

Vookan ETL-prosessi

1. Yleistä	2
<u>2. Tiedonkeruu ja esikäsittely</u>	2
2.1 Rajapinnat	2
2.2 Esikäsittely	3
<u>3. Datat yhdistely ja korjaukset</u>	3
3.1 Yhdistetty kaavadata	3
3.2 Geometriakorjaukset	5
3.3 Kohderekisteritunnusten linkittäminen	6
3.4 Yleiskaavojen kuntarajatörmäysten korjaus kiinteistörajoilla	6
3.5 Ominaisuustietokorjaukset	7
<u>4. Geometris-topologinen vertailu</u>	7
4.1 Kiinteistöpalstojen ja kaavaindeksien vertailu	7
4.2 KTJ- ja kunta-aineistojen vertailu	8
<u>5. Ominaisuustietojen vertailu</u>	10
<u>6. PDF-linkityskonversio</u>	11
6.1 PDF-linkitystaulu	11
6.2 Kaavaliitteiden hakemistorakenne	12
6.3 Kaavaliitetiedostojen uudelleennimeäminen	13
6.4 Liitetiedostojen linkitys paikkatietoaineistoon	14
6.5 PDF/A-konversio	14
<u>7. Kaavatietomallin implementointi</u>	14
<u>8. Jatkokehitystarpeet</u>	15
8.1 OGC API Features maantieteellinen rajaus	15
8.2 Yleiskaavojen kuntarajatörmäys	15
8.3 Kohderekisteritunnukset palstatiedoista	16
8.4 Kiinteistöpalstojen ja kaavaindeksien vertailu	16
8.5 DE-9IM-laskenta	16
8.6 Ominaisuustietojen vertailu	16

1. Yleistä

ETL-prosessi toteutettiin Python-ohjelmointikieltä hyödyntäen ja kaikki kehitetyt ohjelmakoodit ovat MIT-lisenssin piirissä avoimesti saatavilla hankkeen [GitHub-sivuilta](#). GitHubiin on listattu ETL-työkalun tekniset vaatimukset (requirements.txt), joista tärkeimpinä moduuleina toimivat GeoPandas sekä Shapely. Prosessin metodologia yhdistelee sekä algoritmisuutta että heuristista lähestymistapaa.

Ohjelmallinen työnkulku on kuvattu kokonaisuudessaan alla. Erilliset vaiheet taas on kuvattu tarkemmin omissa kappaleissaan. Yleiskuvauksena ETL-prosessi koostui kuudesta päävaiheesta (linkit GitHub-sivuille):

1. [Tiedonkeruu ja esikäsittely](#)
2. [Tiedon yhdistely](#)
3. [Tiedon korjaukset](#)
4. [Tiedon vertailu](#)
5. [PDF-linkityskonversio](#)
6. [Kaavatietomallin implementointi](#)

2. Tiedonkeruu ja esikäsittely

2.1 Rajapinnat

ETL-työkalussa tietojen kerääminen on mahdollista teknisesti kolmesta eri lähteestä:

1. OGC:n WFS-standardin mukaisista rajapinnoista (kuntien kaava-aineistot),
2. Esrin ArcGIS Feature Layereista (kuntien kaava-aineistot) sekä
3. OGC API Features -rajapinnoista (MML:n kiinteistötiedot)

Kaksi ensimmäiseksi mainittua palauttavat syötetyn URL:n mukaisen tiedon kokonaisuudessaan GeoPandas GeoDataframena. OGC API Features -skripti sen sijaan palauttaa API:n sisältöä GeoJSON-formaatissa Etelä-Savon maakunnan rajauksella (bounding box), joka on kovakoodattu toteutukseen. Jatkokehityksessä kovakoodattu maantieteellinen rajaus tulee korvata halutun alueen koordinaateilla. OGC API Features -skriptiin on myös lisätty apufunktioita, joiden avulla MML:n kiinteistötietojen GeoJSON-muotoinen dataskaema voidaan normalisoida tabulaariseksi tiedoksi Pandas DataFrameen.

2.2 Esikäsittely

Erään kunnan osalta WFS-rajapinta oli toteutettu KuntaGML-formaatissa, jonka sisäänluku ei onnistunut perinteisin menetelmin. Ongelma ratkaistiin erillisellä [XML-parser-skriptillä](#), jossa kaavatiedot irrotettiin suoraan merkintäkielen rakenteesta.

Osa kunta-aineistoista saatiin rajapintojen sijaan erillistoimitettuina CAD-piirroksina, joissa ominaisuustiedot oli sidottu pistegeometrioihin varsinaisten kaavarajojen sijasta. Ominaisuustietojen yhdistäminen kaavarajoihin toteutettiin [erillisellä skriptillä](#).

MML:n KTJ-aineistoissa kuntaliitosalueiden kaavoille on ilmoitettu vanha kuntakoodi. Nämä päivitettiin vastaamaan [voimassa olevaa kuntakoodia](#) kattavan vertailun mahdollistamiseksi kunta-aineistojen kanssa.

3. Datan yhdistely ja korjaukset

3.1 Yhdistetty kaavadata

Kunnilta saadut kaava-aineistot vaihtelivat sisällöiltään valtavasti. Stabiilin vertailun ja validoinnin mahdollistamiseksi ETL-työkalussa luodaan sekä KTJ- että kunta-aineistoille yhdistetty kaava-aineisto yhtenäisellä skeemalla (Taulukko 1). Tiedot tallennetaan omina tasoinaan yhteisen master-geopackagen alle. Tasot ovat seuraavat:

- Asemakaavat_kunta
- Asemakaavat_ktj
- Yleiskaavat_kunta
- Yleiskaavat_ktj
- Kuntien_rekisterinpitoalueet_ktj (tarvittaessa)

Asemakaava-tasot kattavat sekä asema- että ranta-asemakaavat.

Kaavalajit on ilmoitettu KTJ-aineistossa valmiina Kiinteistötietojärjestelmän kiinteistörekisterin koodistojen mukaisesti. Ne muunnetaan ETL-työkalussa vastaamaan Rakennetun ympäristön tietojärjestelmän koodeja¹. Kuntien aineistoissa kaavalajeja ei ole ilmoitettu erillisellä ominaisuustietokentällä. ETL-päättelee nämä aineiston kaavaselitteistä. Mikäli selitteitä ei ole aineistoissa kerrottu, kaavalajeiksi ilmoitetaan oletuksena yleiskaavoilla 23 = osayleiskaava, ranta-asemakaavoilla 33 = ranta-asemakaava ja asemakaavoilla 31 = asemakaava.

Taulukko 1. Yhdistetyn kaavadataan ominaisuustiedot ja niiden selitteet. Rasti kuvaa kunta- ja KTJ-aineiston osalta sitä, onko ominaisuustieto mukana aineiston skeemassa.

¹ https://koodistot.suomi.fi/codescheme;registryCode=rytj;schemeCode=RY_Kaavalaji

Ominaisuustieto	Selite	Kunta-aineisto	KTJ-aineisto
FID	Rivin yksilöivä tunniste	X	X
originalref	Lähtöaineiston alkuperäinen koordinaattijärjestelmä epsg-koodina	X	X
vanhakuntakoodi	Kuntaliitosalueiden vanha kuntakoodi		X
kuntakoodi	Tilastokeskuksen mukainen voimassa oleva kuntakoodi	X	X
kuntanimi	Kunnan nimi	X	X
kaavatunnus	Kunnan antama kaavatunnus	X	
kaavatunnus_1	KTJ-aineiston kaavatunnuksen alkuosa		X
kaavatunnus_2	KTJ-aineiston kaavatunnuksen loppuosa		X
kaavaselite	Kaavan tai kaavaindeksin selite/nimi	X	X
kaavalaji	Rakennetun ympäristön tietojärjestelmän koodistojen mukainen kaavalajikoodi	X	X
hyvaksymispvm	Kaavan hyväksymispäivämäärä	X	X
vahvistamispvm	Kaavan vahvistamispäivämäärä	X	X
voimaantulopvm	Kaavan voimaantulopäivämäärä	X	X
kohderekisteriyksikot	Lista kaavaindeksiin liittyvistä kiinteistötunnuksista	X	X
kaavakartta	Kaavakartan PDF-liitteen nimi tai hyperlinkki	X	X
maaraykset	Kaavamääräysten PDF-liitteen nimi tai hyperlinkki	X	X
selostus	Kaavaselostuksen PDF-liitteen nimi tai hyperlinkki	X	X

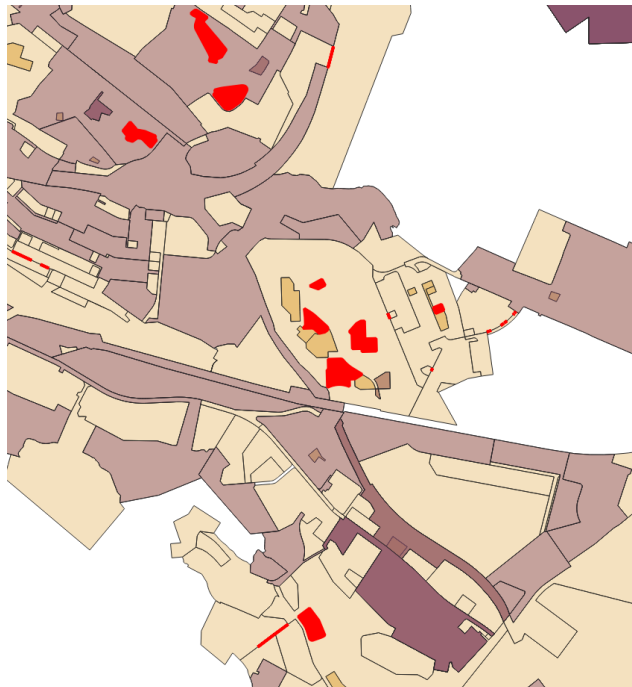
Yhdistetty kaavadata on keskeisessä osassa kaikissa ETL-työkalun vaiheissa, sillä koko toimintalogiikka perustuu muodostetun master-datasetin skeemaan.

3.2 Geometriakorjaukset

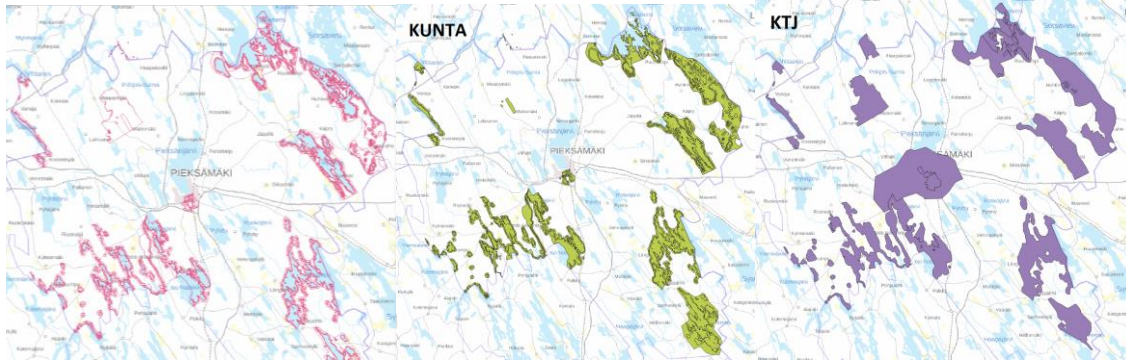
Geometriakorjauksissa ETL-työkalu nojaa vahvasti Pythonin [Shapely-kirjastoon](#). Jokaiselle kaavaindeksille ajetaan explain validity -funktio, joka palauttaa tekstikenttänä tiedon, mikäli geometria itsessään on epävalidi. Esimerkiksi, jos kaavaindeksi risteää itseään tai sisältää ns. sliver-polygonin, funktio palauttaa tekstikentän "Ring Self-intersection" sekä ongelmallisen sijainnin koordinaatit. Mikäli kaavaindeksin geometria on virheellinen, se korjataan Shapelyn make valid -funktiolla.

Usein alkuperäinen geometria tulee räjäyttää useaksi kohteeksi, jotta geometriasta saadaan kelpo. Jos räjäytyksessä tulee luoda useampi kohde samalla geometriatyypillä, palautetaan moniosainen geometria (MultiPolygon). Jos korjauksessa tulee luoda kohteita eri geometriatyypeillä, palautetaan GeometryCollection. Kaavojen tapauksessa vain aluegeometriat ovat sallittuja, joten ETL-työkalu käsittelee mahdollisesti esiintyvät GeometryCollectionit erikseen ja palauttaa ne alueiksi.

Eräässä kunnassa CAD-aineiston muunnos paikkatiedoksi tuotti kaavaindeksille päällekkäisiä "haamugeometrioita". ETL-työkalu suodatti nämä päällekkäisyydet pois ominaisuustietovertailuun pohjautuen.



Kuva 1. Kuntien omat kaavaindeksigeometriat voivat olla haastavia. Tässä esimerkissä kunnan asemakaava- aineistossa oli 852 polygonia, joista löytyi 352 geometriavirhettä.



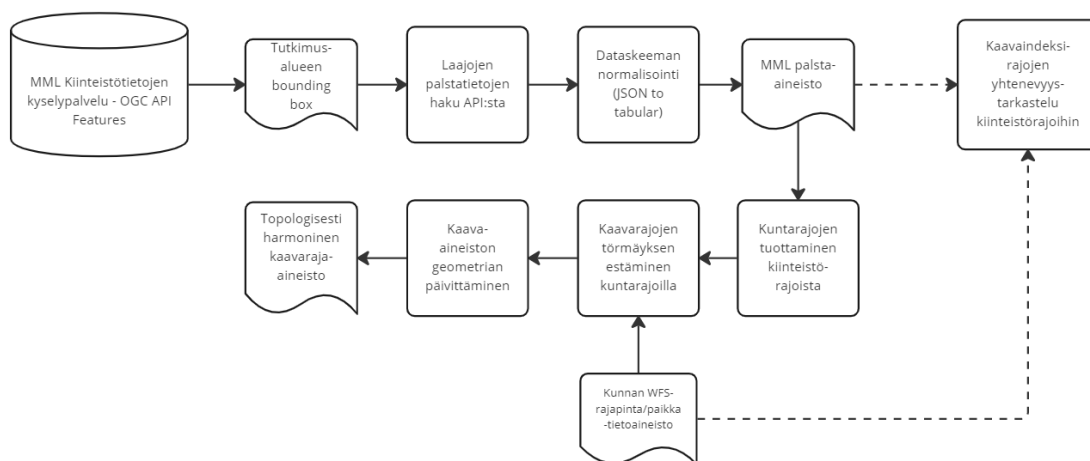
Kuva 2. Viivamaisia rajauksia on hankala automaattisesti muuntaa polygoneiksi, sillä tulkintaa joudutaan tekemään rajausten oikeellisuudesta. Vasemmalla alkuperäinen aineisto. Keskellä polygoneiksi muuntaminen ja oikealla KTJ:stä löytyvät rajaukset.

3.3 Kohderekisteritunnusten linkittäminen

Kiinteistötunnusten sijaintitiedot on saatavilla MML:n kiinteistötietojen OGC API Features -rajapinnasta omana pistetasonaan. ETL-muunnin linkittää jokaiselle kaavaindeksille tiedon siihen liittyvistä kiinteistötunnuksista, mikäli yksittäisen kiinteistötunnuksen sijaintipiste on kaavaindeksin ulkorajojen sisällä.

3.4 Yleiskaavojen kuntarajatörmäysten korjaus kiinteistörajoilla

ETL-prosessin yleistä-osuudessa on mainittu, että prosessin metodologia yhdistelee sekä algoritmisuutta että heuristista lähestymistapaa. Yleiskaavojen kuntarajatörmäysten korjaus kiinteistörajoilla edustaa jälkimmäistä - toteutus perustuu päättelyyn sekä datan arviointiin. Kuva 3 esittää prosessin vaiheet.



Kuva 3. Yleiskaavojen kuntarajatörmäyksen korjausprosessi. Katkoviivalla erotettu osuus on kuvattu tarkemmin osana geometris-topologista vertailua.

Yleiskaavarajojen ETL-muuntimen toimintalogiikka on seuraavanlainen:

1. Kunnalle muodostetaan kuntarajamaski MML:n palstatiedoista saatavista kiinteistörajajoista. Käyttäjä voi halutessaan päättää, otetaanko eksklaavit huomioon vai ei (kiinteistörajat, jotka sijaitsevat varsinaisen kuntarajauksen ulkopuolella).
2. Iteroidaan kaavaindeksit yksittäin läpi ja tarkistetaan, onko kaavaindeksi muodostetun kuntarajamaskin sisällä. Jos indeksi on kokonaan maskin sisällä, ei tehdä mitään (ei ole tarvetta rajamuutoksille).
3. Jos kaavaindeksi ei ole täysin maskin sisällä, se leikkaa kiinteistörajaa osin tai on täysin maskin ulkopuolella (jälkimmäistä tapausta ei esiintynyt Etelä-Savon kunnilla). Tällöin on tarve rajan oikaisemiselle.
4. Lasketaan geometria niille osuuksille, joissa kaavaindeksi risteää kuntarajamaskin kanssa (=intersection).
5. Lasketaan geometria niille osuuksille, joissa kuntarajamaski eroaa kaava-indeksistä (=difference).
6. Filteröidään difference-geometrioita. Hylätään ne difference-geometrian osat, jotka ovat vertailussa olevan kaavaindeksin ulkorajojen sisällä. Lisäksi hylätään ne difference-geometriaosuudet, joissa pinta-ala on poikkeavan suuri. Jokaisella kunnalla pinta-alan suuruudelle on olemassa toleranssi-raja-arvo, jotka on listattu ETL-skriptin docstringiin sekä [Jupyter Notebookiin](#).
7. Lopuksi lasketaan muodostetun intersect-geometrian sekä filteröidyn difference-geometrian liitos (=union), josta muodostuu käsiteltävän kaavaindeksin uusi geometria.

3.5 Ominaisuustietokorjaukset

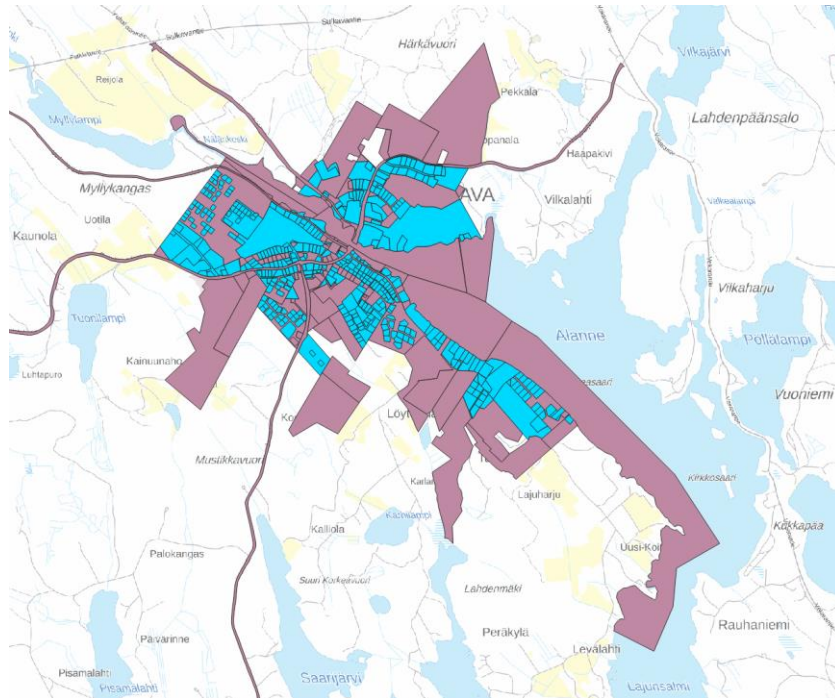
Päivämäärät (esim. kaavan hyväksymispäivämäärä, vahvistamispäivämäärä) oli ilmoitettu kunnilta saaduissa kaava-aineistoissa hyvin vaihtelevissa muodoissa sekä osin puutteellisina. Eri kunnilla oli käytössä päivämääräformaatteina mm. DD.KK.YYYY, YYYYKKDD sekä YYYY. Lisäksi tieto oli usein ilmoitettu päivämäärä-tietotyypin sijasta tekstikenttänä.

Osana ETL-prosessia muodostettiin funktio, joka kääntää tekstimuodossa olevan tabulaarisen tiedon date-tietotyypiksi muotoon DD.KK.YYYY. Mikäli lähtötietona on ainoastaan vuosi, päivämääräksi annettiin kyseisen vuoden tammikuun ensimmäinen päivä.

4. Geometris-topologinen vertailu

4.1 Kiinteistöpalstojen ja kaavaindeksien vertailu

Kiinteistöpalstojen ja kaavaindeksien vertailussa tuodaan esiin, miten hyvin kiinteistörajat yhtyvät kaavaindeksien rajoihin. Topologinen yhtenevyys ilmaistaan prosenttiarvona, joka lasketaan yksittäisen kiinteistöpalstan ja kaavaindeksin risteyskohdan (intersection) pinta-alan suhteena kiinteistöpalstan pinta-alaan.



Kuva 4. Esimerkki palstojen ja kaavaindeksien vertailusta. Sinisellä valittuna palstat, jotka vastaavat kaavaindeksirajoja vähintään 90 % tarkkuudella. Punertavalla palstat, joiden vastaavuus kaavaindeksirajoihin on alle 90 %.

4.2 KTJ- ja kunta-aineistojen vertailu

Geometris-topologinen vertailu tehdään ETL-työkalussa OGC:n DE-9IM -standardin mukaisesti ([Dimensionally Extended 9-Intersection Model](#)), jonka avulla tunnistetaan yhtenevät, päällekkäiset sekä toisistaan eriävät geometriat. Tämän lisäksi ETL-muunnin vertailee kaavarajojen muotoa ja pinta-aloja automaattisesti.

DE-9IM-mallissa kahden geometrian välinen relaatio ilmaistaan 9-merkkisellä mallilla (Taulukko 2). VOOKA-projektissa ylivoimaisesti yleisin relaatiomuoto KTJ:n ja kunta-aineistojen välillä oli "212101212".

Taulukko 2. Esimerkki DE-9IM relaatiomallista.

DE-9IM	Sisusta (Interior)	Raja (Boundary)	Ulkopuoli (Exterior)
Sisusta (Interior)	2	1	2
Raja (Boundary)	1	0	1
Ulkopuoli (Exterior)	2	1	2

Mallia luetaan seuraavasti:

- 2: Geometrioiden risteyskohta on alue (Polygon) eli kaksiulotteinen geometria
- 1: Geometrioiden risteyskohta on viiva (LineString) eli yksiulotteinen geometria
- 0: Geometrioiden risteyskohta on piste (Point) eli nollaulotteinen geometria

Muita DE-9IM-mallissa esiintyviä arvoja voivat olla:

- F: False - risteyskohtaa ei esiinny
- T: True - mikä tahansa ei-false arvo on sallittu (0,1 tai 2)
- *: kaikki sallitaan

Jos kahden geometrian välinen relaatiomalli vastaa muotoa "T*F**FFF*", geometriat ovat täydellisesti samat. VOOKA-projektissa yhdelläkään KTJ:n ja kunta-aineiston vastaavalla kaavalla tämä ehto ei täyttynyt. Tämä johtunee siitä, että kaava-aineistojen digitointi on tapahtunut eri tavalla ja DE-9IM-malli ei salli pienintäkään murto-osan eroavaisuutta. **Tämän takia VOOKA-projektissa vastaavat kaavat tulkittiin geometris-topologisesti yhteneviksi, jos niiden topologinen tarkkuus oli vähintään 98 %.**

Topologinen tarkkuus ilmaistaan ETL-muuntimessa iou-arvolla. Lyhenne tulee sanoista intersection of union, geometrioiden yhtymäkohdan risteyskohta. Mikäli kaksi kaavaa ovat geometris-topologisesti tismalleen samanlaiset, niiden risteyskohdan sekä yhtymäkohdan pinta-alat ovat tismalleen samat. Näiden suhde on täten 100 %. Mikäli kaavoissa on geometris-topologista eroavaisuutta, niiden yhtymäkohta on erisuuri kuin risteyskohta. Tällöin yhtymäkohdan ja risteyskohdan pinta-alojen suhde myös eroaa 100 %:sta.

Geometris-topologista eheyttä ilmaistaan ETL-työkalussa kaiken kaikkiaan Taulukossa 3 ilmaistuun kentin.

Taulukko 3. ETL-työkalussa tuotettavat, geometris-topologista eheyttä ilmaisevat tietokentät.

Ominaisuustieto	Selite
area_ha	Kaavan pinta-ala hehtaareina.
refe_area_ha	Vastaavan kaavan pinta-ala hehtaareina.
de9im_pattern	DE-9IM-mallin mukainen relaatio 9-merkkisellä mallilla.
topo_equal	Boolean-tieto. Onko vertailtavat kaavat topologisesti yhtenevät vai ei. Kyllä, jos iou-arvo on vähintään 98.
iou	Intersection of union -prosentti. Ilmaistaan vertailtavien kaavojen yhtymä- ja risteyskohdan suhteena.
refe_kaavatunnus	Vastaavan kaavan kaavatunnus.
a_delta_%	Vertailtavien kaavojen pinta-alojen suhteellinen

	virheprosentti.
--	-----------------

VOOKA-projektin aikana kuntien datan ja KTJ-datan välillä ei ollut yksilöivää tunnistetta. Täten ominaisuustietojen pohjalta ei voitu tietää, mikä kunnan kaava vastaa KTJ-kaavaa ja päin vastoin. ETL-työkalussa ongelma on ratkaistu siten, että kaavan vastine toisessa datasetissa on se kaava, jonka iou-prosentti on suurin.

Geometris-topologinen vertailu on mahdollista toteuttaa VOOKA-pilotissa tuotetulla työkalulla joko kunta- tai KTJ-aineistopohjaisesti. On kuitenkin suositeltavaa, että KTJ-aineistoa käytetään vertailun pohjana, sillä aineisto on lähtökohtaisesti eheämpää monissa Suomen kunnissa!

5. Ominaisuustietojen vertailu

Siinä missä geometris-topologinen vertailu on mahdollista tehdä myös kunta-aineistoon pohjautuen, ominaisuustietojen vertailu on toteutettu VOOKA-pilotissa vain KTJ-pohjaisesti. KTJ- ja kunta-aineiston ominaisuustietojen yhdenmukaisuutta ilmaistaan ETL-työkalussa kaiken kaikkiaan Taulukossa 4 ilmaistuina kenttinä.

Taulukko 4. ETL-työkalussa tuotettavat, ominaisuustietojen yhtenevyyttä ilmaisevat tietokentät.

Ominaisuustieto	Selite
kl_equal	Boolean-tieto. Onko KTJ-aineiston kaavalaji-tieto sama kuin kunta-aineistossa.
hyv_equal	Kertoo, onko KTJ- ja kunta-aineiston hyväksymispäivämäärä sama.
voim_equal	Kertoo, onko KTJ- ja kunta-aineiston voimaantulopäivämäärä sama.

Ensimmäinen ominaisuustietokenttä "kl_equal" kertoo boolean-tietona onko vertailtavien aineistojen kaavalajit samat. Näissä on ajoittain paljonkin heittoa VOOKA-pilotin eri kunnissa, sillä:

1. KTJ kertoo asemakaavoissa usein, onko kaava ohjeellista tonttijakoa vai ei. Pilotin kunta-aineistoista tämä tieto puuttuu.
2. KTJ:ssa ei ole tietoa oikeusvaikutuksettomista kaavoista. Tämä tieto on myös VOOKA-pilotin kunta-aineistoissa harvinainen, mutta ajoittain esiintyvä.

Toinen ja kolmas ominaisuustietokenttä "hyv_equal" sekä "voim_equal" ilmaisevat hyväksymis- ja voimaantulopäivämäärien yhteneväisyyden sekä niihin liittyvät mahdolliset puutteet. Kentissä voi esiintyä:

1. "Hyvpvm/voimpvm puuttuu molemmista"
 - a. Sekä KTJ- että kunta-aineistosta puuttuu tarkasteltava päivämäärä kaavan osalta.
2. "Hyvpvm/voimpvm puuttuu KTJ"
 - a. KTJ-aineistosta puuttuu tarkasteltava päivämäärä kaavan osalta.
3. "Hyvpvm/voimpvm puuttuu kunta"
 - a. Kunnan aineistosta puuttuu tarkasteltava päivämäärä kaavan osalta.
4. 1 - Tarkasteltavat päivämäärät ovat samat
5. 0 - Tarkasteltavat päivämäärät eroavat toisistaan

Vahvistamispäivämäärät eivät ole mukana vertailussa, sillä ne puuttuvat KTJ-aineistosta kokonaan. Yleisenä periaatteena vertailussa mukana olevat ominaisuustiedot ovat ainoita, joita VOOKA-pilotissa kyettiin vertailemaan.

6. PDF-linkityskonversio

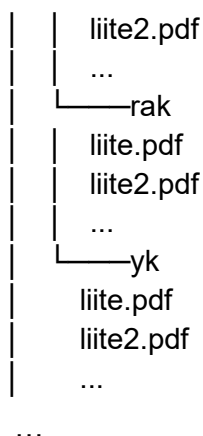
6.1 PDF-linkitystaulu

PDF-liitteiden linkittäminen yhtenäisessä muodossa kaavojen paikkatietoihin on mahdollista, jos tiedetään, mihin kaavaindeksitunnukseen kukin liite kytkeytyy. VOOKA-pilottiprojektissa tämä tieto puuttui lähes poikkeuksetta kaavojen paikkatietoaineistoista. Tästä syystä kaavaliitteiden ja niitä vastaavien kaavaindeksien tunnistustyö toteutettiin manuaalisena työnä.

ETL-työkalun PDF-linkitystä varten koostettiin liitostaulu (Taulukko 5), jossa jokaisella rivillä on tieto kaavan indeksitunnuksesta (KTJ, kunnan aineisto tai molemmat) sekä kaavan dokumenttityypistä (esim. osallistamis- ja arviointisuunnitelma). Kaiken kaikkiaan liitostaulun skeema oli seuraavanlainen:

Taulukko 5. PDF-linkitystaulun ominaisuustietokentät ja niiden selitteet.

Ominaisuustieto	Selite
Kunta	Kunnan virallinen kuntakoodi.
Kunnan indeksitunnus	Kunnan paikkatietomuotoisen kaava-aineiston kaavaindeksitunnus.
KTJ-indeksitunnus	KTJ-aineiston kaavaindeksitunnus (kaavatunnus_1).



ETL-työkalu hyödyntää kyseistä hakemistorakennetta tiedostojen uudelleennimeämisessä!

6.3 Kaavaliitetiedostojen uudelleennimeäminen

Liitetiedostojen uudelleennimeämisen periaatteena on yhdenmukaisten liitenimien muodostus. ETL-työkalu muodostaa vakioidut liitenimet liitostauluun uuteen "new_name" - ominaisuustietokenttään seuraavin periaattein:

1. Liitenimi huomioi virallisen kuntakoodin
2. Liitenimi huomioi Rakennetun ympäristön tietojärjestelmän (Ryhti) koodistojen mukaisesti kaavalajin
3. Liitenimi huomioi Ryhtin koodistojen mukaisesti asiakirjan lajin
4. Liitenimi huomioi kaavatunnuksen
5. Kaksi tai useampi liite ei saa olla samanniminen

Esimerkki ETL-työkalun tuottamasta liitenimestä on "491-31-0304-13030-2.pdf", jossa:

- 491 on Mikkelin virallinen kuntakoodi
- 31 on Ryhtin kaavalaji-koodiston mukaisesti asemakaava
- 0304 viittaa Ryhtin asiakirjan laji -koodiston mukaisesti siihen, että liite kattaa sekä kaavakartan (03) että määräykset (04)
- 13030 on kunnan indeksitunnus
- 2 on liukuva järjestysnumero (491-31-0304-13030-1.pdf on jo olemassa, eli voimassa olevalle kaavalle on löytynyt kaksi samannimistä liitettä)

ETL-työkalu mahdollistaa liitetiedoston uudelleennimeämisen joko KTJ-indeksitunnukseen tai kunnan omaan indeksitunnukseen perustuen.

HUOM! Kaavaliitetiedostojen uudelleennimeämiskäytäntö on ainoastaan ehdotus liitteiden nimeämiselle Ryhtissä! Vakioitu nimeämiskäytäntö voi muuttua Ryhti-toteutusprojektin aikana!

6.4 Liitetiedostojen linkitys paikkatietoaineistoon

ETL-työkalu linkittää uudelleennimetyt tiedostonimet linkitystaulusta paikkatietoaineistoon kaavalajin, kuntakoodin sekä kaavan indeksitunnuksen perusteella (joko KTJ tai kunnan tunnus, riippuen kumman mukaan liitteet on uudelleennimetty). Kaavalaji ja kuntakoodi tarvitaan, koska:

- 1) Kunnan sisällä esim. ranta-asemakaavoilla voi olla sama kunnan indeksitunnus kuin asemakaavoilla
- 2) Eri kunnilla voi olla tismalleen sama kaavatunnus.

Kukin liite löytyy ajon jälkeen paikkatietoaineistosta omasta kentästään (esim. osallistamis- ja arviointisuunnitelmat kentästä "oas").

6.5 PDF/A-konversio

ETL-työkalu konvertoi kunnilta saadut liite-PDF-tiedostot automatisoidusti Ghostscript-muuntimella PDF/A-arkistoformaattiin. Muunnin perustuu soveltuvin osin GitHubista avoimesti saatavilla olevaan [Python-Ghostscript-konvertteriin](#). VOOKA-projektin tarpeisiin muokattu muunnin löytyy [omana versionaan](#) projektin GitHub-sivuilta.

Jos saatu PDF-formaatti on toteutettu versiolla, jota ei saada PDF/A-muotoon, se tallennetaan sellaisenaan. Mikäli jotkin kunnan PDF-tiedostot eivät käänny kunnan aineistosta johtuvista teknisistä syistä PDF/A-muotoon, tallennetaan nämä normaalissa PDF-muodossa.

7. Kaavatietomallin implementointi

ETL-työkalussa [kaavatietomallin](#) implementointi tapahtuu muuntamalla GeoPandas GeoDataFrame -muotoinen kaava-aineisto GeoJSON-formaattiin. ETL-muuntimelle ilmoitetaan aineistolähteeksi joko "kunta" tai "KTJ" sekä molempien aineistolähteiden kaavatunnukset. Osana aineistojen vertailua kohdassa 4.2 on luotu tietokenttä "refe_kaavatunnus", joka kertoo esim. KTJ-pohjaisessa aineistossa kaavan vastaavan kuntatunnuksen. Mikäli vastaavaa tunnusta ei löydy (esim. kaavakohte on vain KTJ:n aineistossa), ETL-muunnin jättää kunnan indeksitunnus-tiedon GeoJSON:ssa tyhjäksi.

ETL-työkalun avulla on lisäksi mahdollista normalisoida kaavatietomallin mukainen JSON-aineisto takaisin geopackage-formaattiin, jotta aineisto on kunnille helpommin luettavissa sekä tarkasteltavissa paikkatieto-ohjelmistoissa.

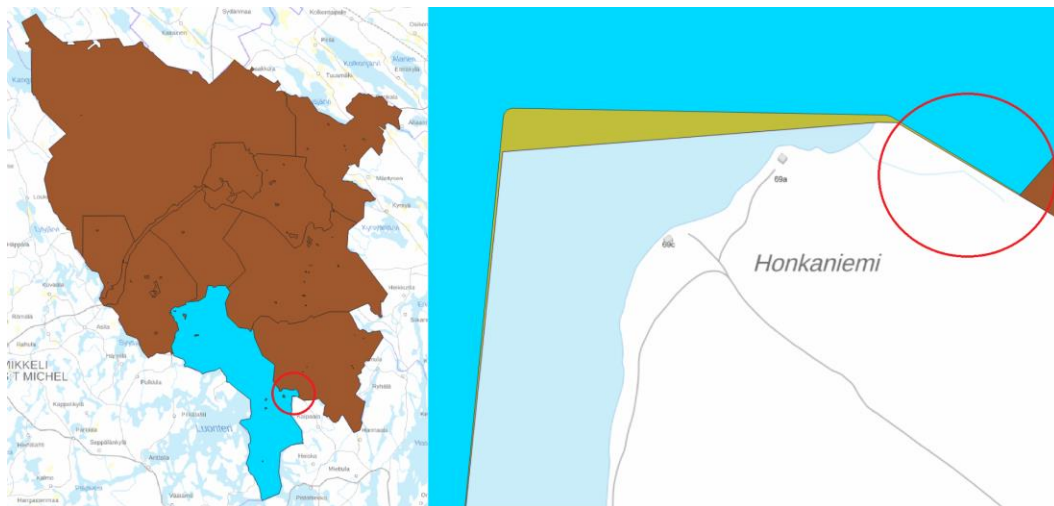
8. Jatkokehitystarpeet

8.1 OGC API Features maantieteellinen raja

Etelä-Savon maakunnan maantieteelliset rajat on kovakoodattu bounding boxina OGC API Features getter-skriptiin. Jatkossa muilla maantieteellisillä alueilla bounding box tulee korvata halutuilla koordinaateilla.

8.2 Yleiskaavojen kuntarajatörmäys

Yleiskaavojen kuntarajatörmäyksen korjauslogiikka toimii hyvin, muttei aukottomasti. VOOKA-projektin aikana tunnistettiin yksi selkeä tapaus, johon logiikan avulla ei ole vielä annettu vastausta (Kuva 5). Kaavaindeksi jää toisen kaavaindeksin rajalla "auki", jolloin laskennassa muodostuva differencen osageometria ei filteröidy pinta-alavertailulla pois.



Kuva 5. Esimerkki tilanteesta, jossa yleiskaavaindeksin kuntarajatörmäyksen korjaus ei toimi aukottomasti. Kuntarajamaskin ja kaavaindeksin differencea laskettaessa indeksi jää toisen kaavaindeksin rajalla "auki" suhteessa maskiin, jolloin eroavaisuus ei filteröidy pinta-alavertailulla pois. Sininen alue (kaavaindeksi) ei yhdisty vihreään (kiinteistö-kuntaraja), vaikka pitäisi.

Pinta-alan suuruuteen pohjautuva difference-geometrioiden filterointi perustuu kokeiluun, sillä kunta-aineistojen spatiaalisissa tarkkuuksissa on paljon eroja. VOOKA-projektin osalta jokaisen kunnan toimiva raja-arvo on testattu erikseen ja listattu referenssiksi jatkokehitystä varten. Vaikka kuntien toleranssiarvoissa onkin yhdenmukaisuutta, niihin ei voi luottaa sokeasti eri maantieteellisillä alueilla, vaan jokaiselle kunnalle tulee löytää toimiva raja-arvo testauksen kautta erikseen. Jatkokehitysideana tähän tarpeeseen voisi myös kouluttaa AI/ML-toimintoja arvioimaan sopivinta toleranssi-arvoa kullekin kunnalle.

8.3 Kohderekisteritunnukset palstatiedoista

Kiinteistötunnusten sijaintitiedot on pistetaso lisäksi tarjolla MML:n OGC API Features -rajapinnasta aluetasona osana palstojen laajoja sijaintitietoja. Aluemaisen referenssitason hyödyntäminen kohderekisteritunnusten linkittämisessä kaavaindekseihin olisi tarkempi tapa kuin pistetaso hyödyntäminen. VOOKA-projektin toteutuksessa osa kiinteistötunnuksista voi jäädä linkittämättä kaavaindekseihin, jotka todellisuudessa sivuavat kiinteistötunnusta, sillä kiinteistötunnuksen sijaintitieto on vaikutusalueen sijasta sidottu vain yhteen pisteeseen.

Lisäksi rakennetun ympäristön tietojärjestelmän kaavatietomallissa halutaan tietää kohderekisteritunnusten osalta, että kuuluvatko ne kokonaan kaavan sisälle sekä sisältyvän alueen pinta-ala. Tätä tietoa ei ole mahdollista saada tietoon ainoastaan pistetasoa hyödyntämällä.

8.4 Kiinteistöpalstojen ja kaavaindeksien vertailu

VOOKA-projektissa kehitetty ETL-työkalu ei tunnista sitä, onko kiinteistörajat muodostettu ennen kaavan voimaantuloa vai sen jälkeen. Näin ollen se ei pysty indikoimaan, onko kaavaindeksin rajan tarkoitus olla yhtenevä kiinteistörajan kanssa vai ei.

8.5 DE-9IM-laskenta

DE-9IM-laskennassa ei oteta huomioon sitä, että vertailun vastaava kaava voi olla false positive. Tästä johtuen esimerkiksi pinta-alojen suhteelliset virheprosentit nousevat pahimmillaan tuhansiin ja iou-prosentti on minimaalinen. ETL-muuntimeen olisi hyvä päivittää raja-arvot, jotka filteröivät false positive -tapauksia pois. Esimerkiksi vastaavaksi kaavaksi ei hyväksytä kaavaa, jolla pinta-alojen suhteellinen virheprosentti ylittää 100 %

8.6 Ominaisuustietojen vertailu

VOOKA-pilotissa ominaisuustietojen vertailu on mahdollista vain kolmelle tietokentälle (kaavalaji, hyväksymispäivämäärä ja voimaantulopäivämäärä) KTJ-aineistopohjaisesti. Muilla maantieteellisillä alueilla ominaisuustiedot voivat olla kattavampia, jolloin vertailuun voi tuoda mukaan myös muita tietokenttiä. Lisäksi KTJ-pohjainen vertailuperiaate on mahdollista laajentaa myös kuntapohjaiseksi.