laatuvaatimukset saivat täytyessään standardin mukaisen Boolean arvon 1 ja jäädessään puutteelliseksi arvon 0. Jokaisella laatuvaatimuksella oli kerroin, jolla boolean arvon paino Käytettävyyden laatuelementissä asetettiin. Kertoimien summa oli 1.

Käytettävyyden elementin sovellustavaksi valittiin käyttäjälähtöinen näkökulma. ISO 19157 (2013) -standardi ohjeistaa kuvaamaan elementillä laatutekijöitä, joita muut laatuelementit eivät sovellu kuvaamaan. Elementin sisällöksi asetettiin Täydellisyyden, Sijaintitarkkuuden ja Topologisen eheyden yhteisvaikutus aineiston käytettävyyteen kolmessa käyttötarkoituksessa. Nämä olivat kartoitus, reititys- ja navigaatio sekä soveltuvuus uuden tiedon lähteeksi. Tarkoitukseen soveltuvuutta oli mahdollista arvioida suuntaa antavasti mitattuihin laatuelementteihin perustuen. Laatuelementin olennaisuus käyttötarkoitukselle ilmaistaan painokertoimella (taulukko 16).

Laatuelementtien arvot painotettiin seuraavin perustein. 1) Kartoituksessa korostuvat kattavuus ja sijaintitarkkuus. 2) Reititykselle ja navigoinnille on keskeisintä verkoston yhtenäisyys. 3) Vertailuaineistosta poikkeava ylimääräinen sisältö voi informoida pyöräilyverkoston suunnittelua.

Painotetut laatutulokset laskettiin paikallisesti aggregaattina $ADQR_i$ kilometriruuduittain (yhtälö 10; taulukko 16). Koko aineiston kattava yhteisprosentti laskettiin yhtälöllä, jossa KK on kartoituksen käytettävyys, RNK reitityksen ja navigoinnin käytettävyys ja TLK käytettävyys tietolähteeksi (yhtälö 11). N on kilometriruutujen lukumäärä. Tällä tavoin aggregoidut laatutulokset vastasivat ISO-standardin kuvaamaa ADQR-lukua.

$$ADQR_i = S_i \cdot ws + P_i \cdot wp + SA_i \cdot wsa + SB_i \cdot wsb + T_i \cdot wt \quad (10)$$

$$KK, RNK, TLK = \frac{\sum_{i=1}^n ADQR_i}{n} \cdot 100 \quad (11)$$

Taulukko 16. Käytettävyysmittauksen muuttujat ja painokertoimet käyttötarkoituksille.

Laatutulos, DQR	Muuttuja	Painokerroin	Kartoitus	Reititys ja navigointi	Uusi tieto
Puuttuminen	<i>P</i>	wp	0,4	0,4	0,5
Sisältyminen	<i>S</i>	ws	0,2	0,1	0,3
Sijaintitarkkuus A	<i>SA</i>	wsa	0,2	0,05	0,05
Sijaintitarkkuus B	<i>SB</i>	wsb	0,1	0,05	0,05
Topologinen eheys	<i>T</i>	wt	0,1	0,4	0,1

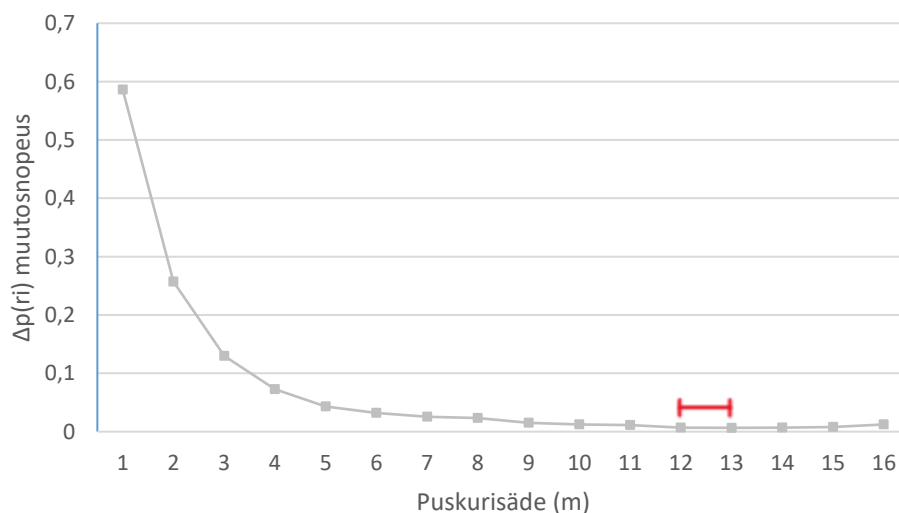
Paikallinen arvio käytettävyydestä esitettiin kilometriruudukossa koropleettikarttana. Aineistotason tulos esitettiin paikallisen arvion kokonaisprosenttina. Arvioitaessa soveltuvuutta tietolähteeksi, Sisältymisen muuttujan arvot käännettiin, jotta suuremmat arvot kuvasivat kartalla parempaa soveltuvuutta.

4 Tulokset

4.1 Aineiston esikäsittely

4.1.1 Optimaalisen puskurisäteen pituus

Segmenttien kappalemäärän ja pituuden suhdelukujen summan $P(r_i)$ muutosnopeus $\Delta P(r_i)$ alkoi tasaantua seitsemässä metrissä, pysähtyi 12 metrin kohdalla ryhmissä r_{12} ja r_{13} ja alkoi sen jälkeen kasvaa (kuva 17). Myös $P(r_i)$ muutos tasaantui ryhmien r_{13} ja r_{14} kohdalla. Ryhmän r_{12} arvojen keskiarvo on 12,5 m, joka osoittaa optimipuskurin säteen. Puskurin läpimitta on 25 metriä.



Kuva 17. Segmenttien kappalemäärän ja pituuden suhdelukujen summan muutosnopeus pysähtyi 12—13 metrin välillä (punainen viiva), jossa puskurisäteittäisten summien keskiarvo oli 12,5 metriä. Keskiarvo osoitti optimaalisen puskurisäteen täydellisyys- ja laatulementin sisällyttämisen alaelementin erottamiselle testiaineistosta.

4.1.2 Spatiaalinen painottuneisuus

Spatiaalisen painottuneisuuden poistaminen on sijaintitarkkuusmittauksen ennakkoehto. Aineistotason painottuneisuuden mittareita olivat mediaanikeskustojen etäisyys sekä segmenttien oskillaatioarvo. Paikallisen painottuneisuuden mittareita olivat kilometriruutujen mediaanikeskustojen etäisyys ja suuntaus. Paikallisten mediaanikeskustojen etäisyys- ja atsimuuttimittausten spatiaalinen autokorrelaatio testattiin globaalilla Moranin I -testillä.

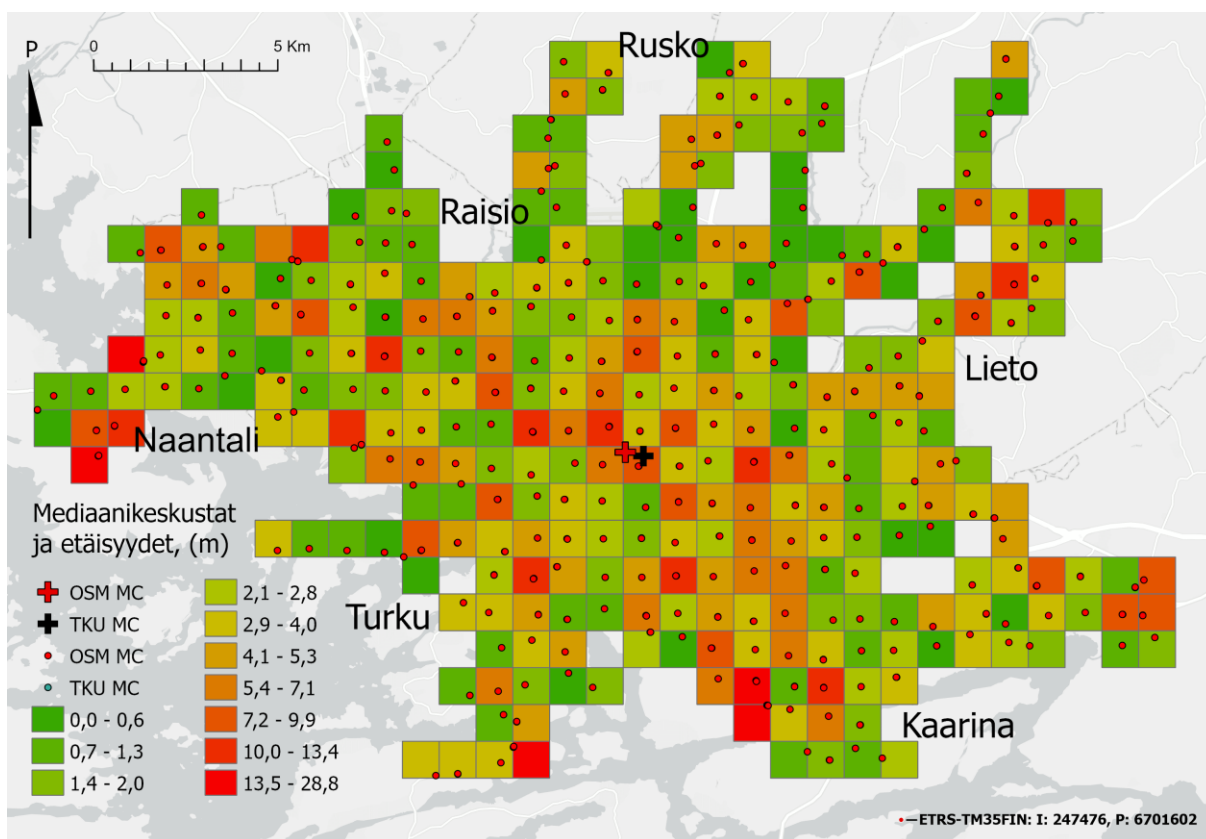
Aineistotason mediaanikeskustat sijoittuvat ETRS-TM35FIN-koordinaatistossa vertailuaineistossa koordinaatteihin (I: 240364, P: 6711447) ja testiaineistossa (I: 239874, P:

6711551). Mediaanikeskustojen etäisyys on 500,8 m. Testiaineiston mediaanikeskusta sijaitsee 489,9 m länteen ja 104,1 m pohjoiseen vertailuaineiston mediaanikeskustasta (kuva 18). Visuaalisessa tarkastelussa ei havaita koko aineiston kattavaa systemaattista painottuneisuutta.

Aineistotason mediaanikeskustojen etäisyysarvo perustelee paikallisen painottuneisuuden varmistamista (kuva 18). Mediaanikeskustojen etäisyys- ja atsimuuttiarvojen kasautuminen kilometriruudukossa testattiin globaalilla Moranin I -testillä, jonka hypoteesit ovat seuraavat:

H_0 : Muuttuja ei autokorreloi. Sen spatiaalinen jakauma on satunnainen.

H_v : Muuttuja autokorreloi. Arvojen kasautuminen on satunnaisuutta voimakkaampaa.



Kuva 18. Testi- ja vertailuaineistojen mediaanikeskustat aineistotasolla (ristit) ja paikallisesti (ympyrät). Paikallinen mediaanikeskustojen etäisyys jatkuvana muuttujana värein. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Globaali Moranin I -testi osoittaa etäisyysarvojen äärimmäisen heikkoa kasautumista.

(Moranin I: 0,009, Z-arvo: 3,02 ja P-arvo: 0,0025). P-arvo on tilastollisesti merkitsevä, joten nollahypoteesi H_0 hylättiin (How spatial... 2023). Moranin I on lähes nolla.

Atsimuuttien globaali Moranin I (0,004), Z-arvo (1,74) ja P-arvo (0,08) osoittavat äärimmäisen heikkoa kasautumista. Tulos ei ole tilastollisesti merkitsevä, joten nollahypoteesia H_0 ei hylätty (How spatial... 2023).

Painottuneisuuden tulokset varmennettiin sijaintitarkkuuden mittauksessa oskillaatioarvolla (Tveite & Langaas 1999). Oskillaatioarvon kuvaaja on aidosti laskeva (kuva 26).

4.2 OSM-aineiston laatu ja sen alueellinen vaihtelu

Aineistosta arvioitiin ISO 19157 -standardin laatulementit Täydellisyys (Sisältyminen ja Puuttuminen), Looginen eheys (Topologinen eheys) ja Sijaintitarkkuus (Ulkoinen sijaintitarkkuus).

Seuraavassa jokaiselle laatulementille ilmoitetaan aineistotason tulokset sekä paikalliset, alueellista vaihtelua kuvaavat tulokset. Samantasoisten paikallisten mittausten kasautuminen esitetään Anselinin Moranin I -testin tulokartoilla.

4.2.1 Täydellisyys

Aineistotason täydellisyys arvioitiin Täydellisyyden laatulementin alaelementeillä Sisältyminen ja Puuttuminen. Sisältyminen ilmaisee testiaineistoon kuuluvia, mutta vertailuaineistosta puuttuvia kohteita. Puuttuminen ilmaisee kattavuutta eli vertailuaineistossa olevia, mutta testiaineistosta puuttuvia kohteita. Sisältymisen alaelementillä arvioitiin myös aineiston käyttöpotentiaalia tietolähteenä.

Visuaalisessa tarkastelussa havaitaan aineistojen huomattavaa yhdenmukaisuutta, mutta myös runsaasti selkeitä eroja. Poikkeavat kohteet keskittyvät testiaineistossa verkoston reuna-alueille ja sijoittuvat verrattain kiinteästi vertailuaineiston verkoston lähistölle.

Käsittlemättömän testiaineiston kokonaispituus on 927,9 km ja vertailuaineiston kokonaispituus 980,8 km.

Puuttumisprosentti, eli testiaineiston kattavuus suhteessa vertailuaineistoon on 95 %.

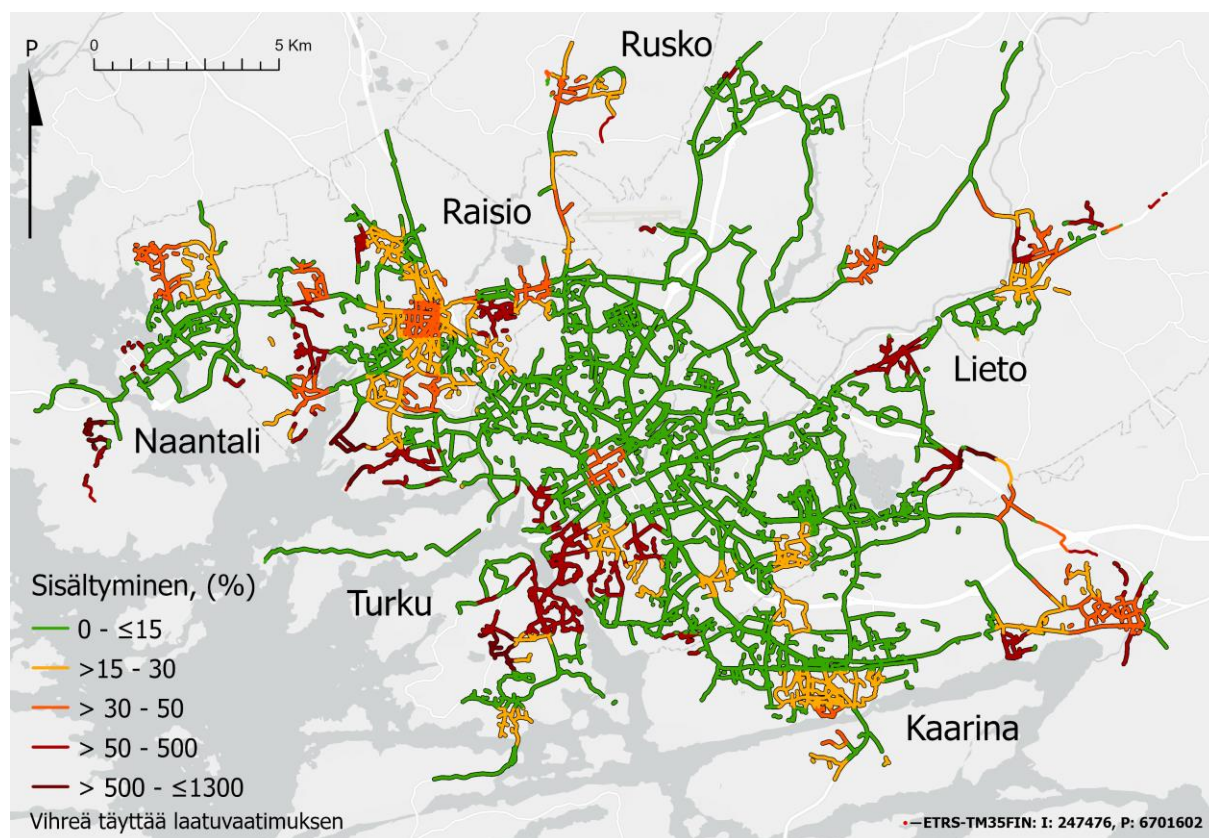
Testiaineistosta puuttuvan verkoston pituus on 53,7 km.

Sisältymisprosentti, eli vertailuaineistosta puuttuvan, mutta testiaineistoon ylimääräisenä sisältyvän verkoston osuus on 12,6 %. Ylimääräisen verkoston pituus on 123,7 km.

Poistettaessa Sisältyminen ylimääräinen osuus testiaineistosta, saadaan kattavuuden tarkka tulos. Testiaineiston verkoston pituus on tällöin 803,4 km ja Puuttumisprosentti 82 %.

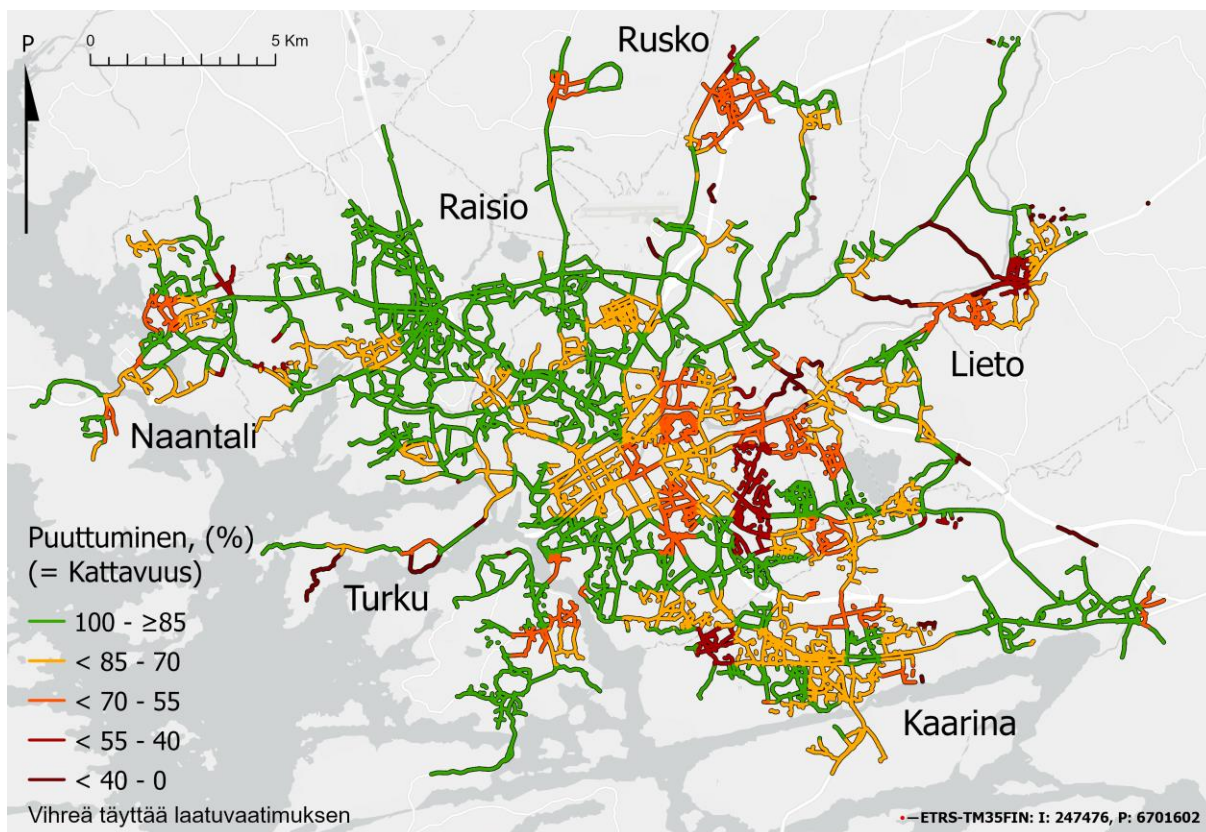
Sisältyminen mitattiin myös osana sijaintitarkkuuden arviota Virhekoodausten tunnusluvulla. Sen tulos on 11,7 % optimipuskurisäteen leveydellä.

Täydellisyyden paikallinen mittaus esitetään Sisältyminen ja Puuttumisen koropleettikartoilla. Sisältymisprosentti on laatuvaatimuksen edellyttämällä tasolla (< 15 %) laajalla alueella (kuva 19). Asutuskeskusten ydinalueiden läheisyydessä sisältyminen moninkertaistuu. Alueille sijoittuu myös paljon vaatimuksen täyttävää verkostoa. Sisältyminen on hyvälaatuinen Turussa laajalla alueella Majakkarakantaa ja Hirvensaloo lukuun ottamatta. Kilometriresoluutiolla sisältyminen alueellisen vaihtelun keskiarvo on 32,8 % ja keskihajonta 115,1 % (taulukko 19).



Kuva 19. Testiaineiston täydellisyyden alaelementti sisältyminen. Vihreä väri osoittaa laatuvaatimuksen täyttävän aineiston. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Puuttumisprosentti alittaa laatuvaatimuksen (> 85 %) laajalla alueella (kuva 20). Raision alueen hyvä kattavuus erottuu muista kuntakeskuksista. Valtaosassa tutkimusaluetta kattavuus pysyttelee välillä 100–70 %. Turun keskustassa kattavuus jää hieman laatuvaatimuksen alle (85–70 %). Keskustasta itään Lietoon saakka levittyy heikon kattavuuden alue (< 85–0 %) ja

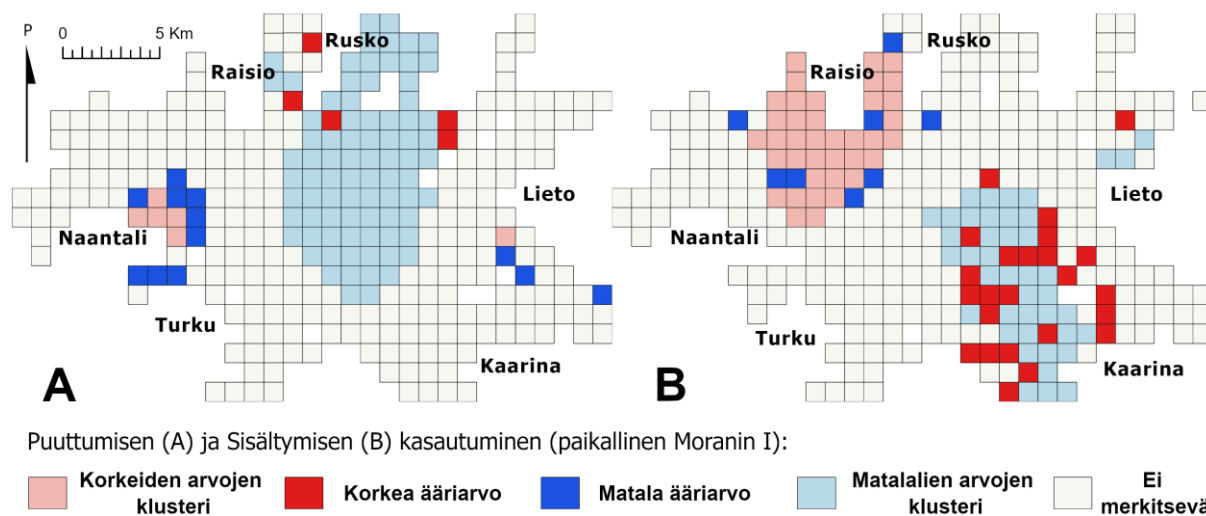


Kuva 20 Testiaineiston täydellisyysalaelementti puuttuminen. Vihreä väri osoittaa laatuvaatimuksen täyttävän aineiston. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Jäkärlässä kattavuus on vain 50–70 %. Heikko kattavuus on korostunut Liedon alueella.

Kaarinassa tilanne on tasalaatuisempi ja laajalti laatuvaatimuksen rajalla.

Kilometriresoluutiolla puuttumisen alueellisen vaihtelun keskiarvo on 82 % ja keskihajonta 14,7 % (taulukko 19).



Kuva 21. Testiaineiston täydellisyysalaelementtien sisällymisen ja puuttumisen spatiaalinen autokorrelaatio (paikallinen Moranin I.) Neliökilometrin resoluutio.

Sisältymisen Anselinin paikallinen autokorrelaatio -testi ilmaisee korkean laadun matalien arvojen kasaamina (kuva 21a). Laatuvaatimuksenmukaisuus ilmenee laajana matalien arvojen kasaamana autokorrelaatiokartan keskiosassa. Raision eteläpuolella esiintyy heikkolaatuinen runsaan sisältymisen kasauma.

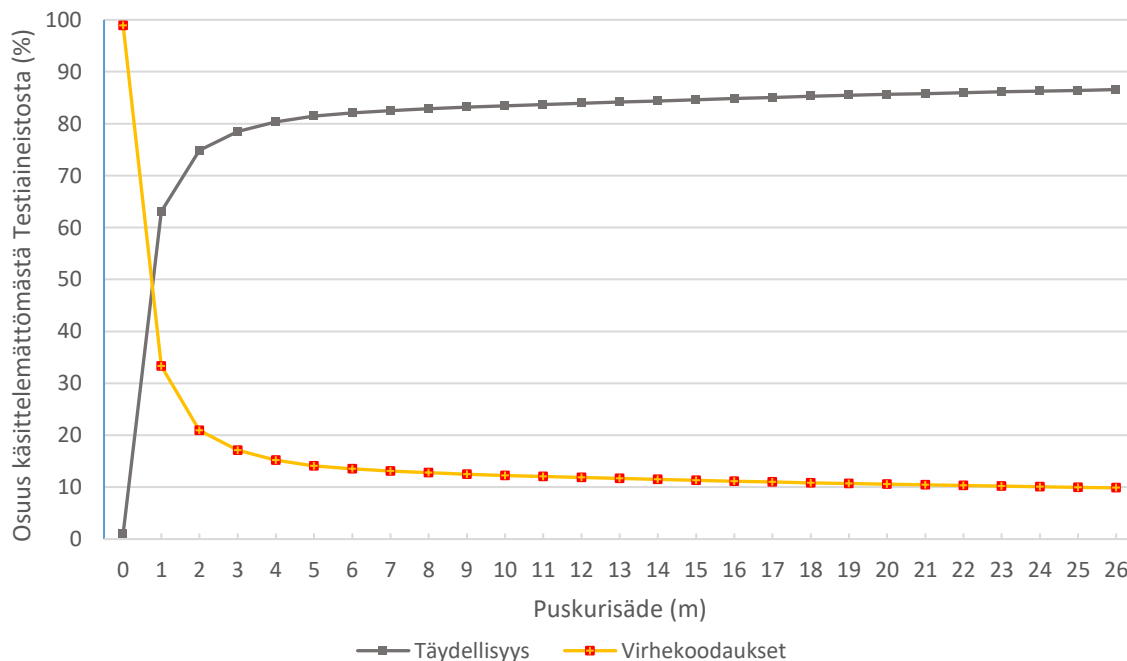
Puuttumisen Anselinin paikallinen autokorrelaatiokartta (kuva 21b) ilmaisee korkean laadun korkeiden arvojen kasaamina. Kattavuus on Raision alueella hyvä. Yllä kuvattu matalan kattavuuden vyöhyke erottuu kartan kaakkoisosassa.

4.2.2 Sijaintitarkkuus

Sijaintitarkkuuden mitaamiseen käytettiin Tveiten ja Langaasin (1999) Päällekkäisten puskureiden tunnusluvuista seuraavia: Täydellisyys (kattavuus), Virhekoodaukset (ylimääräisenä sisältyvä aineisto), keskimääräinen siirtymä (sijaintitarkkuus), Oskillaatio (spatiaalinen painottuneisuus) ja Siirtymätietokaavion tunnusluku (geometrinen tarkkuus).

Sijaintitarkkuusarvion ennakkoehto on aineistojen toisiaan vastaava täydellisyys.

Aineistotason tarkastelussa selvitettiin myös paikalliseen sijaintitarkkuusarvioon soveltuva puskurisäde. Sijaintitarkkuusmittauksen perusteella arvioitiin testiaineiston vastaavuus Maastotietokantaan kahdessa laatuluokassa A (3 m) ja B (12,5 m).



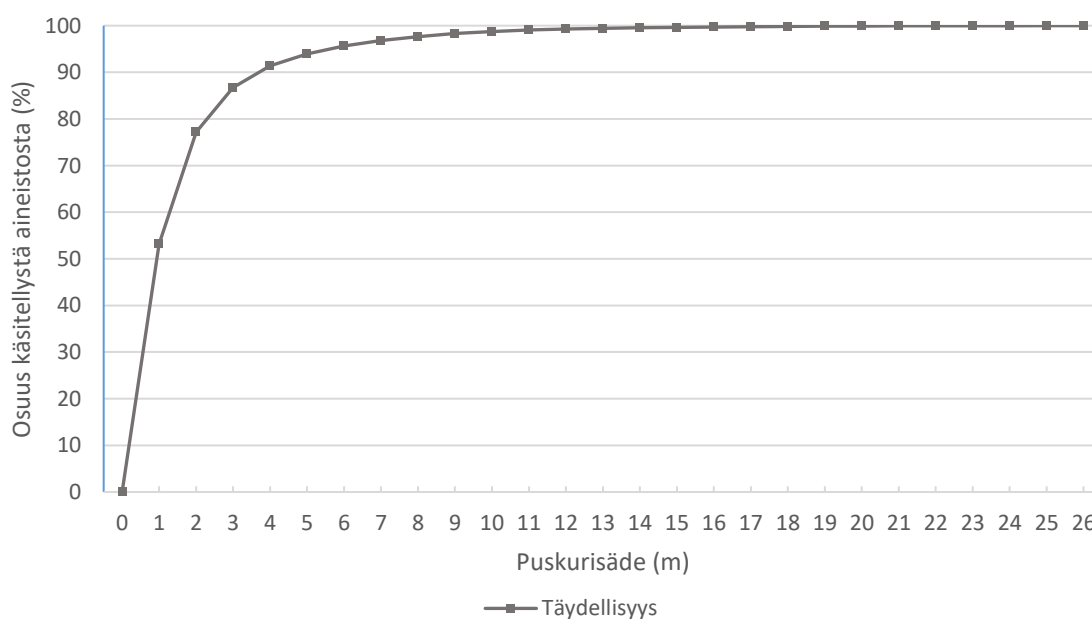
Kuva 22. Täydellisyys: Testiaineiston kattavuus suhteessa vertailuaineistoon. Virhekoodaukset: Vertailuaineistosta puuttuva ylimääräinen testiaineisto, vastaa sisältymistä. Muutos puskurisäteen funktiona. Mittaus suoritettu täydellisysehtoa täyttämättömillä aineistoilla.

Täydellisyyden kuvaaja taittuu puskurisäteen lähestyessä kolmea metriä, jonka jälkeen sen kasvunopeus tasaantuu (kuva 22). Tasaannuttuaan testiaineiston osuus lisääntyy puskurisäteen kasvaessa enää vähittäin. Valtaosa testiaineistosta (78,5 %) sijoittuu kolmen metrin säteelle vertailuaineistosta. Hidas kasvunopeus jatkuu tutkittuun 26 metriin saakka, jossa puskuriin sisältyy 86,6 % testiaineistosta. Puskurin läpimitta on tällöin 52 metriä. Muutostrendi on jatkuva.

Virhekoodausten määrän nähdään vähenevän Päällekkäisten puskurien menetelmän ennustamalla tavalla (Kuva 22). Kuvaaja tasaantuu kolmen metrin läheisyydessä, jonka jälkeen hidas muutosnopeus pitää sen välillä 6–26 metriä tasolla 13,5 % – 9,9 %. Muutostrendi on jatkuva. Sijaintitarkkuuden mittaukseen valitulla 13 metrin puskurisäteellä, virhekoodaukset ovat 11,7 %.

Täsmällinen tulos testiaineiston sijaintitarkkuudesta puskurisäteiden funktiona saadaan, kun mittauksen täydellisyysehto täyttyy, eli aineistojen kattavuus on geometrisesti yhtäläinen.

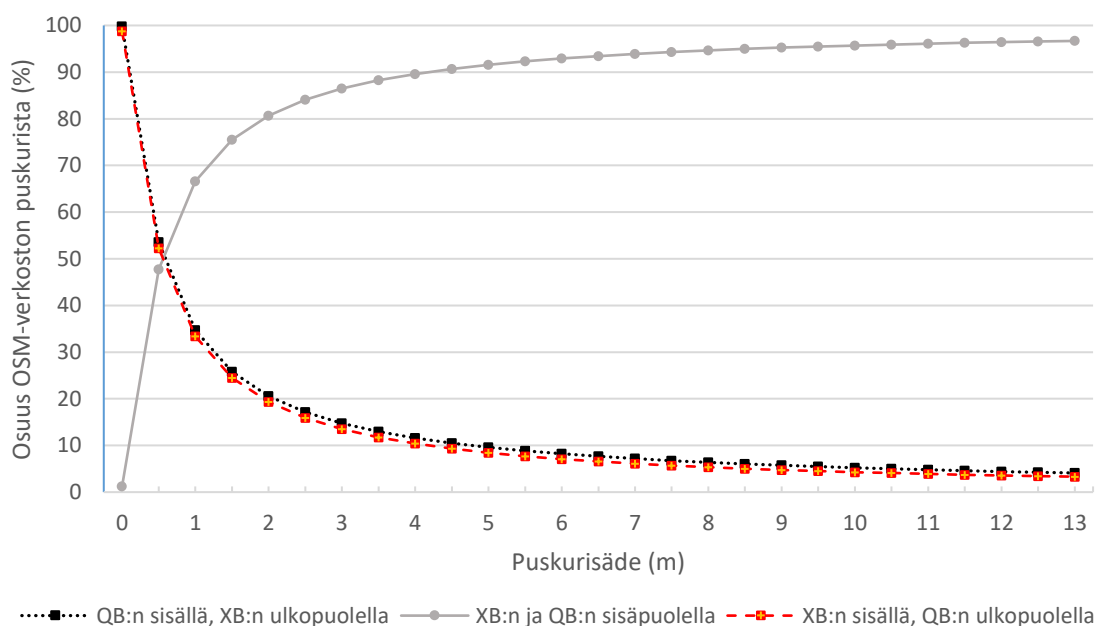
Käsiteltyjen aineistojen vastaava kattavuus varmennettiin Täydellisyyden tunnusluvun mittauksella. Sen kuvaaja osoittaa testiaineiston puskurisäteittäiset osuudet metrin mittausvälein (kuva 23). Kuvaaja nousee nopeasti välillä 0–2 m ja lähestyy 100 % kuuden metrin jälkeen. 86,7 % aineistosta on enintään kolmen metrin etäisyydellä vertailuaineistosta ja 99,3 % enintään 12,5 m etäisyydellä.



Kuva 23. Testiaineiston keskimääräinen sijaintitarkkuus puskurisäteittäisinä osuuksina. Mittaus suoritettu täydellisyysehdon täyttävillä aineistoilla.

Tulokset osoittavat sijaintitarkkuuden paikalliseen arvioon tarvittavan puskurisäteen, joka mahdollistaa vastaavuuden vertaamisen Maastotietokannan kumpaankin laatuluokkaan.

Sijaintitarkkuuden mittauksessa laskettujen polygonien alojen suhdetta kuvaava Siirtymätietokaavio ilmoittaa aineistotasolla testiaineiston tarkkuuden geometrisellä painotuksella suhteessa vertailuaineistoon (kuva 24). Puskuroitujen verkostojen päällekkäisyyden lisääntyessä, eli tarkkuuden parantuessa, alue *XB:n ja QB:n sisäpuolella* kasvaa. Vastaavasti alueet *QB:n sisällä, XB:n ulkopuolella* sekä *XB:n sisällä, QB:n ulkopuolella* pienenevät (kuva 23; kuva 15).



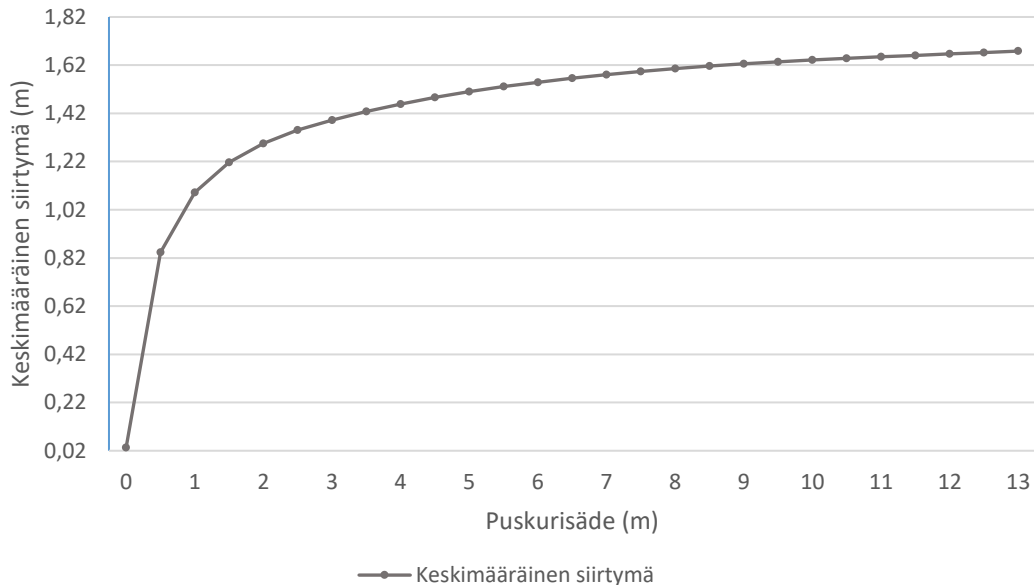
Kuva 24. Siirtymätietokaavion tunnusluku kuvaa verkostojen välistä etäisyyttä huomioiden myös niiden geometrisen eron. Kokonaisalat puskurisäteittäin kolmelle päällekkäisten puskurien leikkausoperaatioissa syntyneille polygoniryhmälle. Mitä suurempi osuus OSM-verkoston puskurista sisältyy ryhmään 'QB:n sisällä, XB:n ulkopuolella', sitä yhdenmukaisempia verkostot ovat.

Testiaineiston tarkkuuden ero vertailuaineistoon on pieni, sillä kuvaajat risteävät jo 0,5 m säteen lähistöllä. Puskureiden yhteinen alue kasvaa ja erillisymueet vähenevät nopeasti ja symmetrisesti. Kuvaajien muutosnopeudet eivät kasva tai vähene odottamattomasti.

Polygonien siirtymätiedon perusteella laskettiin Keskimääräinen siirtymä, jossa eroavuus painottuu viivojen välisellä etäisyydellä. Aineistotason keskimääräinen sijaintitarkkuus suhteessa vertailuaineistoon mitattiin tarkasteluvälillä 0–13 m puskurisäteittäin 0,5 m välein.

Sijaintitarkkuuden ero lisääntyy välillä 0,02–1,67 m puskurisäteen kasvaessa maksimiin (kuva 25), mutta pysyy alhaisella tasolla. Suurin muutos havaitaan puskurisäteiden välillä 0–2 m, jossa epätarkkuus kasvaa nolasta 1,29 metriin nopeasti. Tämän jälkeen kuvaaja tasaantuu

lähestyttäessä koko aineiston keskimääräistä siirtymää. Kolmen metrin puskuriin sisältyvän testiaineiston keskimääräinen etäisyys vertailuaineistosta on 1,39 metriä. 12,5 metrin puskuriin sisältyvän testiaineiston keskimääräinen etäisyys vertailuaineistosta on 1,67 metriä.



Kuva 25. Aineistotason sijaintitarkkuus. Testi- ja vertailuaineistojen välinen keskimääräinen etäisyys toisistaan puskurisäteittäin.

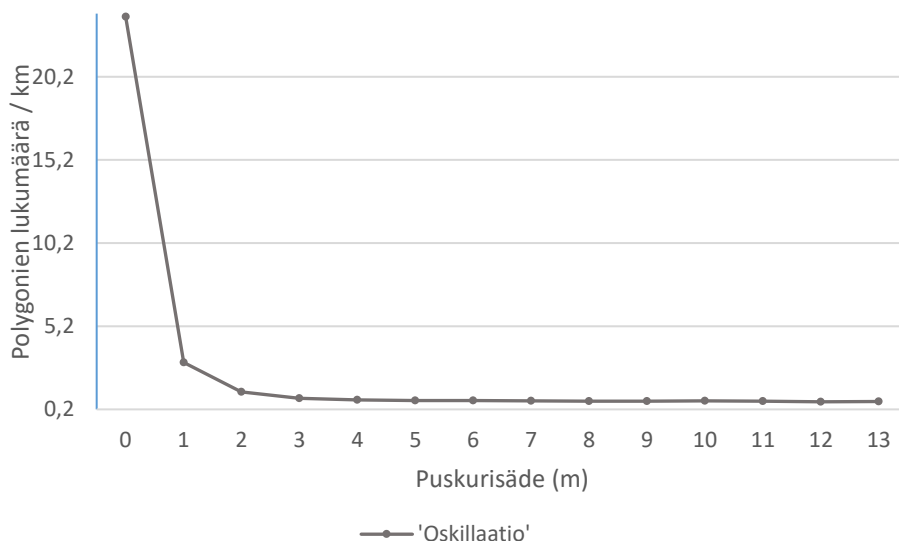
Sijaintitarkkuuden mittaustulokset ja vertailu Maastotietokannan laatuluokkiin on koottu alle taulukkoon (taulukko 17). Siirtymätietokaavion kuvaaman (kuva 23), *XB:n ja QB:n sisäpuolella* olevan alueen tarkastelu yhdessä keskimääräisen siirtymän tunnusluvun kanssa (kuva 25) kertoo sekä testiaineiston sijaintitarkkuuden että tarkkuuden todennäköisyyden puskurisäteittäin.

Taulukko 17. Sijaintitarkkuuden osatekijät Maastotietokannan laatuluokkien mukaisesti. Laskenta suoritettu täydellisyysvaatimuksen täyttävillä aineistoilla. Prosenttiosuus ilmaisee, kuinka suuri osa testatusta aineistosta sisältyy puskuriin 3 metrin ja 12,5 metrin säteillä. Keskimääräinen siirtymä ilmaisee osuuksien keskimääräisen tarkkuuden.

Muuttuja	A-luokka, 3 m	B-luokka, 12,5 m
Ala XB:n ja QB:n sisäpuolella	86,50 %	96,50 %
Keskimääräinen siirtymä	1,39 m	1,67 m

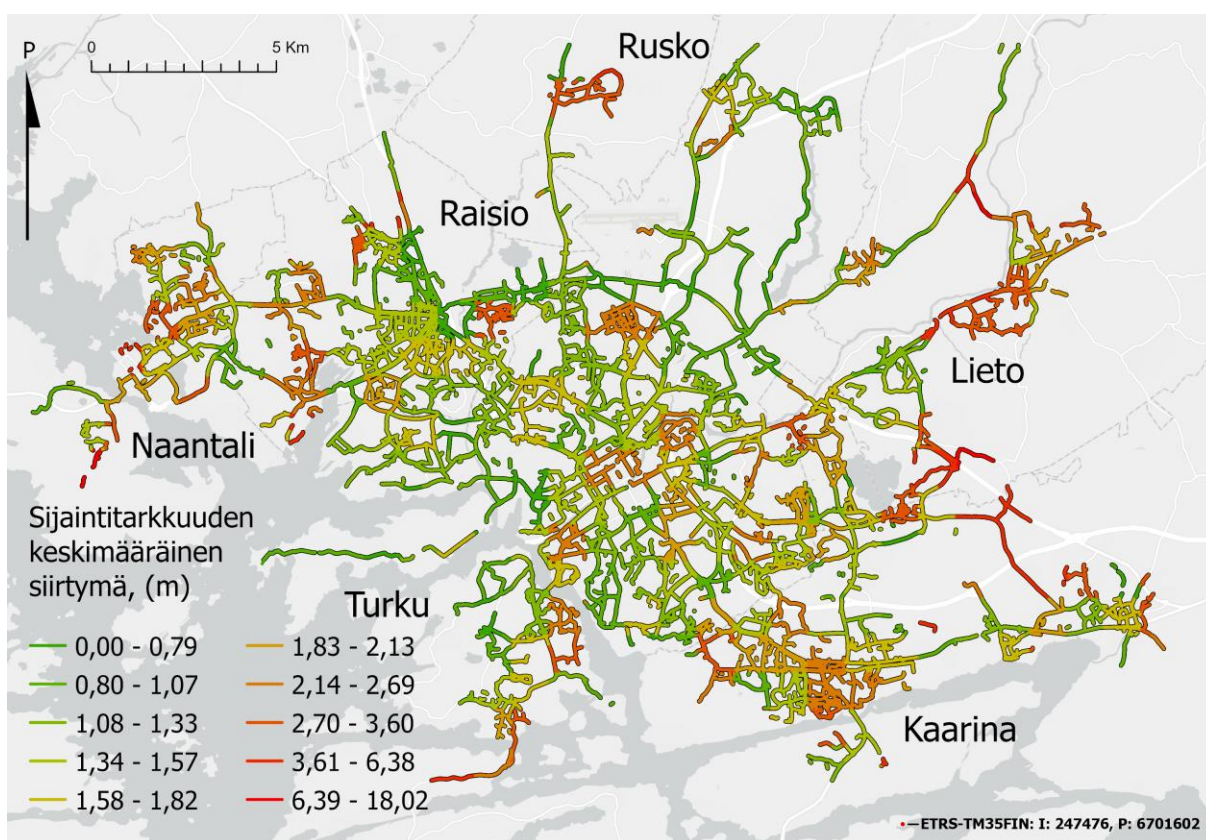
Verkostojen välinen oskillaatio eli toisiaan vastaavien viivojen leikkaaminen keskenään, seuraa aidosti vähenevää trendiä puskurisäteiden vaihteluvälillä 0–13 m (kuva 26).

Puskuroitujen verkostojen leikkausoperaatiossa syntyneiden polygonien lukumäärä kilometrillä on vähäinen (23,8–0,18 Kpl). Voimakkaimmillaan oskillaatio on hyvin pienillä säteillä ja poistuu miltei kokonaan kolmen metrin kohdalla. Kuvaaja laskee koko tarkasteluvälillä, myös välillä 3–13 m.



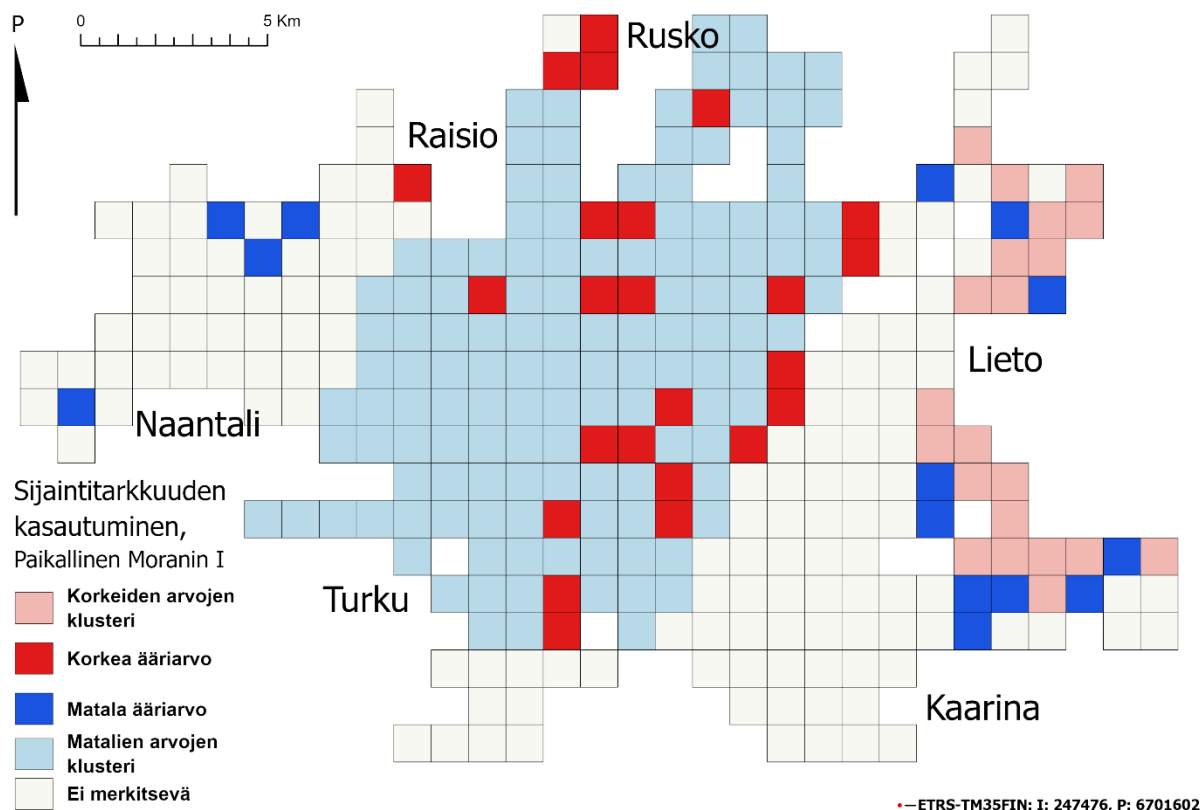
Kuva 26. Oskillaation tunnusluku kuvaa vertailtavien aineistojen tieseementtien leikkauskohtien tiheyttä puskurisäteittän kilometrillä. Suurin tiheys esiintyy pienillä puskurisäteillä. Suuremmilla säteillä leikkauskohtia ei enää havaita. Oskillaatio viittaa spatiaaliseen painottuneisuuteen testiaineistossa, jos kuvaaja ei ole aidosti laskeva.

Sijaintitarkkuus on kokonaisuutena hyvätasoinen. Keskimääräisen siirtymän alueellinen vaihtelu esitetään alla (kuva 27).



Kuva 27. Sijaintitarkkuus jatkuvana muuttujana. Keskimääräinen siirtymä. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

Valtaosassa tutkimusaluetta sijaintitarkkuuden ero vertailuaineistoon on välillä 0–3,6 metriä. Turun, Raision ja pohjoisen Kaarinan alueet erottuvat pienellä vaihteluvälillä 0–1,5 metriä. Epätarkinta aineisto on Liedon ja Kaarinan itäisimmissä osissa, jossa ero vertailuaineistoon vaihtelee välillä 3,6–18 metriä. Keskustojen ydinalueilla epätarkkuus hieman lisääntyy. Kilometriresoluutiolla alueellisen vaihtelun keskiarvo on 1,63 ja keskihajonta 0,84 (taulukko 19).

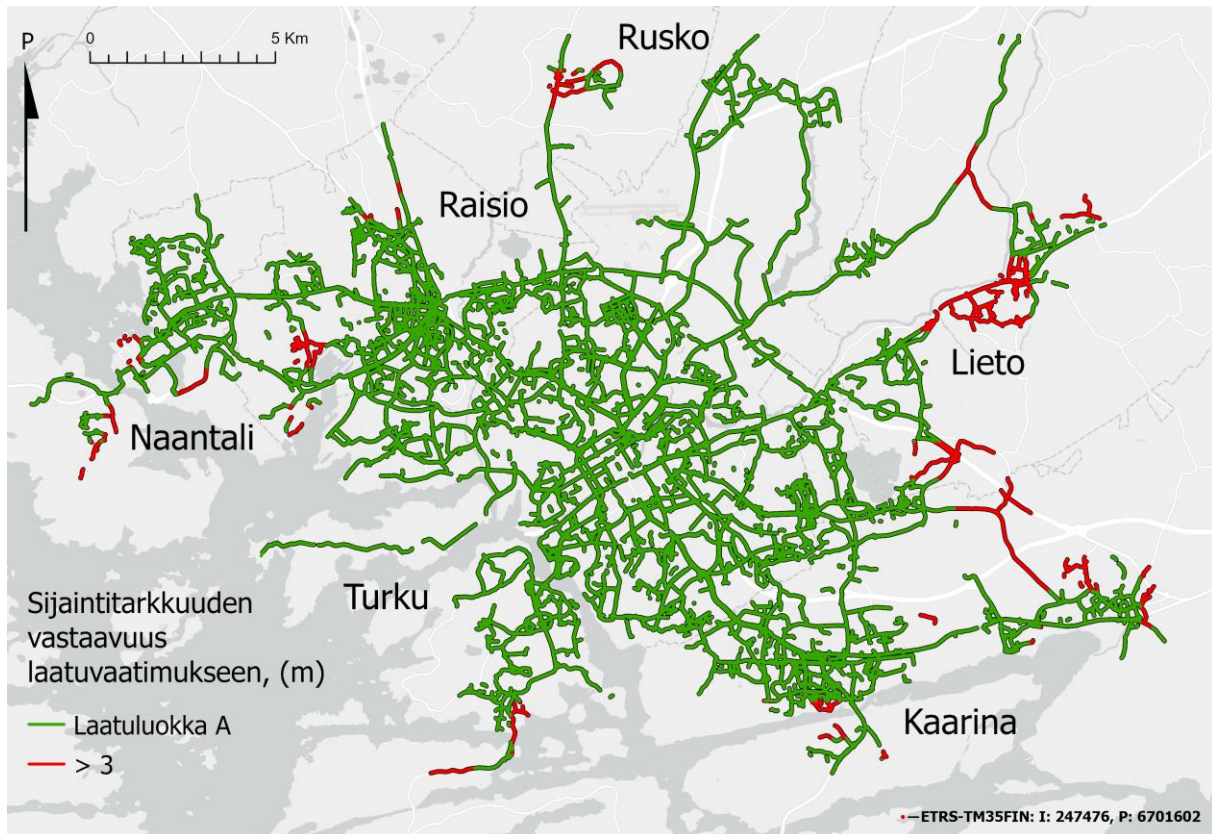


Kuva 28. Sijaintitarkkuusmittausten spatiaalinen autokorrelaatio. Paikallinen Moranin I. Jatkuva muuttuja. Neliökilometrin resoluutio.

Keskimääräisen eroavuuden arvojen spatiaalinen autokorrelaatio esitetään alla (kuva 28).

Matalat arvot ovat ryvästyneet laajalle alueelle tutkimusalueen keskellä. Pääosa tästä vyöhykkeestä sijoittuu Turun alueelle. Epätarkkuuden kasaumat sijoittuvat Turun ympäryskuntiin. Suurimmat kasaumat ovat Liedon ja Kaarinan itäisten osien alueella.

Sijaintitarkkuus on Maastotietokannan laatuluokan A tasolla (< 3 m) kaikkialla paitsi verkoston reuna-alueilla (kuva 29). Tarkkuudeltaan laatuvaatimuksen alittavaa aineistoa on kuitenkin vain vähäisessä määrin. Lieto ja Kaarinan itäismuuet erottuvat epätarkemmalla aineistolla. Laatuluokan B (< 12,5 m) taso täyttyy kaikkialla tutkimusalueella.



Kuva 29. Testiaineiston sijaintitarkkuuden vertailu Maastotietokannan A-laatuluokkaan. Laatuvaatimus <math>< 3\text{ metriä}</math>. Vihreä väri osoittaa laatuvaatimuksen täyttävän aineiston. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

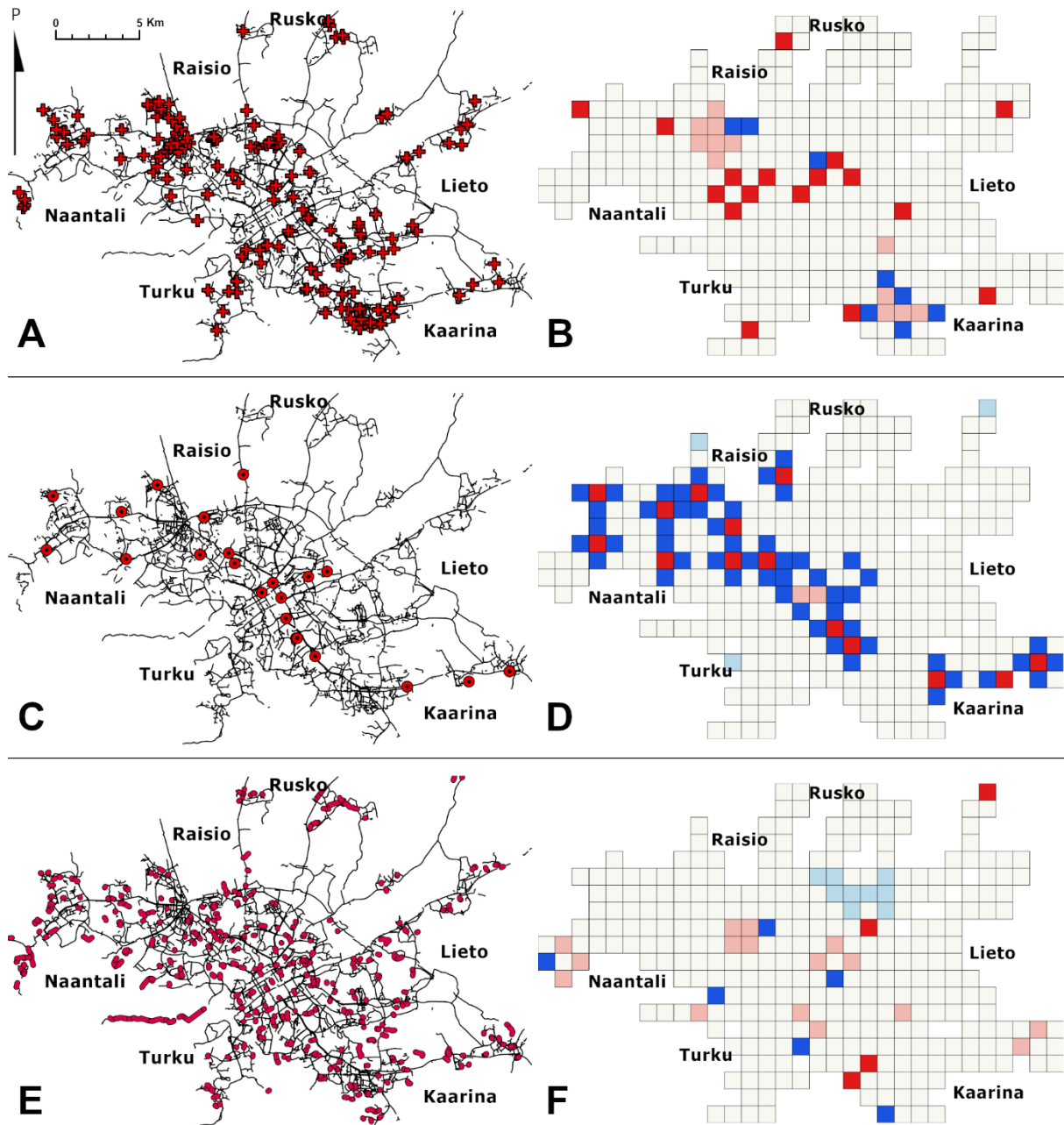
4.2.3 Topologinen eheys

Topologista eheyttä tutkittiin testiaineiston ja topologisen skeeman yhdenmukaisuutena sekä reittien yhdistävyytenä. OSM-aineiston kolme yleisintä topologista virhettä mitattiin (kuva 10).

Risteykset ilman liittymää ovat risteävien viivojen solmuttomia leikkauskohtia.

Aineistotasolla niitä on 172 kpl, eli 7,7 % kaikista testiaineiston risteyksistä. Ne levittyvät tutkimusalueelle verrattain tasaisesti, lukuun ottamatta kahta Raisioon ja Kaarinaan sijoittuvaa kasaumaa (kuva 30a ja kuva 30b).

Melkein liittymät ovat liittymäkohdastaan vajaita tai ylipitkiä solmuttomia viivoja. Niiden lukumäärä testiaineistossa on 22 kpl, eli 2 % kaikista liittymistä. Melkein liittymien etäisyys oikeista liittymäkohdistaan on alle 4,5 metriä. Virheet sijoittuvat Naantalista Kaarinaan ulottuvalle akselille, jolla urbaani rakenne on tiheä (kuva 30c). Spatiaalisen autokorrelaation testissä esiintyy yksittäisiä ääriarvoja. Satunnaisuutta ylittävää kasaantumista ei havaita lukuun ottamatta pientä aluetta Turun keskustassa. (kuva 30d)



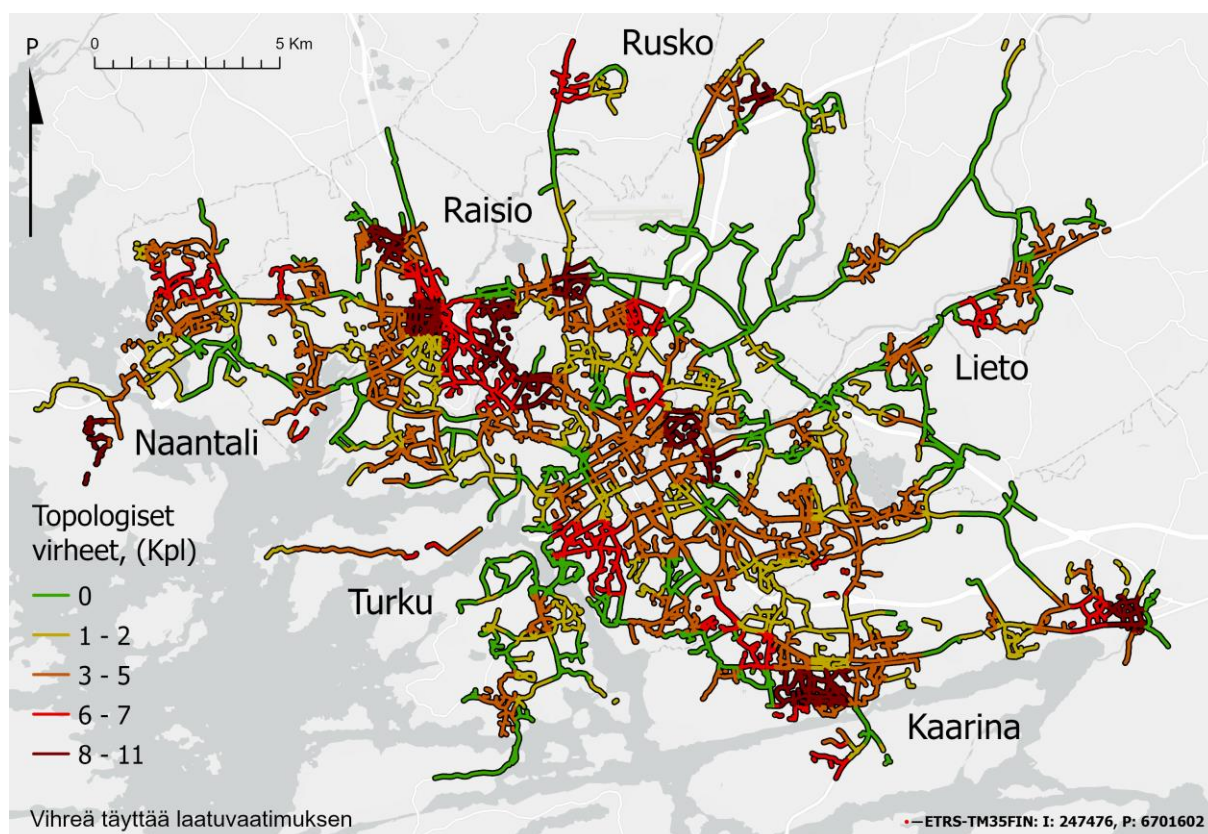
Kuva 30. Topologisten virheiden spatiaalinen jakauma ja spatiaalinen autokorrelaatio (paikallinen Moranin I). Risteykset ilman liittymää: kuvat A–B, Melkein liittymät: C–D, Kelluvat saaret: E–F. Neliökilometrin resoluutio. Miten tulkitaan klustereita.

Erilliset saaret ovat muusta verkostosta irrallisia osuuksia. Niiden lukumäärä testiaineistossa on 394 kpl, yhteispituus 52,8 km, segmenttien pituuden keskiarvo 134,4 m ja keskihajonta 158,6 m. Segmenttien pituuden minimi on 0,5 m ja maksimi 1,2 km. Yhteispituuden osuus testiaineistosta on 0,6 %. Erillisten saarten spatiaalinen jakauma on tasainen (kuva 30e).

Spatiaalisen autokorrelaation testissä havaitaan kuitenkin Naantali–Kaarina-akselille sijoittuvia kasaumia ja yksi matalien arvojen vyöhyke Turun koillisosassa (kuva 30f).

Topologisen eheyden kokonaisvirheen alueellisen vaihtelun keskiarvo on 2,24 ja keskihajonta 2,52 (taulukko 19). Heikon topologisen laadun kasautumista havaitaan tiheän kaupunkirakenteen alueella, vaikka huomattavia kasaumia ei ole.

Topologisten virheiden kokonaismäärä täyttää laatuvaatimuksen (0 virhettä) harvoilla alueilla (kuva 31). Yhdistävyyteen liittyvät ongelmat levittyvät koko tiheimmän urbaanin rakenteen alueelle. Yhtenäistä virheetöntä aluetta esiintyy mainittavasti vain Turun ja Liedon reuna-alueilla. Laajemmat korkeiden virheiden vyöhykkeet sijoittuvat Raision, Turun ja Kaarinan keskustoihin.



Kuva 31. Topologiset virheet jatkuvana muuttujana. Vertailu Maastotietokannan laatuvaatimukseen. Vihreä väri osoittaa laatuvaatimuksen täyttävän testiaineiston. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

4.2.4 Vertailu laatuvaatimukseen

Vertailuaineiston sijaintitarkkuuden kolmen metrin epävarmuusalueelle sisältyvä testiaineisto ylittää Maastotietokannan laatuluokan A tasolle ja 12,5 metrin epävarmuusalueella se täyttää laatuluokan B tarkkuusvaatimuksen.

Aineistotasolla laatulementtien tuloksista yltyvät laatuvaatimukseen sisältyminen ja sijaintitarkkuus (B-luokka) (taulukko 18). Laatuvaatimuksen mukaista kattavuutta, sijaintitarkkuuden A-luokkaa ja täydellistä topologista virheettömyyttä ei saavuteta.

Taulukko 18. Laatuarvion tulokset Maastotietokannan laatuvaatimusta vastaan. Punainen = ei täytä vaatimusta, vihreä = täyttää vaatimuksen.

Laatuelementti	Mittauskohde	Laatuvaatimus	Laatutulos
Täydellisyys	Puuttuminen	≥ 85 %	82 %
Täydellisyys	Sisältyminen	≤ 15 %	12,6 %
Sijaintitarkkuus	Laatuluokka A	95 % ≤ 3 m	86,5 %, 1,39 m
Sijaintitarkkuus	Laatuluokka B	95 % ≤ 12,5 m	96,5 %, 1,67 m
Topologinen eheys	Kaikki virheet	0 kpl	588 kpl

Paikallisessa tarkastelussa laatutaso vaihtelee alueellisesti. Sisältyminen pysyy edelleen hyvällä tasolla laajalla alueella (kuva 19), mutta puuttuminen (kuva 20) ja topologinen eheys (kuva 31) ilmenee alueellisesti rikkonaisempana. Sijaintitarkkuus täyttää laatuvaatimuksen tarkkuusluokassa B koko aineiston osalta. Luokan A tasolle päästään laajalla alueella, mutta ei kaikkialla (kuva 29).

4.3 OSM-aineiston käytettävyys ja sen alueellinen vaihtelu

Painotetuista laatutuloksista kootut aggregaattiarvot kuvaavat käytettävyyttä kolmessa käyttötarkoituksessa (kuvat 32–34). Viitteellinen arvo ilmaisee käytettävyyden suhteellista sisäistä vaihtelua, jota eri laatuelementit painotuksineen lisäävät tai vähentävät. Kokonaisprosentti ilmoittaa keskimääräisen käytettävyyden aineistotasolla.

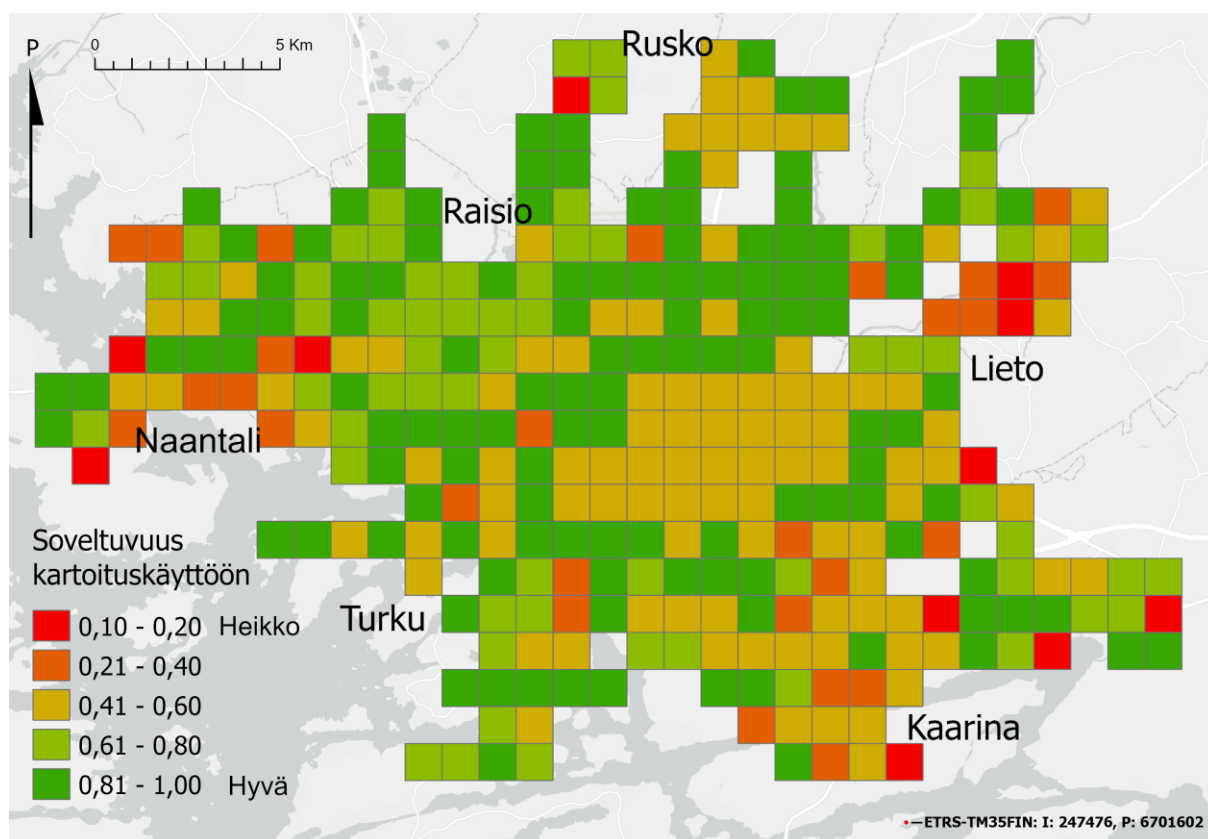
Paras soveltuvuus saadaan kartoitukselle (kuva 32), jonka käytettävyyden kokonaisprosentti on 70 %. Reititys ja navigointi (56 %) (kuva 33) ja uusi tieto (52 %) (kuva 34) saavat kartoitusta selvästi pienemmät arvot.

Taulukko 19. Laatu- ja käytettävyysarvioiden kuvaileva tilastotieto neliökilometrin resoluutiolla.

n = 319	Laatu				Käytettävyys		
	Sisältyminen %	Puuttuminen %	Sij.tarkkuus m	Topologia kpl	Kartoitus 0–1	Reititys 0–1	Uusi tieto 0–1
Keskiarvo	32,8	82	1,63	2,24	0,70	0,56	0,51
Mediaani	6,6	85,3	1,53	2	0,70	0,60	0,60
Keskihajonta	115,1	14,7	0,84	2,52	0,24	0,30	0,29
Maksimi	1281	11,7	18,02	11	1	1	1
Minimi	0	0,001	0,002	0	0,01	0,05	0,05

4.3.1 Käytettävyys peruskartoitukseen

Aineiston soveltuvuutta kartoituksen käyttötarkoituksen kuvaava ADQR-arvo vaihtelee verrattain tasaisesti (kuva 32). Sen keskiarvo on 0,7 ja keskihajonta 0,24 (taulukko 19). Turun ja Liedon välille sijoittuu laaja alue, jolla käytettävyys on keskitasoa ja matalammalla ympäröivään hyvätasoiseen alueeseen verrattuna. Kaarinan ydinalueen seutu korostuu laajuudeltaan ja arvoiltaan saman tyyppisenä. Myös Lieto, Naantali ja Jäkärän seutu osoittavat vastaavaa suuntausta. Raisio näyttäytyy hyvätasoisena poikkeuksena. Yleisesti hyvätasoinen aineisto sijoittuu ydinkeskusten ulkopuolelle. Se ei kuitenkaan muodosta laajoja jatkuvia alueita.

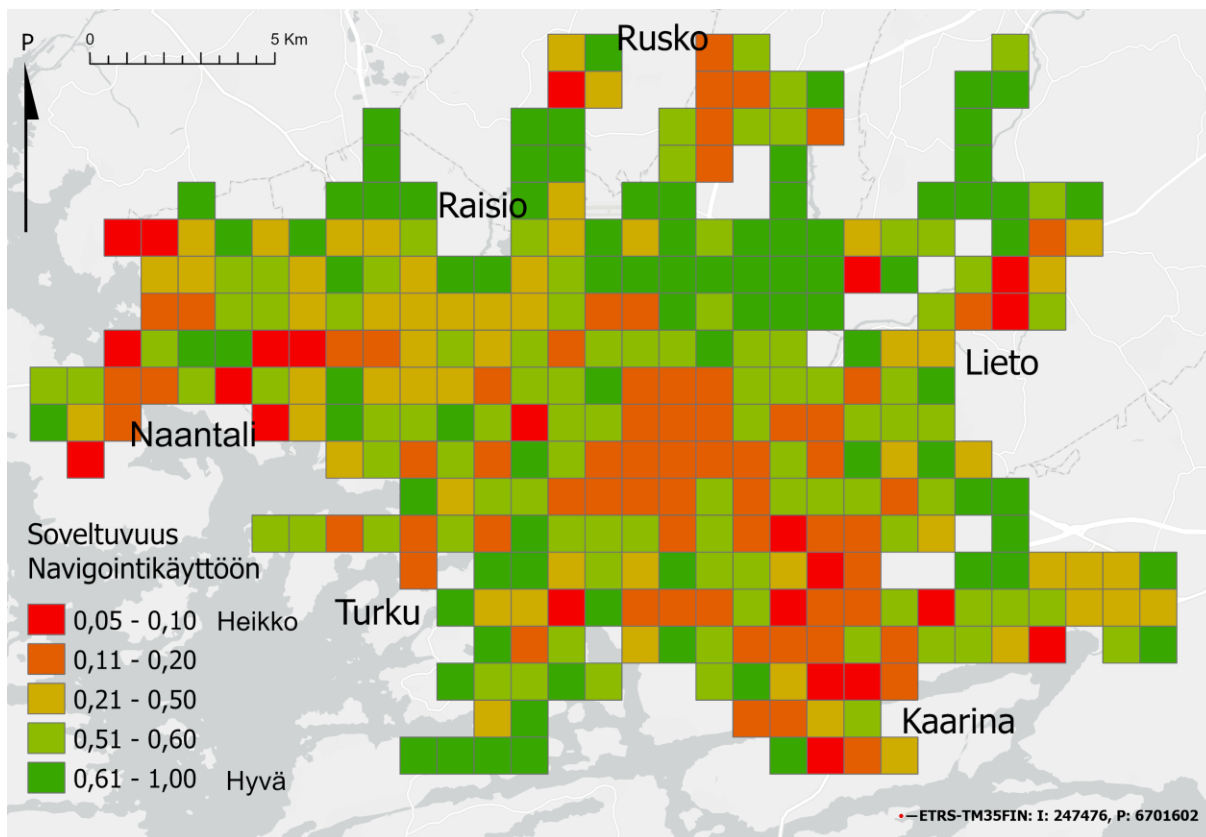


Kuva 32. Testiaineiston soveltuvuus kartoituskäyttöön. Käytettävyyskriteerein painotettujen laatumittaus tulosten aggregaatti ilmaisee soveltuvuuden välillä 0–1. 1 = paras soveltuvuus. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

4.3.2 Reititys- ja navigointikäyttö

Reitityksen ja navigoinnin käyttötarkoituksen laatuvaatimuksia painottanut ADQR-arvo seuraa pääpiirteissään kartoituksen ADQR-arvon yleisiä huomioita (kuva 33). Sen keskiarvo on 0,56 ja keskihajonta 0,30 (taulukko 19). Kuntakeskusten ydinalueilla käytettävyys on

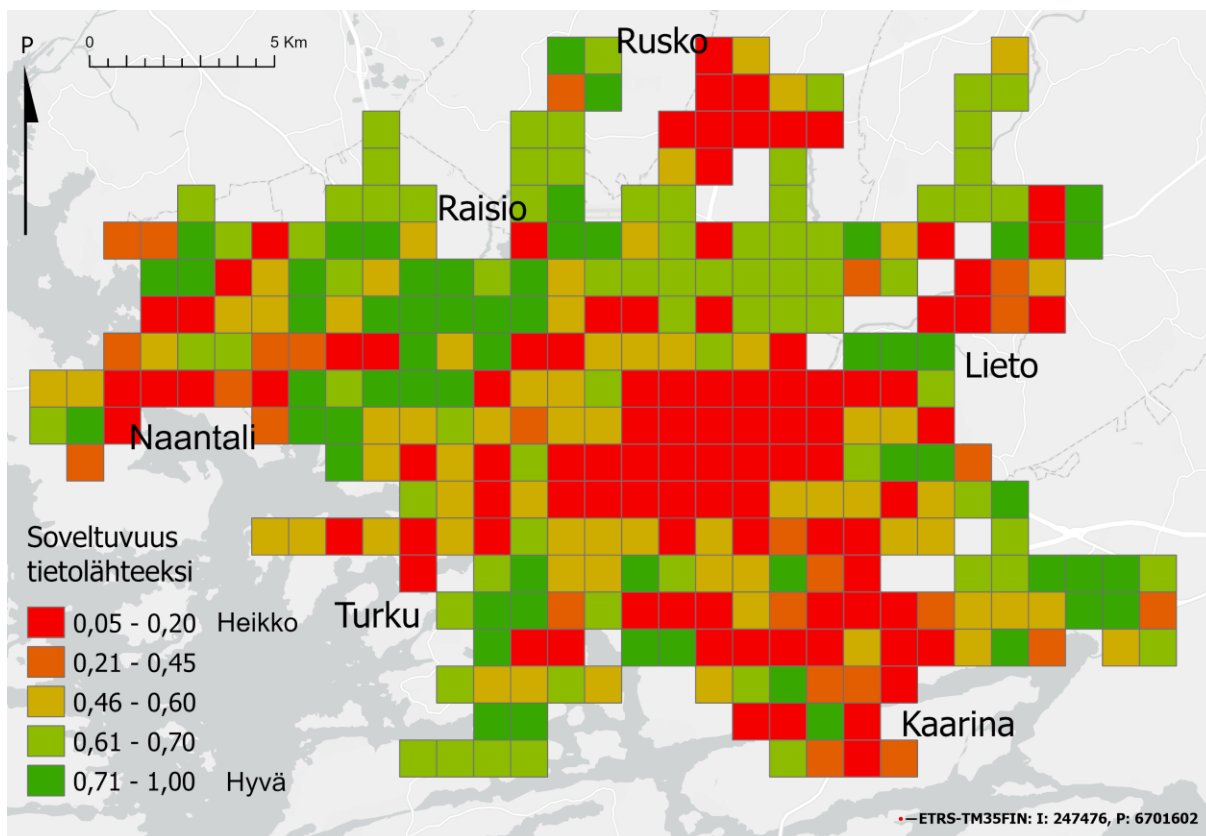
heikkotasoisempaa kuin niiden reuna-alueilla. Raisio muodostaa edelleen poikkeuksen korkeammalla käytettävyydellään. Turun ydinkeskustasta itään, Liedon suuntaan levittyy yhtenäinen heikomman käytettävyyden vyöhyke ympäröiviin alueisiin verrattuna.



Kuva 33. Testiaineiston soveltuvuus navigointikäyttöön. Käytettävyysskriteerein painotettujen laatumittaus tulosten aggregaatti ilmaisee soveltuvuuden välillä 0–1. 1 = paras soveltuvuus. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS

4.3.3 Käytettävyys uuden tiedon lähteenä verkostosuunnittelussa

Verkostosuunnittelukäytössä kuntakeskusten ydinalueet saavat matalia arvoja. Raision keskusta esiintyy tästä poikkeuksena (kuva 34). Uuden tiedon ADRQ-arvon keskiarvo on 0,51 ja keskihajonta 0,29 (taulukko 19). Kaikissa käytettävyysskartoissa erottuva Turun keskustasta itään levittyvä alue muodostaa laajan yhtenäisen matalien arvojen vyöhykkeen. Matalat arvot korostuvat myös Kaarinan alueella. Käytettävyyden korkeammat arvot sijoittuvat Raisiota lukuun ottamatta etäämmälle kuntakeskusten ydinalueista.



Kuva 34. Testiaineiston soveltuvuus tietolähteeksi. Käytettävyysskriteerein painotettujen laatumittaus tulosten aggregaatti ilmaisee soveltuvuuden välillä 0–1. 1 = paras soveltuvuus. Neliökilometrin resoluutio. Taustakartta: Lantmäteriet, Maanmittauslaitos, Esri, Garmin, Foursquare, GeoTechnologies Inc, METI/NASA, USGS.

5 Tulosten tarkastelu

Tutkielman päätavoitteena oli tutkia joukkoistetun paikkatiedon luotettavuutta sekä päivittää käsitystä sen laadusta ja soveltavasta hyödyntämisestä. Luotettavuusarvion muodosti vastaavuus auktoritatiiviseen vertailuaineistoon, käytettävyys erityyppisissä sovelluskohteissa sekä vastaavuuden ja käytettävyyden alueellisen vaihtelun mittaaminen.

Verrattaessa joukkoistettua ja auktoritatiivista aineistoa kysyttiin, 1) **Kuinka laadukasta OSM-aineisto on ISO 19157-standardin avulla arvioituna?** Kysymys tarkasteli joukkoistetun aineiston laatutasoa ja suorituskäkyä suhteessa tuottajan laatuvaatimukseen. Arviolle mahdollisia laatuparametreja OSM-aineistossa olivat täydellisyys, sijaintitarkkuus ja topologinen eheys.

Joukkoistetulle paikkatiedolle tyypillinen tuotantotekijöiden heterogeenisyys ilmenee tilassa esiintyvänä laatuvarhteluna, joka heikentää aineiston luotettavuutta. Tätä selvitettiin kysymällä, 2) **miten OpenStreetMap-aineiston laatu ja käytettävyys vaihtelevat alueellisesti?**

Tuottajan vaatimusten lisäksi kokonaislaadun muodostaa sopivuus käyttötarkoitukseen, sillä laadun merkitykset vaihtelevat eri käyttöyhteyksissä. Tätä tutkittiin kysymällä, 3) **Kuinka käytettävää OSM-aineisto on pyöräilyverkon kartoitussovelluksissa?** Seuraavassa kysymyksiin vastataan liittämällä laatu- ja käytettävyysarvioiden tulosten tarkasteluun niiden alueellisen vaihtelun tarkastelu. Arvioiden tuloksia peilataan samalla myös aiempaan tutkimuskirjallisuuteen. Luotettavuusarvio pohjaa osa-alueista tehtyihin johtopäätöksiin.

5.1 Kokonaislaadun osatekijät ja aiempi tutkimus

Laatu- ja käytettävyysarvioiden tulokset sekä niiden alueellisen vaihtelun tarkastelu osoittavat arvioidun OSM-aineiston olevan kokonaislaadultaan välttävä. Kun kokonaislaadua tarkastellaan osatekijöittäin ja niiden rakenneosat edelleen arvioidaan erillisinä laatulementteinä ja käyttötarkoituksina, näyttäytyy laatu parempana. Aineistotason laatu vaihtelee laatulementtien välillä huomattavasti, mikä rajaa aineiston käytettävyyttä vaihtelevissa sovelluksissa. Lisäksi käyttötarkoituksen monimutkaisuus heikentää käytettävyyttä, voimistaen jo olemassa olevia tuottajan kvantitatiivisen laadun heikkouksia. Aineiston luotettavuuden voi tulosten perusteella katsoa rakentuvan kolmesta tekijästä, joissa vaihtelu esiintyy: laadun ja käytettävyyden tasoista, laatulementtien alueellisesta vaihtelusta

ja alueellisen vaihtelun eroista laatulementtien välillä. Yleisesti voi todeta vaihtelun olevan huomattavaa tilassa ja elementtien välillä. Rakennneosien erillisessä, ja alueellisessa, tarkastelussa laadun havaitaan olevan osittain hyvätasoinen.

Laadun rakenneosia voidaan lopulta verrata aiempaan tutkimukseen, jossa laadun osittainen tarkastelu on hallitseva. OSM-liikenneverkostojen laatu tutkimus on painottunut moottoriliikennöityihin teihin. Kotimaista laatu tutkimusta OSM-pyöräilyverkostoista ei löydy, kansainvälisiä löytyy täydellisyydestä ja sijaintitarkkuudesta (Hochmair ym. 2014; Ludwig 2011). Moottoriliikenneverkostoja koskevat tulokset on otettu mukaan tarkasteluun, koska ne antavat viitteen lineaarikohteiden laadun tavanomaisesta tasosta OSM-palvelussa ja laatuun vaikuttavista tekijöistä. Tematiikan lisäksi mittakaavat poikkeavat tieverkostojen arvioiden ja tutkielman välillä, joten tulokset eivät ole täysin vertailukelpoisia. Tieverkostojen arvioissa urbaani-ruraali-aluejako ja väestötiheyden vaikutus korostuvat, sillä tutkimusalueet ovat maakunta- ja osavaltiomittakaavoissa. Tutkimukset sijoittuvat globaalin pohjoisen kehittyneisiin maihin, vastaten tältä osin tutkielman tuloksia. Laatu taso on tyypillisesti korkea kaupunkialueilla ja laskee maaseudulle siirryttäessä huomattavasti (Ludwig ym. 2011, Girres & Touya 2010, Haklay 2010, Wu ym. 2020).

Täydellisyyden tuloksista erottuvat seuraavat yleiset huomiot. Täydellisyys on aineistotasolla hyvin lähellä laatuvaatimusta, vaikka ei puuttumisen osalta (kattavuus) täysin yllä siihen. Alueellisesti sekä puuttuminen että sisältyminen vaihtelevat kuntien ydinkeskustojen osittaisten matalan laadun kasaumista ympäristöjen laajoihin hyvälaatuisiin alueisiin (kuvat 20a ja 20b). Girres & Touya (2010) painottivat keskihajonnan tärkeyttä luotettavuuden osoittajana. Täydellisyyden molempien alaelementtien arvojen vaihtelu keskiarvon ympärillä (taulukko 19) ja samalla laatuvaatimuksen rajalla on suurta. Toisaalta laajoilla yhtenäisillä alueilla laatuvaatimus täyttyy tai täydellisyys on lähellä sitä. Sisältyminen kohdalla tämä on todennäköisesti osin myös temaattisen tarkkuuden puute.

Moottoriliikennöityjen OSM-tieverkostojen täydellisyys (puuttuminen) on Saksan ja Quebecin kaupunkialueilla yli 90 % tasolla (Ludwig ym. 2011; Moradi ym. 2021). Maakuntamittakaavassa kattavuus on urbaani-ruraali-liukumalla Skånessa 87 % – 69 % ja Saksassa 79,8 % – 50,8 % (Abdolmajidi ym. 2015; Ludwig ym. 2011). Hohmairin ym. (2014) OSM-pyöräilyverkostojen laatu arvio tulokset Portlandin ja Miamin kaupunkialueilla soveltuvat hyvin täydellisyyden tuloksien vertailuun. Molemmissa kaupungeissa sekä pyöräilykaistojen puuttuminen (95,3 % ja 99,2 %) että sisältyminen (2,8 % ja 0,1 %) ovat

erittäin hyvätasoisia. Pyöräilypolkujen puuttumisen tulos on sen sijaan (86,4 % ja 22,8 %) ja sisällyttämisen (22 % ja 1,4 %).

Moottoriliikennöityjen teiden tutkielman tuloksia paremmat puuttumisen tulokset voivat selittyä maantieteellisen epävarmuuden vähäisyydellä. Tiet voivat olla luonteeltaan pyöräilyreitistöä paremmin määriteltyjä kohteita ja näkemykset kohteiden määrittelyssä auktoritatiivisen tuottajan ja OSM-yhteisön välillä ovat todennäköisesti lähempänä toisiaan. Teillä on myös suurempi käyttäjäkunta ja näin ollen suurempi kartoitustarve, jolloin niihin myös kohdistuu suurempi huomio.

Tutkielman puuttumisen tulokset vertautuvat tasoltaan Hochmairin ym. (2014) pyöräilypolkujen tuloksiin Portlandissa. Merkittyjen kaistojen laatu on kuitenkin molemmissa kaupungeissa paremmalla tasolla. Polut on tutkimuksessa määritelty autoliikenteestä erotetuksi erillisiksi pyöräilyreiteiksi, jotka sallivat myös kävelyn. Kaistat ovat autoliikenteen väylistä erotettuja pyöräilykaistoja. Tutkielman pyöräilyverkoston määritelmä puolestaan kattoi pyöräilylle merkityt reitit, joiden tyyppi voi vaihdella kaistoista metsäteihin. OSM-käyttäjäkuntaan liittyvien heterogeenisyystekijöiden lisäksi pyöräilyinfrastruktuurien erot ja luokittelun temaattisen tarkkuuden vaihtelu (ISO 19157 2013) viranomaisten ja tutkimusten määritelmässä voivat selittää eroja. Peilattaessa tutkielman täydellisyyden tuloksia yleisesti aiempaan tutkimukseen, voi tulosten todeta olevan samaa luokkaa, mutta hieman heikompia.

Havaitun perusteella täydellisyyden laatua voi pitää yleisesti melko hyvänä huolimatta siitä, ettei puuttuminen (kattavuus) täytä laatuvaatimusta täysin. Alueellinen laatuvahtelu kuitenkin heikentää täydellisyyden luotettavuutta aineistotasolla selvästi.

Huomionarvoista on, ettei aiemmassa tutkimuksessa havaittu täydellisyyden sekä urbaanirakenteen ja väestön tiheyden positiivinen korrelaatio saa vahvistusta tuloksista visuaalisessa tarkastelussa (Haklay 2010, Moradi ym. 2021). Tulokset osoittavat yleisesti täydellisyyden olevan heikompaa urbaanirakenteen ytimissä, kun taas reuna-alueilla havaitaan parempi laatutaso. Jos havainto on ilmentymä kaupunkirakenteesta laatuindikaattorina, sen poikkeavuutta aiemmista havainnoista voi selittää mittakaava, sillä kirjallisuudessa esitetty käsitys urbaani-ruraali-liukumasta sellaisena on muodostunut tarkasteltaessa kansallisen ja maakuntatason tutkimusalueita (Ludwig ym. 2011, Girres & Touya 2010, Wu ym. 2020). Toinen mahdollinen selittäjä voi olla ajantasaisuus eli hiljattain keskustoissa toteutettu pyöräilyverkoston kehitystyö. Varauksena on painotettava johtopäätöksen nojaavan visuaaliseen tarkasteluun. Laatuindikaattorit eivät olleet tutkielmassa mittauskohteena.

Sijaintitarkkuuden ja täydellisyysden tulokset muistuttavat yleiseltä tasoltaan toisiaan. Täydellisyysden tavoin sijaintitarkkuus on aineistotasolla melko hyvä, mutta ei täysin yllä laatuvaatimukseen (taulukko 18). Sijaintitarkkuuden alueellinen vaihtelu ei kuitenkaan kasaudu täydellisyysden tavoin heikkoina keskittyminä kuntakeskuksiin, vaan tasalaatuisuus on huomattava. Laatuvaatimus täyttyy miltei koko aineiston laajuudelta. Harvat heikon laadun alueet ovat verkoston reuna-alueilla. Tämän perusteella sijaintitarkkuuden luotettavuutta voidaan pitää hyvänä ja se näyttäytyy parempana kuin aineistotason tuloksessa. Tasalaatuisuuden ero täydellisyysden ja sijaintitarkkuuden välillä viittaa temaattiseen epätarkkuuteen OSM-aineistossa tai mahdollisesti tutkimuksen metodologiassa. Epätarkkuus ilmenee täydellisyysden kohdalla, sillä sijaintitarkkuuden mittaus edellytti vertailtavien aineistojen vastaavuutta. Tuloksen voi tulkita osoittavan, että kattavuuden kartoittaminen on OSM-yhteisölle sijaintia haasteellisempaa.

Kanadassa 83 % – 77 % moottoriliikennöidyistä teistä on havaittu sijoittuvan alle viiden metrin etäisyydelle vertailuaineistosta (Moradi ym. 2021, Jacobs 2018, Zhang 2018) ja Saksassa 73 % (Ludwig ym. 2011). Quebeckissä 60 % teistä sijoittuu alle 3 metrin säteelle. Sijaintitarkkuus vaihtelee urbaani–ruraali-liikumalla kaupunkialueiden hyväksi sekä korreloi positiivisesti väestötiheyden ja tulotason mukaan. (Moradi ym. 2021, Girres & Touya 2010, Haklay 2010). Saksassa OSM-pyöräilyverkostosta 61,5 % on ollut alle viiden metrin säteellä ja 38,5 % 10–30 metrin säteellä vertailuaineistosta (Ludwig ym. 2011).

Tutkielman OSM-aineistosta kolmen metrin etäisyydelle vertailuaineistosta sijoittuu 86,5 %, viiden metrin etäisyydelle 91,5 % ja 12,5 metrin etäisyydelle 96,5 %. Tulokset vertautuvat aiempaan tutkimukseen huomattavasti parempina. Syitä eroon voivat olla joko ajan myötä parantuneet OSM-yhteisön kyvykkyys tai kartoituksen työkalut. Merkit voivat viitata myös kollektiivisen älykkyyden ilmentymään, jonka läsnäolosta olisi editointihistoriaan nojaten mahdollista esittää johtopäätöksiä (Lévy 1999; Raymond 1999).

Sijaintitarkkuuden laadun alueellisen vaihtelun voi katsoa seuraavan aiempia havaintoja urbaani-ruraali vaihtumisen ja väestötiheyden yhteydestä laatuvahteluun. Toisaalta hyvälaatuinen alue seuraa yleisesti Turun kunnan rajoja (Kuva 28), mutta korkea asutustiheys levittyy myös Raisio-Kaarina-akselilla. On mahdollista, että Turun kunnan alueen hyvä laatu voi selittyä myös kuntien välisillä tuotantoprosessien eroilla, sillä jokainen ympäröivä kunta tuottaa oman verkostoaineistonsa ja toimittaa ne Turulle, joka koostaa yhtenäisen mallin (Pärssinen 2022). Havainnot kaupunkirakenteen vaihtumisen

vaikutuksesta on tehty Turun kaupunkialuetta laajemmilla tutkimusalueilla maakuntamittakaavassa, jolloin ne eivät välttämättä tue valtaosin urbaanista ympäristöstä saatujen tulosten tulkintaa.

Topologinen eheys on aineistotasolla tasaisesti heikkolaatuinen, vaikka heikko laatu kasautuukin vähäisessä määrin tiheimmän kaupunkirakenteen alueelle. Laatuvaatimuksen edellyttäessä ehdotonta virheettömyyttä, virheiden vähäinenkin lukumäärä ja alueellinen vaihtelu heikentävät luotettavuutta huomattavasti.

Verratessa topologisen eheyden tulosta aiempaan tutkimukseen, virheiden määrä OSM-testiaineistossa on 0,8–1,9 kertaluokkaa suurempi. Zacharopoulou ym. (2021) havainnoivat tielinkkien päällekkäisyyttä, mikä mahdollistaa tulosten vertailun risteysiin ilman liittymää ja osin melkein liittymiin (overshoots). Heidän tutkimuksessaan OSM-moottoriliikenneverkostojen topologisten virheiden määrä kuudessa suuressa eurooppalaisessa kaupungissa oli 0,1 ‰. Testiaineistossa risteysia ilman liittymää oli 7,7 ‰, melkein liittymiä 2 ‰. Tutkimusaineistojen temaattisesta erosta huolimatta, kertaluokkaerot osoittavat aineiston olevan topologisesti huomattavan heikkolaatuinen.

Girres ja Touya (2010) havaitsivat Ranskan moottoriliikennöidyn tieverkoston yhdistävyuden olevan OSM-aineistossa 95 %. Koska tutkielman tutkimusalue oli valtaosin urbaani ja tematiikka poikkeava, erillisten saarten 0,6 ‰ tulos ei ole täysin vertailukelpoinen. Sen voi kuitenkin tulkita vahvistavan käsitystä laadun riippuvuudesta urbaali-ruraali-liikumasta myös topologisen eheyden kohdalla (Haklay 2010; Wu ym. 2020). Barrington-Leigh ja Millard-Ball (2017) sekä Basiri ym. (2016) pitivät topologisen eheyden sekä sijaintitarkkuuden ja täydellisyyden välistä korrelaatiota voimakkaana. Tutkielman sijaintitarkkuuden hyvät, ja täydellisyyden kohtalaiset tulokset eivät näytä kytkeytyvän heikkoon topologiseen laatuun samalla tavoin. Testiaineiston topologisen eheyden heikko tulos on yllättävä siksi, että OSM-yhteisö tarjoaa automaattiset työkalut ja ohjeistuksen topologian laadunvarmistukseen (Quality... 2023). Topologisen skeeman soveltaminen jää kuitenkin käyttäjien huolellisuuden varaan (OSM XML 2023), koska laadunvarmistus ei ole pakollinen osa tuotantoprosessia (Sehra ym. 2020).

Käytettävyyssarvio edustaa eksploratiivista tutkimusta ja sen tulokset ovat yleisesti suuntaa antavia. Niiden pohjana olevat laatuarvion tulokset on painotettu tutkimuskirjallisuuteen perustuen käyttötarkoituksen kannalta keskeisillä laatuparametreilla (Zieltsra & Willem 2014; Neis 2015; El-Ashmawy 2016). Painokertoimien voimakkuus nojaa kuitenkin subjektiiviseen

näkemykseen. Käytettävyyssarvio antaa mahdollisuuden tarkastella aineiston käyttötarkoitukseen sopivuuden suhteellista sisäistä vaihtelua alueellisesti ja ilmentää laadun vaihtelevia merkityksiä käyttäjän näkökulmasta. Vaikka tiettyjen laatupiirteiden merkitys käytettävyydelle korostuukin, nojaa se aina kaikkien piirteiden kokonaisuuteen.

Aineistotasolla paras soveltuvuus käyttötarkoitukseen on kartoituskäytössä, jonka keskiarvo on 0,70 ja keskihajonta 0,24 (taulukko 19). Sen kohdalla painotettiin täydellisyyttä ja sijaintitarkkuutta, joiden laatutaso on yleisesti hyvä. Täydellisyyden ja sijaintitarkkuuden vaikutuksen katsottiin olevan keskeisempi kuin topologisen eheyden.

Reitityksen ja navigoinnin sekä uuden tiedon käyttötarkoituksissa vaihtelu on suurempaa. Niissä yksittäisen laatuparametrin vaikutus käytettävyyteen on asetettu merkittävämmäksi. Edellistä selittää heikko topologinen laatu ja jälkimmäistä vertailuaineistoon kuulumaton aineisto, joka ei testiaineistossa jakaudu tilassa tasaisesti, vaan verkoston reuna-alueille. Uuden tiedon tulos ei täysin onnistu kuvaamaan sen käyttötarkoituksen merkitystä. Jos sen taso olisi kaikkialla 100 %, tutkimusalue olisi täynnä vertailuaineistoon kuulumatonta testiaineistoa, mikä ei pyöräilyverkoston kehitystoimenpiteiden kannalta ole realistista. Aineistotason tulokset eivät rohkaise navigoinnin sovelluksiin, mutta verkostosuunnitteluun se voi tarjota mielenkiintoisia avauksia.

Informatiivisinta tuloksissa on käytettävyyden alueellinen vaihtelu tarkasteltuna luotettavuuden perspektiivistä (kuvat 31–33). Yhteenvedona käyttötarkoitusten tuloksista voidaan huomioda miten tummanvihreät korkean käytettävyyden alueet ovat jopa kartoituskäytössä vähäiset verrattuna eri tasoihin matalan käytettävyyden alueisiin. Yhtenäisiä laajoja hyvän käytettävyyden alueita ei erotu, vaan voimakas vaihtelu rikkoo niitä koko tutkimusalueen laajuudella. Matala käytettävyys on tyypillistä kuntakeskusten ydinalueille. Toteutetulla tavalla arvioituna, luotettavuutta ei voi pitää tutkituissa käyttöyhteyksissä mainittavana.

Tulokset seuraavat Frankin (2007) ajatusta aineistomuodostuksen aikana tehtyjen sitoumusten kumuloituvasta vaikutuksesta lopputuotteen laatuun. Havainnot viittaavat kaikkien arvioitujen laatuparametrien olevan käytettävyyden kannalta olennainen osa kokonaislaatua, riippumatta käyttötarkoituksesta. Erillisinä arvioidut laadun osa-alueet voivat näyttäytyä hyvätasoisina, mutta käyttötarpeiden monimutkaistuessa, nähdään virheiden vaikutuksen nopeasti kasvavan.

Paikkatiedon käyttötarkoitusten, erityisesti reitityksen, kannalta vähäisilläkin virheillä voi olla huomattava merkitys. Toisaalta, vaikka aineisto ei täytä reitityskäytön topologisia vaatimuksia, mahdollistaa se kartoituskäytön huomattavasti paremmin. Tällöin monipiirteisen joukkoistetun paikkatiedon tehokkaassa hyödyntämisessä korostuu oikean aineiston, tai osa-aineiston, käyttäminen oikeassa asiayhteydessään. Tämä voi kuitenkin nostaa aineiston käsittelykustannuksia ja siten käyttökynnystä.

Uuden tiedon käyttötarkoituksen voi katsoa olevan teoreettinen julkisen ja kaupallisen paikkatietotuotannon kannalta. Yleisesti se kuitenkin ilmentää joukkoistetun paikkatiedon potentiaalia kuvata yhteiskunnan “ruohonjuuritasoa” koskevaa informaatiota (McCullagh & Jackson 2013). OSM-yhteisön kartoittama pyöräilyverkosto poikkeaa julkisesta näkemyksestä ja voi näin tarjota ihmisten mobiliteetteihin liittyvää uutta tietoa.

Katsottaessa laatuarvion tuloksia kokonaisuutena nähdään, että laadun urbaani-ruraali-vaihtuminen alkaa esiintyä laajemmassa alueellisessa mittakaavassa, kuten maakuntatasolla, mutta urbaanien ympäristöjen ja niiden lähialueiden paikallisessa tarkastelussa sitä ei havaita ainakaan yhtä voimakkaana. Lisäksi vaihtuminen käyttäytyy laatulementtien välillä eri tavoin. Tähän liittyen on huomionarvoista, etteivät aiemmassa tutkimuksessa havaitut laatulementtien väliset korrelaatiot saa vahvistusta tutkielman tuloksista. Havaintojen voi tulkita viittaavan mittakaavatasojen merkityksellisyyteen laadun vaihtelevassa ilmenemisessä.

5.2 Menetelmälliset haasteet ja riskit

Tutkimuksen validiteetti pyrittiin varmistamaan aineiston riippumattomuudella ja vakiintunein mittausten menetelmin, joiden ennakkoehtojen täytyminen huomioitiin. Tutkimusasetelman soveltuvuutta joukkoistetun aineiston ja kokonaislaadun arvioon tuki kirjallisuudessa vallinnut konsensus. Seuraavassa näkökohdat avataan tarkemmin.

Laatututkimuksen asetelma vastasi paikkatietoalan ISO 19157 (2013) -standardiin perustunutta vertailevaa ulkoista laatuarviota, joka soveltuu suorituskykymittaukseen (Fonte ym. 2017). Vertailuaineistona käytettiin vakioitua laadunhallintaa noudattaen tuotettua auktoritatiivista aineistoa (Paikkatieto... 2023, KMTK... 2020). Tällöin informaatiolähteiltään riippumattoman vertailun validiteettia voidaan pitää luotettavana (Mocnik ym. 2018). Lähestymistavaksi valittiin joukkoistetun paikkatiedon heterogeenisten lähtökohtien edellyttämä alueellisen vaihtelun huomioiva kokonaistutkimus (Longley ym. 2015: 42–43).

Riippumattomuutta voivat jossain määrin heikentää OSM-tietokantaan tehdyt viranomaisaineiston massatuonnit (Import... 2023). Aineistoilla on mahdollisena yhteisenä lähteenä Maanmittauslaitoksen Maastotietokanta, josta Digiroad aineisto saa geometriatietonsa. Maastotietokannan tiegeometriaa on tuotu vaihtelevalla aktiivisuudella vuodesta 2013 lähtien OSM-tietokantaan. Projekti on edelleen ensimmäisessä vaiheessaan, jossa tuodaan ainoastaan ei-päällekkäistä paikkatietoa. Tietoa Turun ydinkaupunkialueella tehdystä tuonnista ei löydy. OSM-yhteisön koostuessa yksilöistä ja institutionaalisista toimijoista (Yu-Wei Lin 2011), osittaisen riippuvuuden voi kuitenkin nähdä luonnollisena lähtötilanteena. Näkemystä tukee Zielstran ym. (2013) laadunvarmistuksessa havaitsema synerginen tehokkuus yksilökartoittajien jälkeisjalostaessa massapäivityksiä.

Testiaineistosta mitattavaksi mahdollisten laatuparametrien analyysiin valittiin kirjallisuuskatsauksen perusteella (taulukot 5, 7 ja 8) aiemmassa tutkimuksessa vakiintuneimmat mittarit (taulukko 12). Tämä vahvisti tulosten luotettavuutta ja vertailtavuutta aiempaan tutkimukseen. Laatuä pyrittiin tutkimaan kokonaisvaltaisesta perspektiivistä yhdistäen kirjallisuudessa sen tekijöiksi tunnistetut tuottajan ja käyttäjän näkökulmat (Devillers & Jeansoulin 2006: 31–41; Mocnik ym. 2018).

Alueellinen laatu vaihtelu mitattiin kilometriresoluutiolla. Viitaten yleisesti muuntuvan alueyksikön ja ryhmitymisen ongelmiin, on mahdollista, että toinen mittausyksikkö olisi tuottanut osin erilaiset tulokset. Tutkimuskirjallisuudessa resoluutiovalinnat on perusteltu heikosti. Samoin spatiaalisen autokorrelaatiotestien tulokset ovat riippuvaisia naapuruston määrittelystä. Sen määrittelyyn valittiin, kirjallisuudessa menetelmän yhteydessä yleisesti käytetyin shakkitermein, kuningattaren tapaus etäisyyden mukaan heikentyvällä koko aineiston kattavalla vaikutussäteellä. Tämän katsottiin kuvaavan parhaiten OpenStreetMap-aineistotuotannon paikallisten prosessien vaikutusta 319 km² tutkimusalueella.

Täydellisyysmittauksen tarkkuutta paransi puuttuvan ja sisältyvän aineiston erottaminen Liun ym. (2015) menetelmällä. Tämä tuotti monia aiempia tutkimuksia täsmällisemmän tuloksen. erottamiseen käytetyn puskurin leveys oli 12,5 metrin säteellä 25 metriä. Se on enemmän kuin liikenneväylien leveyden mahdollinen vaihteluväli, mutta ei sisältänyt vertailuaineistoon kuuluvaa aineistoa. Säteen voi näin odottaa erottavan laatuparametrit täsmällisesti. Sijaintitarkkuuden arvion edellyttämä täydellisyysehto voitiin myös täyttää perustuen tähän arvioon. Mittaus varmennettiin Tveiten ja Langaasin (1999) virhekoodausten tunnusluvulla.

Sijaintitarkkuuden mittaamiseen valittu menetelmä huomioi geometrisen eroavuuden ja sieti ääriarvoja (Tveite & Langaas 1999). Santos ym. (2015) ja França ym. (2018) olivat myös havainneet sen vaihtoehtoista tarkimmaksi, jonka tarkkuus vastasi auktoritatiivisessa tuotannossa käytettyä pistepohjaista menetelmää (Santos ym. 2015). Laatuvaatimukseen vertailu edellytti tarkkuuden epävarmuustason selvittämistä, mihin Mozas-Calvache (2021) suositteli puskurimenetelmiä pistepohjaisten sijaan. Mittauksen tarkkuus varmistettiin Tveiten (2023) laatimalla automatisoidulla versiolla Päällekkäisten puskurien tunnusluvut -menetelmästä.

Sijaintitarkkuuden täsmälliseen arviontiin soveltuvan puskurisäteen valintaa perusteli täydellisyyden tunnusluvun 13 metrin puskurisäde, joka sisälsi 99,4 % testiaineistosta ja ylitti Maastotietokannan laatuluokan B kattavuuden todennäköisyyden 95 %. Luvut ovat noin neljä kertaa vertailuaineiston keskimääräinen tarkkuus, jota Tveite ja Langaas (1999) suosittelivat säteeksi. Käsittelemättömän testiaineiston täydellisyydsmittaus myös osoitti kuvaajan muutostrendin olevan jatkuva, mutta lisääntyvän vain verkkaisesti aina mitattuun 26 metrin säteeseen saakka saavuttamatta 100 prosenttia. Tämä perusteli sijaintitarkkuusarvion tarkasteluvälin lyhentämistä.

Sijaintitarkkuutta ja tarkkuusarvion luotettavuutta heikentävää spatiaalista painottuneisuutta ei havaittu (Tveite & Langaas 1999). Mediaanikeskustojen etäisyysarvojen heikko kasautuminen tulkittiin käytännössä merkityksettömäksi. Atsimuuttiarvojen heikko kasautuminen tulkittiin vain yhdeksi esimerkiksi lukemattomista kasautumisen satunnaisista vaihtoehtoista.

Oskillaatioarvo viittasi painotuksista vapaaseen aineistoon ja vahvasti mediaanikeskustojen etäisyyksien ja suuntausten tuloksia. Kolme samansuuntaista tulosta tulkittiin spatiaalisen painottuneisuuden poissaoloksi OSM-aineistossa. Myöskään sijaintitarkkuusarvion Siirtymätietokaavion kuvaajien muutosnopeudet eivät kasvaneet tai vähentyneet odottamattomasti, mikä viittasi spatiaalisista painotuksista vapaaseen aineistoon. Kuvaajien käyttäytyminen osoitti verkostojen geometrian muistuttavan läheisesti toisiaan.

Sijaintitarkkuuden tulokset varmennettiin algoritmisesti QGIS-liitännäisellä neljän mittausyksikön otoksella. Mittauksen ja varmennuksen välinen ero oli pienempi kuin varmennuksen tulosten sisäinen ero puolen metrin välein mitatuilla puskurisäteiden arvoilla. Sijaintitarkkuuden mittaustuloksia pidettiin tämän perusteella luotettavina.

Topologisen eheysmittauksen luotettavuus nojaa korkealuokkaisella tuotteella suoritettuun automaattiseen validaatioon sekä virhetyyppien ja algoritmien keskinäisen vastaavuuden oikeellisuuteen. Mitattujen topologisten virheiden valinta pohjasi tutkimuskirjallisuudessa yleisimmiksi tunnistettuihin topologisiin poikkeamiin (Wu ym. 2020; Kukulska-Koziel ym. 2018). Näitä vastaavat topologisen skeeman säännöt luotiin, ja niiden validointi suoritettiin ArcGIS-algoritmein tuotteen dokumentaatioon perustuen (Creating a topology 2023).

Topologisten tulosten tarkkuutta parannettiin erottamalla poikkeamista todelliset virheet visuaalisella tarkastelulla. Risteys ilman liittymää -virheen automaattinen mittausta täydennettiin visuaalisella arviolla auktoritatiiviseen alikulkutietoon perustuen. Melkein liittymä -virheet arvioitiin osaksi manuaalisesti. Tässä arvio nojasi Brasselin ym. (1995) toteamukseen, jonka mukaan virhe on topologinen, jos se ei ole johdonmukainen aineiston topologisessa kokonaisuudessa. Muutoin se luetaan täydellisyyden virheeksi. Arviossa johdonmukaiseksi Melkein liittymäksi tulkittiin alle 10 metrin säteellä liittymäkohdistaan olleet virheet. Etäisyyden katsottiin suhteutuvan pyöräilyverkoston rakenteelliseen mittakaavaan sekä olevan Liun ym. (2015) menetelmällä saadun puskurisäteen luokassa. Virheiden kasautuminen 4,5 metrin säteelle tuki tätä näkemystä.

Käytettävyyttä arvioitiin alan standardin kuvaamalla tavalla (ISO 19157 2013), joka salli menetelmän sovellukselle laajan liikkumavaran. Mittaus perustui käyttötarkoitukselle olennaisten laatuparametrien painottamiseen kaikkien parametrien aggregaatissa. Painotukset asetettiin subjektiivisesti, mutta käyttötarkoituksille olennaisten parametrien valinta perustui tutkimuskirjallisuuteen. Zielstra ja Willem (2014) sekä Neis (2015) tunnistivat reitityskäytölle olennaisiksi laatuparametreiksi yhdistävyyden ja kattavuuden. El-Ashmawy (2016) piti sijaintitarkkuutta olennaisimpana kartoituskäytössä. Uuden tiedon käyttötarkoituksella oli vahvasti kokeellinen funktio.

Käytettävyyden tulos yhdistää tuottajan ja käyttäjän laatuäkökulmat kokonaislaatuun ilmaisevaksi arvoksi (Devillers & Jeansoulin 2006: 31–41). Tulosten merkitys on kokonaislaadun alueellisen vaihtelun kuvaamisessa, ja ne ovat luotettavia sisäisesti suhteellisen vaihtelun indikaattoreina. Niillä on myös merkitystä eksploratiivisessa vaiheessa jatkotutkimussuuntien osoittajina sekä käytettävyyden ilmenemismallien alustavassa tunnistamisessa. Tutkielmassa sovelletulla tavalla käytettävyyden tulokset jäävät väistämättä yleisluontoisiksi, eikä laatu tulosten subjektiivinen painotus salli käytettävyydestä yleistyksiä.

5.3 Jatkotutkimustarpeet

Kaupallinen sektori käyttää joukkoistettua paikkatietoa palveluissaan ja Euroopassa julkinen hallinto tutkii mahdollisuuksia sen hyödyntämiseen (Who... 2023; Olteanu-Raimond 2017). Myös Suomessa julkisen hallinnon strateginen suunnittelu on tunnistanut joukkoistetun paikkatiedon datatalouden arvokkaana raaka-aineena (Karlsson 2018). Tutkielman tulokset linjautuvat julkisen strategisen suunnittelun tavoitteeseen tunnistaa miten erilaiset paikkatietovarannot tukevat datatalouden mekanismien syntyä. (Kansallinen... 2021; Karlsson 2018).

Julkisen paikkatietoinfrastruktuurin vakioitu tuotantomalli tarjoaa paikkatietoa hyödyntävälle datataloudelle keskeisen tukirakenteen, jonka joukkoistettu vastine on OpenStreetMap (ERT... 2017; Mooney ja Minghini 2017). Mahdollisuuksia joukkoistetun paikkatiedon hyödyntämiseen molempien välillä on nähty mm. peruspaikkatietojen päivityksen, liikennetiedon jalostamisen, muutoshavainnoinnin, virheiden korjauksen ja temaattisesti uudenlaisen tiedon keräämisen alueilla (Olteanu-Raimond 2017; See ym. 2017). Kaupalliset toimijat käyttävät OSM-tietoa myös pohjakarttoina ja navigointiin (Major... 2023).

Haasteen hyödyntämiselle muodostavat kuitenkin validoinnin kustannukset, käyttöoikeuskysymykset, tietolähteen pysyvyys sekä datan laatu ja luotettavuus (Olteanu-Raimond 2017). Joukkoistetun paikkatiedon monet sovelluskohteet ja vaihtelevat ominaisuudet perustelevat laatututkimusta hyödyntämisyökkimysten pysyvänä toimintona. Spielmanin (2014) huomiot kollektiivisen älykkyyden ehdollisesta läsnäolosta joukkoistetuissa projekteissa ja Goodchildin (2009) tekemä ero joukkoistetun ja asiantuntijalähtöisen paikkatiedon välillä korostavat tätä tarvetta.

Tutkielman tulosten perusteella suomalaisten viranomaisten hahmottelemista OSM-tiedon ja paikkatietoinfrastruktuurin yhteentoimivuusskenaarioista realistisia ovat auktoritatiivisen peruspaikkatiedon ja joukkoistetun paikkatiedon yhteistoiminnallinen malli sekä joukkoistetun paikkatiedon käyttäminen referenssinä. Tulosten osoittama luotettavuus ei tue viranomaispaikkatiedon pohjaamista ensisijaisesti joukkoistettuun lähteeseen. (Tietoaineistojen... 2016). Havainnot tukevat Cinnamonin (2015) näkemystä hybridimallista hyödyntämisen käytännöllisimpänä lähestymistapana. Tulokset viittaavat kahden soveltuvan skenaarion voivan mahdollistaa kustannussäästöt valtakunnallisen pyöräilyverkotomallin luomisessa, jonka edellyttämä verkoston inventointi on nähty pienissä kunnissa resurssihaasteena (Laitinen ym. 2015).

Kunnianhimoisimmissa visioissa joukkoistettua paikkatietoa hyödynnetään reaaliaikaisesti (Laine ym. 2013). Tämä voi useissa tapauksissa vaatia laatuindikaattoreihin tukeutumista (Antoniou & Skopeliti 2015). Niitä käyttävät sovellukset edellyttävät kuitenkin laatuvalidoinnin pitkälle kehittyntä automaatiota. Tätä tukemaan tarvitaan lisätutkimuksia laatuun vaikuttavien data-, käyttäjä- ja ympäristötekijöiden soveltuvuudesta laatuindikaattoreiksi suomalaisessa yhteydessä. Indikaattoritutkimus puolestaan edellyttää ensivaiheessaan myös kansallisesti kattavaa, kaupunki- ja maaseutu-ympäristöissä toteutettavaa laatutason suorituskykymittausta.

Menetelmätasolla OpenStreetMapin jatkotutkimustarpeita voidaan tunnistaa alueelliseen mittaukseen ja spatiaaliseen autokorrelaatiotestaukseen liittyen, tarkemmin muuntuvan alueyksikön ja ryhmittelyn, sekä naapuruston määrittelyn ongelmiin. Niiden suhde joukkoistetun paikkatietotuotannon tilasidonnaisuuteen tarjoaa mielenkiintoisia, kirjallisuudessa etenkin suomalaisessa yhteydessä koskemattomia kohteita OpenStreetMapin jatkotutkimukselle. Lisäksi mittakaavatasojen vaikutuksen selvittäminen laatutuloksiin on perusteltu.

Käytettävyyssarvion tulokset antavat aiheen tutkia laatuparametrien merkitystä eri käyttöyhteyksissä. Suorien tapauskohtaisten käytettävyyssarvioiden lisäksi käytettävyystudkimus ja alueellinen tarkastelu voivat tarjota vapausasteita tyypeiltään ja teemoiltaan moninaisen joukkoistetun paikkatiedon soveltavaan hyödyntämiseen (See ym. 2017; McGullagh & Jackson 2013), jotta laadultaan vaihtelevakin joukkoistettu paikkatieto voidaan ottaa käyttöön. Laatuarvion tulokset osoittavat laatutason voivan vaihdella osatekijöiden välillä suuresti. Aineisto voi kokonaislaadultaan alittaa laatuvaatimuksen, mutta olla joiltain osin laatuparametreiltään, tai alueellisesti, korkealaatuinen ja soveltua näin riittävän hyvin käyttötarkoitukseen, jossa nämä tekijät ovat olennaisia.

The Economist julisti datan uudeksi öljyksi (The worlds... 2017). Joukkoistetun karttapaikkatiedon hyödyntämissyrkimykset ovat usein koskeneet pääasiassa aktiivisesti kerättyä ja peruspaikkatietoja muistuttavaa joukkoistettua paikkatietoa (See ym. 2017). Arjen digitalisaation synnyttämät datalähteet ja erityisesti automaation jatkuva kehitys antavat kuitenkin luvan odottaa passiivisesti kerättyjen, ja peruspaikkatiedoista poikkeavien tietolähteiden avautuvan lisääntyvässä määrin datatalouden käytettäväksi. Laatututkimus on näiden uusien lähteiden luotettavan hyödyntämisen keskeinen mahdollistaja.

Paikkatietoinfrastruktuuriin integroitavuuden alueella tehtävä yhteentoimivuustutkimus, josta vertailussa olleiden tietojärjestelmien saatavuus-, ylläpito- ja tiedonvälityskysymysten lisäksi huomattava osa on käytännössä paikkatietomallien täydellisyyden sekä niiden käsitteellisen, arvojoukko- ja formaattieheyden laatututkimusta, luo lähtökohtia joukkoistettujen aineistojen laajemmalle hyödynnettävyydelle. Usein tekniset kysymykset ovat helpoimmin ratkaistavia. Haastavampaa voi olla yhteensovittaa joukkoistetun ja auktoritatiivisen tuotantotavan toimintaympäristöön liittyvät ja organisatoriset erot. Jälkimmäiset ovat joukkoistetun paikkatiedon luonteenomaisia piirteitä ja sen heterogeenisen laadun pääasiallinen aiheuttaja. Niiden ymmärtäminen ja hallinta edellyttävät syvällistä tuntemusta paitsi teknologiasta, myös joukkoistetun paikkatietotuotannon tilasidonnaisesta toimintakulttuurista ja rakenteista.

Ohjelmistokehitysmallien etuja vertaillessaan Raymond (1999) näki joukkoistuksessa basaarin, siinä missä keskitetty toimintatapa muistutti katedraalia. Vaikka katedraalit rakennettiin kaupunkilaisten voimin, olivat rakentajat usein ammattilaisia ja aikaa uhmaavaa, monimutkaista ponnistusta ohjattiin keskitetysti. Tämä muistuttaa Cinnamonin (2015) näkemyksestä joukkoistetun toimintatavan tarvelähtöisestä, mutta hallitusta sovittamisesta auktoritatiiviseen kehukseen. Tällöin vastaus McCullaghin ja Jacksonin (2013) kysymykseen siitä, lasketaanko amatöörit temppeliin, saattaa kuulua: Tulkoot sellaisina kuin ovat.

Kiitokset

Kiitän Pro Gradu -työni ohjaajaa Niina Käyhköä. Lisäksi kiitän Johanna Nylundia.

Lähteet

- 421/2009 Laki paikkatietoinfrastruktuurista (2009) Finlex. Viitattu 5.9.2023.
<<https://www.finlex.fi/fi/laki/alkup/2009/20090421>>
- 725/2009 Valtioneuvoston asetus paikkatietoinfrastruktuurista (2009) Finlex. Viitattu 5.9.2023. <<https://www.finlex.fi/fi/laki/alkup/2009/20090725>>
- 906/2019 Laki julkisen hallinnon tiedonhallinnasta (2019) Finlex. Viitattu 5.9.2023.
<<https://www.finlex.fi/fi/laki/alkup/2019/20190906>>
- Aalders, H. G. J. L., (2002) The Registration of Quality in a GIS. Teoksessa Shi, W., Fisher, P.F. & Goodchild, M.F. (toim.). *Spatial Data Quality*, 186–299. Taylor and Francis, London.
- Abdolmajidi, E., Mansourian, A., Will, J., & Harrie, L. (2015) Matching authority and VGI road networks using an extended node-based matching algorithm. *Geo-spatial Information Science*, 18(2–3) 65–80. <https://doi.org/10.1080/10095020.2015.1071065>
- Albuquerque, J. P., Herfort, B. & Eckle, M. (2016b) The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing* 8(10) 1–22. <https://doi.org/10.3390/rs8100859>
- Albuquerque, J. P., Herfort, B., Eckle, M., & Zipf, A. (2016a). Crowdsourcing geographic information for disaster management and improving urban resilience: An overview of recent developments and lessons learned. Teoksessa Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F. & Purves, R. (toim.), *European handbook of crowdsourced geographic information*, 309–321. Ubiquity Press, London. <http://dx.doi.org/10.5334/bax.w>
- AlphaBeta (2017) *The economic impact of geospatial services: how consumers, businesses and society benefit from location-based information*. AlphaBeta Advisors Pty Ltd. <https://accesspartnership.com/wp-content/uploads/2023/03/GeoSpatial-Report_Sept-2017.pdf>
- Anderson, C. (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired* 16(7) 106–129. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Anderson, J., Sarkar, D., Palen, L. (2019) Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information* 8(5) 232. <https://doi.org/10.3390/ijgi8050232>

- Annala, S. (2020) Uusi laki muuttaa julkisen hallinnon tiedonhallintaa – miten käy arkkitehtuurin? Suomidigi. Viitattu 5.9.2023.
<<https://www.suomidigi.fi/artikkelit/uusi-laki-muuttaa-julkisen-hallinnon-tiedonhallintaa-miten-kay-arkkitehtuurin>>
- Antoniou, V. & Schlieder, C. (2018) Addressing uneven participation patterns in VGI through gamification mechanisms. Teoksessa Ahlqvist, O. & Schlieder, C. (toim.) *Geogames and Geoplay: Game-based Approaches to the Analysis of Geo-Information*, 91–110. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-22774-0_5
- Antoniou, V. & Skopeliti, A. (2015) Measures and indicators of VGI quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5 345–351. <https://doi.org/10.5194/isprsannals-II-3-W5-345-2015>
- Arsanjani, J. J., Mooney, P., Helbich, M. & Zipf, A. (2015) An exploration of future patterns of the contributions to OpenStreetMap and development of a contribution index. *Transactions in GIS* 19(6) 896–914. <https://doi.org/10.1111/tgis.12139>
- Ash, J., Kitchin, R., & Leszczynski, A. (2018) Digital turn, digital geographies? *Progress in Human Geography* 42(1) 25–43. <https://doi.org/10.1177/0309132516664800>
- Azouzi, M. (2000) Introducing the concept of reliability in spatial data. Teoksessa Lowell, K. & Jaton, A. (toim.) *Spatial Accuracy Assessment* 139–144. CRC Press, Boca Raton.
- Barrington-Leigh, C., & Millard-Ball, A. (2017) The world's user-generated road map is more than 80 % complete. *PloS One* 14(10) e0180698.
<https://doi.org/10.1371/journal.pone.0224742>
- Barron, C., Neis, P. & Zipf, A. (2014) A Comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS* 18(6) 877–895.
<https://doi.org/10.1111/tgis.12073>
- Basiri, A., Jackson, M., Amirian, P., Pourabdollah, A., Sester, M., Winstanley, A., Zhang, L. (2016) Quality assessment of OpenStreetMap data using trajectory mining. *Geo-spatial Information Science* 19(1) 56–68.
<https://doi.org/10.1080/10095020.2016.1151213>
- Bégin, D., Devillers, R. & Roche, S. (2013) Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours. *Proceedings of the 8th international symposium on spatial data quality ISSDQ* 149–154. Viitattu 5.9.2023.
<<http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-2-W1/149/2013/isprsarchives-XL-2-W1-149-2013.pdf>>

- Bicycle Tags Map (2022) OpenStreetMap Wiki. Viitattu 2.7.2023
<https://wiki.openstreetmap.org/wiki/Bicycle_tags_map>
- Blakemore, M. 1984. Generalization and error in spatial databases. *Cartographica: The International Journal for Geographic Information and Geovisualization* 21(2–3) 131–39.
- Bordogna, G. & Carrara, P., Kliment, T., Frigerio, L. & Sterlacchini, S. (2015) Spatial data infrastructures empowered by interoperable volunteered geographic information. *Plurimondi* 2015(16).
- Brassel, K., Bucher, F., Stephan, E.-M., & Vckovski, A. (1995) Completeness. Teoksessa Guptill, S., C. & Morrison, J., L. (Toim.) *Elements of spatial data quality*, 81–108. Pergamon, Oxford.
- Bryant, A. & Raja, U. (2014) In the realm of big data. *First Monday* 19.
<https://doi.org/10.5210/fm.v19i2.4991>
- Budhathoki, N.R., Haythornthwaite, C. (2013) Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist* 57(5) 548–575. <https://doi.org/10.1177/0002764212469364>
- Capineri, C. (2016) The nature of volunteered geographic information. Teoksessa Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F. & Purves, R. (toim.) *European Handbook of Crowdsourced Geographic Information*, 15–33. Ubiquity Press, London. <http://dx.doi.org/10.5334/bax.b>
- Castells, M. (1996) *The Rise of the Network Society, The Information Age: Economy, Society and Culture*. Oxford, Blackwell.
- Chaturvedi, K., Matheus, A., Nguyen, S. H. & Kolbe, T. H. (2019) Securing Spatial Data Infrastructures for Distributed Smart City applications and services. *Future Generation Computer Systems* 101 723–736. <https://doi.org/10.1016/j.future.2019.07.002>
- Chrisman, N. R. (1984) The role of quality information in the long-term functioning of a geographic information system. *Cartographica: The International Journal for Geographic Information and Geovisualization* 21(2) 79–87. Viitattu 5.9.2023.
<https://www.researchgate.net/publication/238350661_The_role_of_quality_information_in_the_long_term_functioning_of_a_Geographic_Information_System>
- Ciepluch, B., Jacob, R., Mooney, P. & Winstanley, A. C. (2010) Comparison of the accuracy of openstreetmap for ireland with google maps and bing maps. *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources*

- and Environmental Sciences 20-23rd July 2010*, Leicester, England, 337–340.
<http://eprints.maynoothuniversity.ie/2476/>
- Cinnamon, J. (2015) Deconstructing the binaries of spatial data production: Towards hybridity. *Canadian Geographies* 59(1) 35–51. <https://doi.org/10.1111/cag.12119>
- Collins, F.C. & Smith J. L. (1994) Taxonomy for error in GIS. *Proceedings of the International Symposium on the Spatial Accuracy of Natural Resource Data Bases*, 1–7. 16–20 May 1994. Williamsburg, USA, ASPRS.
- Couclelis, H. (2003) The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS* 7(2) 165–175. <https://doi.org/10.1111/1467-9671.00138>
- Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an infrastructure for spatial information in the European Community (INSPIRE) (2007) European Parliament and of the Council of the European Union. Viitattu 5.9.2023. <<https://inspire.ec.europa.eu/documents/directive-20072ec-european-parliament-and-council-14-march-2007-establishing>>
- Creating a topology (2023) ESRI, ArcMap 10.8 documentation. Viitattu 5.9.2023. <<https://desktop.arcgis.com/en/arcmap/latest/manage-data/topologies/creating-a-topology.htm>>
- Crisuolo, L., Carrara, P., Bordogna, G., Pepe, M., Zucca, F., Seppi, R., Oggioni, A. & Rampini, A. (2016) Handling quality in crowdsourced geographic information. Teoksessa Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F. & Purves, R. (toim.) *European Handbook of Crowdsourced Geographic Information*, 57–74. Ubiquity Press, London.
<http://dx.doi.org/10.5334/bax.e>
- David, B. & Fasquel, P. (1997) Qualité d'une base de données géographique: concepts et terminologie. *Bulletin d'information de Institut géographique national* 67(2).
- Devilleers, R. & Jeansoulin, R. (2006; toim.). *Fundamentals of Spatial Data Quality*. Antony Rowe Ltd, Chippenham.
- Digiroad-aineiston ylläpito (2023) Väylävirasto. Viitattu 5.9.2023. <<https://vayla.fi/vaylista/aineistot/digiroad/yllapito>>
- Digiroad-tietotuotteen laadunhallintajärjestelmä (2012) Väylävirasto. Viitattu 16.4.2022. <<https://docplayer.fi/16184670-Laatumalli-liikennevirasto.html>>
- Download OpenStreetMap data for this region: Finland (2022) Geofabrik GmbH. Viitattu 25.3.2022. <<http://download.geofabrik.de/europe/finland.html>>

- Echeverry-Galvis, M., Ramírez, P. & Amaya-Espinel, J. (2023) Long-term Christmas Bird Counts describe neotropical urban bird diversity. *PloS one* 18(2) e0272754. <https://doi.org/10.1371/journal.pone.0272754>
- Editing Standards and Conventions (2023) OpenStreetMap Wiki. Viitattu 5.9.2023. <https://wiki.openstreetmap.org/wiki/Editing_Standards_and_Conventions>
- El-Ashmawy, K. L. A. (2016) Testing the positional accuracy of OpenStreetMap data for mapping applications. *Geodesy and Cartography* 42(1) 25–30. <https://doi.org/10.1080/10095020.2016.1151213>
- Elias, E. N. N., Fernandes, V. O., Alixandrini Jr., M. J., Schmidt, M. M. R. (2020) The quality of OpenStreetMap in a large metropolis in northeast Brazil: Preliminary assessment of geospatial data for road axes. *Bulletin of Geodetic Sciences* 26(3) 1. <https://doi.org/10.1590/s1982-21702020000300012>
- ERT position on building the european data economy (2017) European Commission. Viitattu 5.9.2023. <<https://ert.eu/pdf-information/2017-05-2017-05-12-building-the-european-data-economy-position-paper-pdf/>>
- Evans, B. J. (1997) Dynamic display of spatial data-reliability: does it benefit the map user. *Computers & Geosciences* 23(4) 409–422. [https://doi.org/10.1016/S0098-3004\(97\)00011-3](https://doi.org/10.1016/S0098-3004(97)00011-3)
- Export topology checker results as table/shapefile in QGIS (2020) StackExchange. Viitattu 5.9.2023. <<https://gis.stackexchange.com/questions/372184/export-topology-checker-results-as-table-shapefile-in-qgis>>
- Fagerholm, N., Käyhkö, N., Ndumbaro, F. & Khamis, M. (2012) Community stakeholders' knowledge in landscape assessments – Mapping indicators for landscape services. *Ecological Indicators* 18. 421–433. <https://doi.org/10.1016/j.ecolind.2011.12.004>
- Ferreira, S., L., F., C. & Cintra, J. P. 1999. Quantificação de Discrepâncias entre Feições Lineares por Retângulos Equivalentes. *Revista Brasileira de Cartografia*, 51 1–8.
- Find topology errors with the error inspector (2023) ESRI, ArcMap 10.8 documentation. Viitattu 5.9.2023. <<https://desktop.arcgis.com/en/arcmap/latest/manage-data/editing-topology/finding-topology-errors-with-the-error-inspector.htm>>
- Fisher, P. F. (1999) Models of uncertainty in spatial data. Teoksessa Longley P., Goodchild, M., Maguire D. & Rhind D. (toim.) *Geographical Information Systems: Principles, Techniques, Management and Applications*, 1 191–205. John Wiley & Sons, New York.

- Fisher, P., Comber, A. & Wadsworth, R. (2006) Approaches to uncertainty in spatial data. Teoksessa Devillers, R. & Jeansoulin, R. (toim.) *Fundamentals of spatial data quality*, 43–56. <https://doi.org/10.1002/9780470612156.ch3>
- Flanagin, A.J. & Metzger, M.J. (2008) The credibility of volunteered geographic information. *GeoJournal* 72(3) 137–148. <https://doi.org/10.1007/s10708-008-9188-y>
- Fonte, C. C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J. J., Bayas, C. L., See, L. M. & Vatsava R. (2017) Assessing VGI data quality. Teoksessa, Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A. M., Fonte, C. C. & Antoniou, V. (toim.) *Mapping and the Citizen Sensor*, 137–163. Ubiquity Press, London. <http://doi.org/10.5334/bbf.g>
- Forghani, M. & Delavar, M. R. (2014) A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information* 3(2) 750–763. <https://doi.org/10.3390/ijgi3020750>
- França, L., Ferreira, S., Luiz, F. (2018) Comparison between the double buffer method and the equivalent rectangle method for the quantification of discrepancies between linear features. *Boletim de Ciencias Geodesicas* 24(3) 300–317. <https://doi.org/10.1590/s1982-21702018000300020>
- Frank, A. U. (2007) Data quality ontology: an ontology for imperfect knowledge. *Proceedings of the 8th Conference on Spatial Information Theory*, 406–420. COSIT. https://doi.org/10.1007/978-3-540-74788-8_25
- Fritz, S., See, L. & Brovelli, M. (2017) Motivating and sustaining participation in VGI. Teoksessa Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C. & Antoniou, V. (toim.) *Mapping and the Citizen Sensor*, 93–117. Ubiquity Press. London. <https://doi.org/10.5334/bbf.e>
- Gardner, Z., Mooney, P., Dowthwaite, L. & Foody, G. (2018) Gender differences in OpenStreetMap contributor activity, editing and tagging behaviour. GISRUK conference 17th–20th April 2018. Viitattu 5.9.2023. https://www.researchgate.net/publication/327445924_Gender_differences_in_OpenStreetMap_contributor_activity_editing_and_tagging_behaviour
- Geoinformatiikan sanasto 2018. Maanmittauslaitos. Viitattu 5.9.2023. <https://sanastokeskus.fi/tiedostot/pdf/GeoinformatiikanSanasto.pdf?file=pdf/GeoinformatiikanSanasto.pdf>
- Gibson, J. J. (1977) The theory of affordances. Teoksessa Shaw, R. & Bransford, J. (toim.) *Perceiving, acting, and knowing* 127–143. Lawrence Erlbaum Associates, Hillsdale.

- Girres, J.-F. & Touya, G. (2010) Quality assessment of the french OpenStreetMap dataset. *Transactions in GIS* 14(4) 435–459. <https://doi.org/10.1111/j.1467-9671.2010.01203.x>
- Goodchild, M. & Li, L. (2012) Assuring the quality of volunteered geographic information. *Spatial Statistics* 1 110–120. <https://doi.org/10.1016/j.spasta.2012.03.002>
- Goodchild, M. (2009) NeoGeography and the nature of geographic expertise. *Journal of Location Based Services* 3(2) 82–96. <https://doi.org/10.1080/17489720902950374>
- Goodchild, M. F. & Hunter, G. J. (1997) A simple positional accuracy measure for linear features. *ISPRS International Journal of Geographical Information Science* 11(3) 299–306. <https://doi.org/10.1080/136588197242419>
- Goodchild, M. F. (2007) Citizens as Sensors: the world of volunteered geography. *GeoJournal* 69(4) 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Hahmann, S. & Burghardt, D. (2013) How much information is geospatially referenced? Networks and cognition. *ISPRS International Journal of Geographical Information Science* 27(6) 1171–1189. <https://doi.org/10.1080/13658816.2012.743664>
- Haklay, M. & Weber, D. (2008) OpenStreetMap: user-generated street maps. *IEEE Pervasive Computing*, 7(4) 12–18. <https://doi.org/10.1109/MPRV.2008.80>
- Haklay, M. (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning. B: Planning & Design* 37(4), 682–703. <https://doi.org/10.1068/b350>
- Haklay, M. (2013) Neogeography and the Delusion of Democratisation. *Environment and Planning A: Economy and Space* 45(1). <https://doi.org/10.1068/a45184>
- Haklay, M. E. (2016) Why is participation inequality important? Teoksessa Capineri, C., Haklay, M., Huang, H., Antoniou, V., Kettunen, J., Ostermann, F. & Purves, R., (toim.) *European Handbook of Crowdsourced Geographic Information*, 35–44. Ubiquity Press, London. <https://doi.org/10.5334/bax.c>
- Haklay, M., Antoniou, V., Basiouka, S., Soden, R., Deparday, V., Ryan, M. & Mooney, P. (2018) *Identifying Success Factors in Crowdsourced Geographic Information Use in Government*. Policy brief, World Bank. Viitattu 5.9.2023. <https://www.gfdr.org/en/publication/identifying-success-factors-crowdsourced-geographic-information-use-government-0>
- Harvey, F., (2013) To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information, Teoksessa Sui, D., Elwood, S., Goodchild, M. (toim.) *Crowdsourcing Geographic Knowledge*, 31–42. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4587-2_3

- Hashemi, P., Abbaspour, R. (2015) Assessment of logical consistency in OpenStreetMap based on the spatial similarity concept. Teoksessa Jokar Arsanjani, J., Zipf, A., Mooney, P., Helbich, M. (toim.) *OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography*, 19–36. Springer, Cham.
https://doi.org/10.1007/978-3-319-14280-7_2
- Herfort, B., Lautenbach, S., Albuquerque, J., Anderson, J. & Zipf, A. (2021) The evolution of humanitarian mapping within the OpenStreetMap community. *Scientific Reports* 11(1) 3037. <https://doi.org/10.1038/s41598-021-82404-z>
- Hochmair, H., Zielstra, D., & Neis, P. (2014) Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States. *Transactions in GIS* 19(1) 63–81. <https://doi.org/19.10.1111/tgis.12081>
- How Spatial Autocorrelation (Global Moran's I) works (2023) ArcGIS Pro dokumentaatio. Viitattu 17.9.2023. <<https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>>
- Hunter, G. J. & Beard, M. K. (1992) Understanding error in spatial databases. *The Australian Surveyor* 37(2) 108–119. <https://doi.org/10.1080/00050326.1992.10438784>
- Import (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.
 <<https://wiki.openstreetmap.org/wiki/Import>>
- INSPIRE Data specification on transport networks – technical guidelines (2014) European Commission. Viitattu 5.9.2023. <<https://inspire.ec.europa.eu/id/document/tg/tn>>
- INSPIRE-velvoitetulle (2023) Maanmittauslaitos. Viitattu 5.9.2023.
 <<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/paikkatietojen-yhteentoimivuus/inspire/inspire-velvoitetulle>>
- ISO 19157 (2013) *Geographic information — data quality*. International Organization for Standardization. Viitattu 26.9.2021. <<https://www.iso.org/standard/32575.html>>
- Jääskeläinen (2018) *Kävelyn ja pyöräilyn edistämishjelma*. Liikenne- ja viestintäministeriön julkaisuja 5/2018. <http://urn.fi/URN:ISBN:978-952-243-549-1>
- Jacobs, K. T. (2018) Quality assessment of volunteered geographic information: an investigation into the Ottawa-Gatineau OpenStreetMap database. Master's Thesis, Carleton University. <https://doi.org/10.22215/etd/2018-13302>
- Jacobsson, A. (2015) Laadun tilannekatsaus, Maanmittauslaitos. Viitattu 5.9.2023.
 <<https://docplayer.fi/45993697-Laadun-tilannekatsaus-paikkatietoverkosto-laadun-karkiteema-kick-off-antti-jakobsson-maanmittauslaitos-pj.html>>

- JHS 160 Paikkatiedon laadunhallinta (2009). Julkisen hallinnon tietohallinnon neuvottelukunta. Viitattu 26.9.2021. <<https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset/jhs-160-paikkatiedon-laadunhallinta-vanhentunut>>
- JHS-suositukset (lakkautetut) (2023) Suomidigi. Viitattu 5.9.2023. <<https://www.suomidigi.fi/ohjeet-ja-tuki/jhs-suositukset-lakkautetut>>
- Jokar Arsanjani, J., Mooney, P., Zipf, A. & Schauss, A. (2015) Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. Teoksessa Jokar Arsanjani, J., Zipf, A., Mooney, P. & Helbich, M. (toim.) *OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography*, 37–58. Springer, Cham. https://doi.org/10.1007/978-3-319-14280-7_3
- Jokar Arsanjani, J., Zipf, A., Mooney, P. & Helbich, M. (2015) An introduction to OpenStreetMap in geographic information science: experiences, research, and applications, Teoksessa Jokar Arsanjani, J., Zipf, A., Mooney, P. & Helbich, M. (toim.), *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography*, 1–18. Springer International Publishing, Cham.
- Kansallinen paikkatietostrategia 2022–2025 (2021) Geoforum. Viitattu 5.9.2023. <<https://geoforum.fi/wp-content/uploads/2022/01/Kansallinen-paikkatietostrategia-2022%E2%80%932025.pdf>>
- Karlsson, K.-P. (2018) *Paikkatietopoliittinen selonteko*. Maa- ja metsätalousministeriön julkaisuja 4a/2018. <http://urn.fi/URN:ISBN:978-952-453-978-4>
- Kitchin, R. & McArdle, G. (2016) What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society* 3(1). <https://doi.org/10.1177/2053951716631130>
- Kitchin, R. (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1). <https://doi.org/10.1177/2053951714528481>
- KMTK laadunhallintajärjestelmä (2020) Maanmittauslaitos. Viitattu 5.9.2023. <<https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2020/03/KMTK%20laadunhallintajarjestelma.pdf>>
- Kotsev, A., Minghini, M., Tomas, R., Cetl, V. & Lutz, M. (2020) From spatial data infrastructures to data spaces—a technological perspective on the evolution of european SDIs. *ISPRS International Journal of Geo-Information* 9(3) 176. <https://doi.org/10.3390/ijgi9030176>

- Koukoletsos, T., Haklay, M., & Ellul, C. (2012) Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS* 16(4) 477–498. <https://doi.org/10.1111/j.1467-9671.2012.01304.x>
- Kukulka-Koziel, A., Tomasz, S., Cegielska, K., & Szylar, M. (2018) Methodology of evaluation and correction of geometric data topology in QGIS software. *Acta Scientiarum Polonorum. Formatio Circumiectus*. 17(1). <https://doi.org/10.15576/ASP.FC/2018.17.1.137>
- Laine, T., Metsäranta, H. & Saarinen, H. (2013) *Liikennetiedon arvo. Esiselvitys*. Liikenneviraston tutkimuksia ja selvityksiä 48/2013. Viitattu 5.9.2023. https://www.doria.fi/bitstream/handle/10024/121175/lts_2013-48_978-952-255-365-2.pdf?sequence=1&isAllowed=y
- Laitinen, L., Mattila M., Metsäpuro, P. & L. Nykänen. (2015) *Pyöräväylien tiedot ja laatutaso. Esiselvitys valtakunnallisesta pyöräväylien tiedonhallintamallista*. Liikenneviraston tutkimuksia ja selvityksiä 24/2015. Viitattu 5.9.2023. https://www.doria.fi/bitstream/handle/10024/121360/lts_2015-24_978-952-317-091-9.pdf?sequence=1&isAllowed=y
- Lehtonen, K., Piippo, H., Marttila, J., Ryynänen, M., Liimatainen, A., Aaltonen, I. & Massinen S. (2021) Tien poikkileikkauksen suunnittelu. *Väyläviraston ohjeita* 16/2021. https://ava.vaylapilvi.fi/ava/Julkaisut/Vaylavirasto/vo_2021-16_tien_poikkileikkauksen_web.pdf
- Lévy, P. (1999). *Collective intelligence: mankind's emerging world in cyberspace*. Perseus Publishing, Cambridge, MA, United States.
- Lin, Y.-W. (2011) A qualitative enquiry into OpenStreetMap making. *New Review of Hypermedia and Multimedia*, 17(1) 53–71. <https://doi.org/10.1080/13614568.2011.552647>
- Liu, C., Xiong, L., Hu, X., & Shan, J. (2015) A progressive buffering method for road map update using OpenStreetMap data. *ISPRS International Journal of Geo-Information* 4(3) 1246–1264. <https://doi.org/10.3390/ijgi4031246>
- Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2015) *Geographic Information Science and Systems*. 4. laitos. Wiley & Sons. Kendallville.
- Ludwig, I., Voss, A., & Krause-Traudes, M. (2011) A Comparison of the Street Networks of Navteq and OSM in Germany. Teoksessa Geertman, S., Reinhardt, W., Toppen, F. (toim.) *Advancing Geoinformation Science for a Changing World I* 65–84. Springer, Berlin. <https://doi.org/10.1007/978-3-642-19789-5>

- Ma, D., Sandberg, M. & Bin, J. (2015) Characterizing the heterogeneity of the OpenStreetMap data and community. *ISPRS International Journal of Geo-Information* 4(2) 535–550. <https://doi.org/10.3390/ijgi4020535>
- Maastotietojen laatumalli (1995) Maanmittauslaitos. Viitattu 5.9.2023.
<https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/old/Maastotietojen_laatumalli.pdf>
- Maastotietokanta, laatumallitiedot (2023) Maanmittauslaitos. Viitattu 5.9.2023
<<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/asiantuntevalle-kayttajalle/tuotekuvaukset/maastotietokanta-0>>
- Major OpenStreetMap Consumers (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.
<https://wiki.openstreetmap.org/wiki/Major_OpenStreetMap_consumers>
- Map features - OpenStreetMap Wiki. (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.
<https://wiki.openstreetmap.org/wiki/Map_features>
- Marjamäki, V.-M. (2014) *Tiedon hyödyntämisen ekosysteemi*. Liikenneviraston tutkimuksia ja selvityksiä 56/2014. <http://www.doria.fi/handle/10024/121643>
- Martin, S., Gautier, P., Turki, S. & Kotsev, A. (2021) *Establishment of sustainable data ecosystems. Recommendations for the evolution of spatial data infrastructures*. Publications Office, European Commission, Luxembourg.
<https://doi.org/10.2760/04462>
- McCullagh, M. & Jackson, M. (2013) Crowdsourced mapping - letting amateurs into the temple? *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. 40. XL-1/W1, 399–432.
<https://doi.org/10.5194/isprsarchives-XL-1-W1-399-2013>
- Min, D., Zhilin, L. & Xiaoyong, C. (2007) Extended Hausdorff distance for spatial objects in GIS. *ISPRS International Journal of Geographical Information Science* 21(4) 459–475. <https://doi.org/10.1080/13658810601073315>
- Minghini, M., Kotsev, A. & Lutz, M. (2019) Comparing INSPIRE and OpenStreetMap data: how to make the most out of the two worlds. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W14 167–174. <https://doi.org/10.5194/isprs-archives-XLII-4-W14-167-2019>
- Mocnik, F.-B., Mobasheri, A., Griesbaum, L., Eckle, M., Jacobs, C., & Klonner, C. (2018) A grounding-based ontology of data quality measures. *Journal of Spatial Information Science* (16) 1–25. <http://dx.doi.org/10.5311/JOSIS.2018.16.360>

- Mõisja, K., Uuemaa, E. & Oja, T. (2018) The Implications of Field Worker Characteristics and Landscape Heterogeneity for Classification Correctness and the Completeness of Topographical Mapping. *ISPRS International Journal of Geo-Information* 7(6) 205. <https://doi.org/10.3390/ijgi7060205>
- Mooney, P. & Minghini, M. (2017) A review of OpenStreetMap data. Teoksessa Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C. & Antoniou, V. (toim.) *Mapping and the Citizen Sensor*, 37–59. Ubiquity Press, London. <https://doi.org/10.5334/bbf.c>
- Mooney, P., (2015) Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets, Teoksessa Arsanjani, J. J., Zipf, A., Mooney, P., & Helbich, M. (toim.) *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography*, 319–324. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-14280-7_3
- Moradi, M., Roche, S., & AbolfazlMostafavi, M. (2021) Exploring five indicators for the quality of OpenStreetMap road networks: a case study of Québec, Canada. *Geomatica* 75(4) 178–208. <https://doi.org/10.1139/geomat-2021-0012>
- Mozas-Calvache, A. T. (2021) Positional quality assessment based on linear elements. *Revista cartográfica*, (103) 11–31. <http://orcid.org/0000-0001-5847-4338>
- Napolitano, M. & Mooney, P., (2012) MVP OSM: a tool to identify areas of high quality contributor activity in OpenStreetMap. *The Bulletin of the Society of Cartographers* 45(1) 10–18.
- Neis, P. & Zipf, A. (2012) Analyzing the contributor activity of a volunteered geographic information project — the case of OpenStreetMap. *ISPRS International Journal of Geo-Information* 1(3) 146–165 <http://doi.org/10.3390/ijgi1020146>
- Neis, P. (2015) Measuring the reliability of wheelchair user route planning based on volunteered geographic information. *Transactions in GIS* 19(2) 188–201. <https://doi.org/10.1111/tgis.12087>
- Neis, P. (2023) *OSMstats - Statistics of the free wiki world map*. Viitattu 5.9.2023. <https://osmstats.neis-one.org/?item=members>
- Nielsen, J. (2006) Participation Inequality: The 90-9-1 rule for social features. Nielsen Norman Group. Viitattu 5.9.2023. <https://www.nngroup.com/articles/participation-inequality/>

Normit ja käytännöt (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.

<https://wiki.openstreetmap.org/wiki/Fi:Normit_ja_k%C3%A4yt%C3%A4nn%C3%B6t>

O'Reilly, T. (2007) What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies* 65(1) 17.

<https://ssrn.com/abstract=1008839>

Olteanu-Raimond, A.-M., Laakso, M., Antoniou, V., Fonte, C. C., Fonseca, A., Grus, M., Harding, J., Kellenberger, T., Minghini, M., Skopeliti, A., (2017) VGI in national mapping agencies: experiences and recommendations. Teoksessa Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C. & Antoniou, V. (toim.) *Mapping and the Citizen Sensor*, 299–326. Ubiquity Press.

<https://www.jstor.org/stable/j.ctv3t5qzc>

OSM XML (2023) OpenStreetMap wiki. Viitattu 5.9.2023.

<https://wiki.openstreetmap.org/wiki/OSM_XML>

Paikkatietoalan standardit ja suositukset (2023) Maanmittauslaitos. Viitattu 5.9.2023.

<<https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/paikkatietojen-yhteentoimivuus/standardit-ja-suositukset>>

Pärssinen, T. (2022) Kysymyksiä pyöräilyverkon tietomallista. Henkilökohtainen sähköpostiviesti M. Lahtiselle. 17.3.2022.

Peltonen, T. (2016) Digiroad- ja OpenStreetMap-aineistojen yhteiskäyttö joukkoliikennepysäkeissä. *Opinnäytetyö* 3/2016, Väylävirasto.

<https://www.doria.fi/handle/10024/121944>

Peruspaikkatietojen ylläpito (2023) Maanmittauslaitos. Viitattu 5.9.2023.

<<https://www.maanmittauslaitos.fi/peruspaikkatietojen-yllapito>>

Quality assurance (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.

<https://wiki.openstreetmap.org/wiki/Quality_assurance>

Quattrone, G., Mashhadi, A., Quercia, D., Smith-Clarke, C. & Capra, L. (2014) Modelling growth of urban crowd-sourced information. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* 563–572. New York, NY, USA <http://doi.org/10.1145/2556195.2556244>

Quattrone, G., Mashhadi, M. & Capra, L. (2014) Mind the map: the impact of culture and economic affluence on crowd-mapping behaviours. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* 934–944. Baltimore Maryland USA, Feb. 2014. <https://doi.org/10.1145/2531602.2531713>

- Ramm, F. (2021) *OpenStreetMap data in layered GIS format*. Geofabrik GmbH. Viitattu 24.3.2022. <<https://www.geofabrik.de/data/geofabrik-osm-gis-standard-0.7.pdf>>
- Raymond, E. S. (1999) *The cathedral and the bazaar: musings on Linux and open source by an accidental revolutionary*. O'Reilly, Cambridge.
- Rifkin, J. (2014) *The zero marginal cost society: the internet of things, the collaborative commons, and the eclipse of capitalism*. St. Martin's Press, New York.
- Roper, B. (2023) *Investigating the properties of OpenStreetMap provenance graphs*. PhD Thesis. University of Southampton.
- Santos, A., Medeiros, N., Santos, G. & Rodrigues, D. (2015) Controle de qualidade posicional em dados espaciais utilizando feicoes lineares. *Boletim de Ciências Geodésicas* 21(2) 233–250. <http://dx.doi.org/10.1590/S1982-21702015000200013>
- Sarretta, A., Napolitano, M., & Minghini, M. (2023) OpenStreetMap as an input source for producing governmental datasets: The case of the Italian military geographic institute. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-4/W7-2023, 193–200. <https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-193-2023>
- See, L., Estima, J., Pödör, A., Arsanjani, J. J., Bayas, J.-C. L. & Vatsava, R. (2017) Sources of VGI for mapping. Teoksessa Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C. C. & Antoniou, V. (toim.) *Mapping and the Citizen Sensor*, 13–35. Ubiquity Press, London. <https://doi.org/10.5334/bbf.b>
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H. Y., Milčinski, G., Nikšič, M., Painho, M., Pödör, A., Olteanu-Raimond, A.-M. & Rutzinger, M. (2016) Crowdsourcing, citizen science or Volunteered Geographic Information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information* 5(5) 55. <https://doi.org/10.3390/ijgi5050055>
- Sehra, S. S., Singh, J., Rai, H.S., Anand, S.S. (2019) Extending processing toolbox for assessing the logical consistency of OpenStreetMap data. *Transactions in GIS* 24(1) 44–71. <https://doi.org/10.1111/tgis.12587>
- Senaratne, H., Mobasher, A., Ali, A. L., Capineri, C., & Haklay, M. (2017) A Review of volunteered geographic information quality assessment Methods. *ISPRS International Journal of Geographical Information Science* 31(1) 139–167. <https://doi.org/10.1080/13658816.2016.1189556>

- Seo, S. & O'Hara, C., G. (2009) Quality assessment of linear data. *ISPRS International Journal of Geographical Information Science* 23(12) 1503–1525.
<https://doi.org/10.1080/13658810802231456>
- Servigne, S., Ubeda, T., Puricelli, A., & Laurini, R. (2000) A methodology for spatial consistency improvement of geographic databases. *GeoInformatica*, 4(1) 7–34.
<https://doi.org/10.1023/A:1009824308542>
- Shapefiles (2022) Geofabrik GmbH. Viitattu 24.3.2022.
<https://www.geofabrik.de/data/shapefiles.html>
- Shi, W. (2010) *Principles of modeling uncertainties in spatial data and spatial analyses*. CRC Press, Taylor & Francis Group, Boca Raton.
- Sieber, R. (2006) Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers*, 96(3) 491–507.
<https://doi.org/10.1111/j.1467-8306.2006.00702.x>
- Spielman, S. E. (2014) Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information. *Cartography and Geographic Information Science* 41(2) 115—124. <https://doi.org/10.1080/15230406.2013.874200>
- Stephens, M. (2013) Gender and the GeoWeb: divisions in the production of user-generated cartographic information. *GeoJournal* 78(6) 981–996. <https://doi.org/10.1007/s10708-013-9492-z>
- Sui, D. & Cinnamon, J. (2017) Volunteered geographic information. Teoksessa Richardson, D., Castree, N., Goodchild, M. F., Kobayashi, A. L., Liu, W., Marston, R. A. (toim.) *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1–13. John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118786352.wbieg0913>
- Sundaram, R. C. & Naghizade, E., Renata, B.-G. & Tomko, M. (2021) Can you fixme? An intrinsic classification of contributor-identified spatial data issues using topic models. *ISPRS International Journal of Geographical Information Science*. 36(1) 1–30.
<https://doi.org/10.1080/13658816.2021.1893323>
- Suomen pinta-alat kunnittain 1.1.2023 (2023) Maanmittauslaitos. Viitattu 16.7.2023.
https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/attachments/2023/02/Vuoden_2023_pinta-alatilasto_kunnat_maakunnat.pdf
- Suomi/Piirto-ohjeet (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.
<https://wiki.openstreetmap.org/wiki/Fi:Suomi/Piirto-ohjeet>
- Surowiecki, J. (2004). *The Wisdom of Crowds*. Anchor, New York.

- Taeuscher, K. (2016) Leveraging collective intelligence: how to design and manage crowd-based business models. *Business Horizons* 60(2) 237–245.
<https://doi.org/10.1016/j.bushor.2016.11.008>
- Tavana, M., Liu, W., Elmore, P., Petry, F. & Bourgeois, B. (2016) A practical taxonomy of methods and literature for managing uncertain spatial data in geographic information systems. *Measurement* 81 123. <https://doi.org/10.1016/j.measurement.2015.12.007>
- The world's most valuable resource is no longer oil, but data (2017) *The Economist*. Viitattu 15.8.2023. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclid=EAIaIQobChMI3qTZlojW_wIVAWWEYCh3zWQQREAAAYASAAEgLh8fD_BwE&gclsrc=aw.ds>
- Tietoa Digiroadista (2023) Väylävirasto. Viitattu 5.9.2023. <<https://vayla.fi/tietoa-digiroadista>>
- Tietoaineistojen yhteentoimivuus: Digiroad OpenStreetMap:illa -OpenStreetMap Digiroadilla? (2016) Väylävirasto. Viitattu 5.9.2023.<https://vayla.fi/documents/25230764/0/Loppuraportti_OSM.pdf/a88b49af-125b-43d0-81d9-d7c414bd2184>
- Tobler W. A., (1970) Computer movie simulating urban growth in the detroit region. *Economic Geography* 46(2) 234–240. <https://doi.org/10.2307/143141>
- Torkkeli, M., Rajala, P. & Aaltonen, I. (2019) Väylä4000 2023. Ajantasaisen tie- ja katuverkon keskilinja-aineiston ylläpito Suomessa (SURAVAGE-prosessi). Väyläviraston ohjeita 25/2019. Viitattu 5.9.2023.
 <https://ava.vaylapilvi.fi/ava/Julkaisut/Vaylavirasto/vo_2019-25_ajantasaisen_tie-katuverkon_web.pdf>
- Truong, Q. T., de Runz, C., & Touya, G. (2019) Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *ISPRS International Journal of Geographical Information Science*, 33(8) 1651–1682.
<https://doi.org/10.1080/13658816.2018.1556395>
- Tunnuslukuja väestöstä alueittain 1990–2022 (2022) Tilastokeskus. Viitattu 16.7.2023.
 <https://statfin.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__vaerak/statfin_vaerak_pxt_11>

- ra.px/table/tableViewLayout1/?loadedQueryId=45ba5f20-2525-44b7-96f0-0459758e1b5b&timeType=top&timeValue=1>*
- Turner, A. (2006) *Introduction to neogeography*. O'Reilly Media, Cambridge.
- Turun kaupunkiseudun yhteistyö tiivistyy yhteisen kuntayhtymän perustamisella (2022) Turun kaupunki. Viitattu 24.3.2022. <https://www.turku.fi/uutinen/2022-03-02_turun-kaupunkiseudun-yhteistyö-tiivistyy-yhteisen-kuntayhtymän-perustamisella>
- Turun pyöräilyn kehittämissuunnitelma 2029 (2018) Turun kaupunki. Viitattu 20.3.2022. <<https://ah.turku.fi/kh/2018/1126026x/Images/1663910.pdf>>
- Turun pyöräliikenteen verkkosuunnitelma (2021) Turun kaupunki. Viitattu 20.3.2022. <https://www.turku.fi/sites/default/files/atoms/files/turun_pyoraliikenteen_verkkosuunnitelma_luonnos.pdf>
- Turun seudun pyöräilykatsaus (2021). Turun kaupunki, Valonia. Viitattu 5.9.2023. <<https://valonia.fi/materiaali/turun-seudun-pyorailykatsaus-2021/>>
- Turun seudun pyöräilyverkoston lähireitit (2021b) Turun kaupunki. Viitattu 25.3.2022. <<https://www.avoindata.fi/data/fi/dataset/turun-seudun-pyorailyverkoston-lahireitit>>
- Turun seudun pyöräilyverkoston pääreitit (2021a) Turun kaupunki. Viitattu 25.3.2022. <<https://www.avoindata.fi/data/fi/dataset/turun-seudun-pyorailyverkoston-paareitit>>
- Tveite, H. (2020) BOS. QGIS Python Plugins Repository. Viitattu 17-9-2023. <<https://plugins.qgis.org/plugins/BOS/>>
- Tveite, H. & Langaas, S. (1999) An accuracy assessment method for geographical line data sets based on buffering. *ISPRS International Journal of Geographical Information Science* 13(1) 27–47. <https://doi.org/10.1080/136588199241445>
- Vaarala, R., Palo, N., Manelius, L., Keisanen, R. & Pulkkinen, M. (2020) Pyöräliikenteen suunnittelu. *Väyläviraston ohjeita* 18/2020. <https://ava.vaylapilvi.fi/ava/Julkaisut/Vaylavirasto/vo_2020-18_pyoraliikenteen_suunnittelu_web.pdf>
- Valtakunnallinen henkilöliikennetutkimus (2018) Liikenne- ja viestintävirasto. Viitattu 17.3.2022. <<https://www.traficom.fi/fi/ajankohtaista/julkaisut/valtakunnallinen-henkiloliikennetutkimus?toggle=Ty%C3%B6matkat&toggle=Liikkuminen%20eri%20el%C3%A4m%C3%A4ntilanteissa&toggle=Liikkumisen%20erot%20eri%20puolilla%20Suomea>>
- Van Niel, T. & McVicar, T. (2002) Experimental evaluation of positional accuracy estimates from a linear network using point- and line-based testing methods. *ISPRS*

- International Journal of Geographical Information Science* 16(5) 455–473.
<https://doi.org/10.1080/13658810210137022>
- Van Oort, P. (2006) Spatial data quality: from description to application. PhD Thesis, Wageningen University. Optima, Rotterdam. <https://edepot.wur.nl/38987>
- Veregin, H. & Hargitai, P. (1995) Chapter nine - an evaluation matrix for geographical data quality. Teoksessa Guptill, S., C. & Morrison J. L. (toim.) *Elements of Spatial Data Quality*, 167–188. International Cartographic Association, Pergamon.
<https://doi.org/10.1016/B978-0-08-042432-3.50016-1>
- Veregin, H. (2005) Data quality parameters. Teoksessa Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (toim.) *Geographical Information Systems: Principles, Techniques, Management and Applications*, 2. ed., 177–189, Abridged. Wiley, New York.
- WFS-latausrajapinta (2022). Väylävirasto. Viitattu 25.3.2022.
 <<https://vayla.fi/vaylista/aineistot/digiroad/aineisto/rajapinnat>>
- What is the correct way to cite an ArcGIS Online basemap? (2023) Esri Technical support. Viitattu 13.9.2023. <<https://support.esri.com/en-us/knowledge-base/faq-what-is-the-correct-way-to-cite-an-arcgis-online-ba-000012040>>
- What we do (2023) Humanitarian OpenStreetMap team. Viitattu 5.9.2023.
 <<https://www.hotosm.org/what-we-do>>
- Who Uses OpenStreetMap? Welcome Mat (2023) OpenStreetMap Wiki. Viitattu 5.9.2023.
 <https://wiki.openstreetmap.org/wiki/Major_OpenStreetMap_consumers>
- Wiggins, A. & Crowston, K. (2011) From conservation to crowdsourcing: a typology of citizen science. *Proceedings of the Annual Hawaii International Conference on System Sciences*. 1–10. <https://doi.org/10.1109/HICSS.2011.207>
- Wilson, M., & Graham, M. (2013) Neogeography and volunteered geographic information: a conversation with Michael Goodchild and Andrew Turner. *Environment and Planning A: Economy and Space* 45(1) 10–18. <https://doi.org/10.1068/a44483>
- Woolley, A., Chabris, C., Pentland, A., Hashmi, N. & Malone, T. (2010) Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330(6004) 686—688. <https://doi.org/10.1126/science.1193147>
- World Light Gray Base (2012) Esri. Viitattu 13.9.2023 <<https://doc.arcgis.com/en/data-appliance/2022/maps/world-light-gray-base.htm>>
- Wu, H., Lin, A., Clarke, K. C., Shi, W., Cardenas-Tristan, A., & Tu, Z. (2020) A comprehensive quality assessment framework for linear features from Volunteered

- Geographic Information, *ISPRS International Journal of Geographical Information Science* 35(9) 1826–1847. <https://doi.org/10.1080/13658816.2020.1832228>
- Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020) Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *ISPRS International Journal of Geographical Information Science*, 34(9) 1765–1791. <https://doi.org/10.1007/s41651-020-00067-2>
- Yang, A., Fan, H. & Jing, N. (2016) Amateur or professional: assessing the expertise of major contributors in OpenStreetMap based on contributing behaviors. *ISPRS International Journal of Geo-Information* 5(2) 21. <https://doi.org/10.3390/ijgi5020021>
- Zacharopoulou, D., Skopeliti, A., Nakos, B. (2021) Assessment and visualization of OSM consistency for european cities. *ISPRS International Journal of Geographical Information Science* 10(6) 361. <https://doi.org/10.3390/ijgi10060361>
- Zhang, G. (2021) Volunteered geographic information. Teoksessa Wilson, J., P. (toim.) *The Geographic Information Science & Technology Body of Knowledge* (1st Quarter 2021 Edition), <http://doi.org/10.22224/gistbok/2021.1.1>
- Zhang, H., Malczewski, J. (2018) Accuracy evaluation of the canadian OpenStreetMap road networks. *ISPRS International Journal of Geospatial and Environmental Research* 5(2) 1. <https://dc.uwm.edu/ijger/vol5/iss2/1>
- Zhou, Q. 2018 Exploring the relationship between density and completeness of urban building data inOpenStreetMap for quality estimation. *ISPRS International Journal of Geographical Information Science* 32(2) 257–281. <https://doi.org/10.1080/13658816.2017.1395883>
- Zielstra, D. W. G. (2014) *Assessing the quality, suitability and usability of user generated content for routing applications*. PhD Thesis, University of Florida.
- Zielstra, D., Hochmair, H. H., Neis, P., & Tonini, F. (2014) Areal delineation of home regions from contribution and editing patterns in OpenStreetMap. *ISPRS International Journal of Geo-Information* 3(4). <https://doi.org/10.3390/ijgi3041211>
- Zielstra, D., Hochmair, H.H. & Neis, P. (2013) Assessing the effect of data imports on the completeness of openstreetmap—A United States case study. *Transactions in GIS* 17(3) 315–334. <https://doi.org/10.1111/tgis.12037>
- Taeihagh, A. (2017) Crowdsourcing, sharing economies, and development. *Journal of Developing Societies*. 33(2) 191–222. <http://dx.doi.org/10.1177%2F0169796X17710072>