

ETL-processen i VOOKA

1. Allmänt	3
2. Datainsamling och förbehandling	3
2.1 Gränssnitt	3
2.2 Förbehandling	4
3. Kombination och korrigerig av data	4
3.1 Kombinerade plandata	4
3.2 Geometrikorrigeringar	6
3.3 Länkning av objektregisterbeteckningar	7
3.4 Korrigerig av kommungränskollisioner i generalplaner vid fastighetsgränser	7
3.5 Korrigeringar av egenskapsdata	8
4. Geometrisk-topologisk jämförelse	9
4.1 Jämförelse av fastighetstomt och planindex	9
4.2 Jämförelse av FDS- och kommunmaterial	9
5. Jämförelse av egenskapsuppgifter	11
6. PDF-länkningskonvertering	12
6.1 PDF-länkningstavla	12
6.2 Katalogstruktur för planbilagor	14
6.3 Nya namn på planbilagor	14
6.4 Länkning av bilagor till geodatamaterial	15
6.5 PDF/A-konvertering	15
7. Implementering av plandatamodellen	15
8. Behov av vidareutveckling	16
8.1 OGC API Features geografisk avgränsning	16
8.2 Kommungränskollisioner på generalplaner	16
8.3 Objektregisterbeteckningar från tomtuppgifterna	17
8.4 Jämförelse av fastighetstomt och planindex	17
8.5 DE-9IM-beräkning	17
8.6 Jämförelse av egenskapsuppgifter	17

1. Allmänt

ETL-processen genomfördes med hjälp av Python-programmeringsspråket och alla programkoder som utvecklats finns öppet tillgängliga under MIT-licensen på projektets [GitHub-sidor](#). I GitHub finns en förteckning över ETL-verktygets tekniska krav (requirements.txt), varav de viktigaste modulerna är GeoPandas och Shapely. Processens metodologi kombinerar både algoritm och ett heuristiskt förhållningssätt.

Det programmässiga arbetsflödet beskrivs i sin helhet nedan. De separata faserna beskrivs däremot närmare i egna avsnitt. Som allmän beskrivning bestod ETL-processen av sex huvudfaser (länkar till GitHub-sidorna):

1. [Datainsamling och förbehandling](#)
2. [Sammanställning av uppgifter](#)
3. [Korrigeringar av uppgifter](#)
4. [Jämförelse av uppgifter](#)
5. [PDF-länkningskonvertering](#)
6. [Implementering av plandatamodellen](#)

2. Datainsamling och förbehandling

2.1 Gränssnitt

I ETL-verktyget är det tekniskt möjligt att samla in uppgifter från tre olika källor:

1. från gränssnitt enligt OGC:s WFS-standard (kommunernas planmaterial),
2. från Esri ArcGIS Feature Layers (kommunernas planmaterial) samt
3. OGC API Features -gränssnitt (LMV:s fastighetsuppgifter)

De två förstnämnda returnerar den inmatade informationen enligt URL i sin helhet som GeoPandas GeoDataFrame. Skriptet OGC API Features returnerar däremot API:s innehåll i GeoJSON-format med en avgränsning av landskapet Södra Savolax (bounding box), som har hårdkodats för genomförandet. I den fortsatta utvecklingen ska den hårdkodade geografiska avgränsningen ersättas med koordinater för det önskade området. I OGC API Features-skriptet har man också lagt till hjälpfunktioner med vilka LMV:s dataschema i geoJSON-format för fastighetsuppgifter kan normaliseras till tabulär information i Pandas DataFrame.

2.2 Förbehandling

För en viss kommun hade WFS-gränssnittet genomförts i KommunGML-format, som inte kunde läsas in med traditionella metoder. Problemet löstes med ett separat [XML-parser-skript](#), där planinformationen lösgjordes direkt från märkspråkets struktur.

En del av kommunmaterialet erhöles som CAD-ritningar som levererades separat i stället för som gränssnitt, där egenskapsuppgifterna var bundna till punktgeometrier i stället för de egentliga plangränserna. Egenskapsuppgifterna kombinerades med plangränserna med ett [separat skript](#).

I LMV:s FDS-material anges den gamla kommunkoden för planerna för kommunsammanslagningssområden. Dessa uppdaterades så att de motsvarar [den gällande kommunkoden](#) för att möjliggöra en omfattande jämförelse med kommunmaterialet.

3. Kombination och korrigerig av data

3.1 Kombinerade plandata

Innehållet i planmaterialet från kommunerna varierade enormt. För att möjliggöra stabil jämförelse och validering skapas planmaterial i ETL-verktyget som kombinerats med både FDS- och kommunmaterial med ett enhetligt schema (Tabell 1). Uppgifterna sparas som egna nivåer under den gemensamma master-geopackage. Nivåerna är följande:

- Detaljplaner_kommun
- Detaljplaner_fds
- Generalplaner_kommun
- Generalplaner_fds
- Kommunernas_registerföringsområden_fds (vid behov)

Detaljplanenivåerna omfattar både detaljplaner och stranddetaljplaner.

Planslagen har meddelats i FDS-materialet i enlighet med koduppsättningarna i Fastighetsdatasystemets fastighetsregister. De omvandlas i ETL-verktyget så att de motsvarar koderna i datasystemet för den byggda miljön¹. I kommunernas material har plantyperna inte angetts i ett separat egenskapsfält. ETL-avgör dessa utifrån planförklaringarna i materialet. Om inga förklaringar har angetts i materialet anges som plantyper enligt antagandet i generalplanerna 23 = delgeneralplan, i stranddetaljplanerna 33 = stranddetaljplan och detaljplanerna 31 = detaljplan.

¹ <https://koodistot.suomi.fi/codescheme;registryCode=rytj;schemeCode=RY> Kaavalaji

Tabell 1. Egenskapsuppgifter för kombinerade plandata och förklaringar till dem. Krysset beskriver för kommun- och FDS-materialets del om egenskapsuppgiften ingår i materialets schema.

Egenskapsuppgifter	Förklaring	Kommunmaterial	FDS-material
FID	Kod som identifierar raden	X	X
originalref	Utgångsmaterialets ursprungliga koordinatsystem som epsg-kod	X	X
den gamla kommunkoden	Den gamla kommunkoden för områden med kommunsammanslagningar		X
kommunkod	Gällande kommunkod enligt Statistikcentralen	X	X
kommunnamn	Kommunens namn	X	X
planbeteckning	Planbeteckning som kommunen gett	X	
kaavatunnus_1	Första delen av planbeteckningen för FDS-material		X
kaavatunnus_2	Slutdelen av planbeteckningen för FDS-material		X
planbeskrivning	Beskrivning av/namn på plan eller planindex	X	X
planslag	Kod för planslag enligt koduppsättningen i datasystemet för den byggda miljön	X	X
godkännandedatum	Datum för godkännande av planen	X	X
fastställersedatum	Datum då planen fastställdes	X	X
ikraftträdandedatum	Datum då planen träder i kraft	X	X
objektregisterenheter	Förteckning över fastighetsbeteckningar i anslutning till planindex	X	X
plankarta	Namn på plankartans PDF-bilaga eller hyperlänk	X	X
bestämmelser	Namn på plankartans PDF-bilaga eller hyperlänk	X	X
beskrivning	Namn på planbeskrivningens PDF-bilaga eller hyperlänk	X	X

Kombinerade plandata har en central roll i alla skeden av ETL-verktyget, eftersom hela verksamhetslogiken grundar sig på ett schema i ett masterdataset.

3.2 Geometrikorrigeringar

Vid geometrikorrigeringar stöder sig ETL-verktyget starkt på Pythons [Shapely-bibliotek](#). För varje planindex körs explain validity-funktionen som returnerar uppgiften som textfält om geometrin i sig är ogiltig. Om formelindexet till exempel korsar sig själv eller innehåller en s.k. sliver-polygon, återställer funktionen textfältet "Ring Self-intersection" samt koordinaterna för den problematiska positionen. Om planindexets geometri är felaktig, korrigeras den med Shapelys make valid-funktion.

Ofta ska den ursprungliga geometrin delas upp i flera objekt för att geometrin ska bli duglig. Om flera objekt ska med samma geometrityp ska skapas vid uppdelning återställs geometrin i flera delar (MultiPolygon). Om objekt ska skapas med olika geometrityper, återställs GeometryCollection. I fråga om planerna är endast områdesgeometrier tillåtna, varmed ETL-verktyget behandlar eventuella GeometryCollections separat och återställer dem till områden.

I en kommun producerade konverteringen av CAD-materialet till geodata överlappande "spökgeometrier" för planindexen. ETL-verktyget filtrerade bort dessa överlappningar utifrån jämförelsen av egenskapsuppgifter.

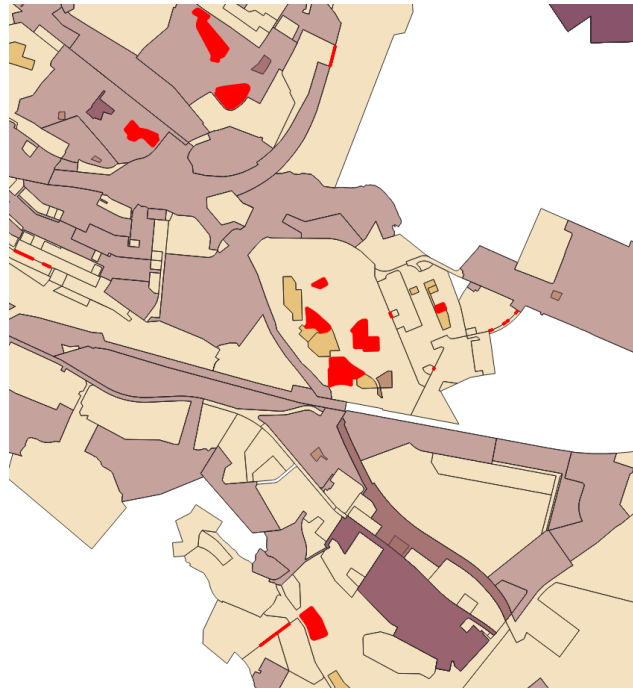


Bild 1. Kommunernas egna planindexgeometrier kan vara utmanande. I det här exemplet fanns det 852 polygoner med 352 geometrifel i kommunens detaljplanematerial.

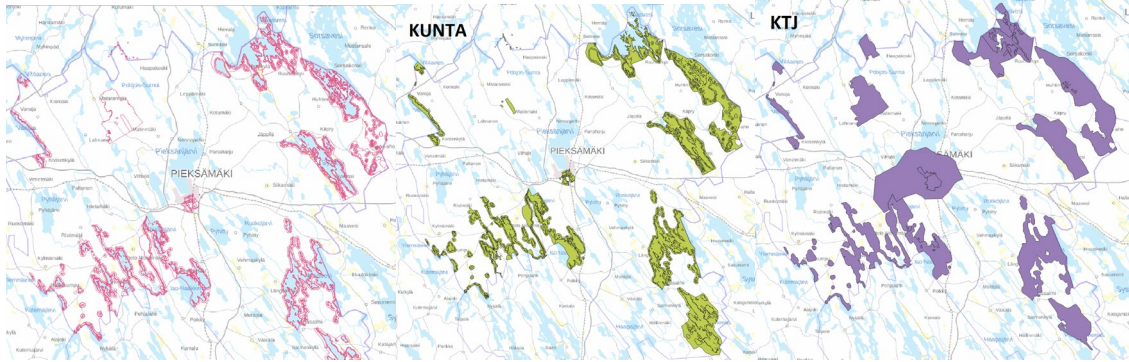


Bild 2. Det är svårt att automatiskt omvandla sträckliknande avgränsningar till polygoner, eftersom man måste tolka avgränsningarnas riktighet. Till vänster originalmaterialet. I mitten omvandling till polygoner och till höger i FDS visas avgränsningar.

3.3 Länkning av objektregisterbeteckningar

Lokaliseringsuppgifter för fastighetsbeteckningarna finns i LMV:s OGC API Features-gränssnitt för fastighetsuppgifter som en egen poängnivå. ETL-omvandlaren länkar till varje planindex information om fastighetsbeteckningar i anslutning till den huruvida lokaliseringpunkten för en enskild fastighetsbeteckning är inom planindexets yttre gränser.

3.4 Korrigering av kommungränskollisioner i generalplaner vid fastighetsgränser

I ETL-processens avsnitt allmänt, nämns att processens metodologi kombinerar både algoritm och ett heuristiskt förhållningssätt. Korrigeringen av kommungränskollisioner i generalplanerna vid fastighetsgränserna representerar det senare - genomförandet grundar sig på slutledning och bedömning av data. Bild 3 visar processens faser.

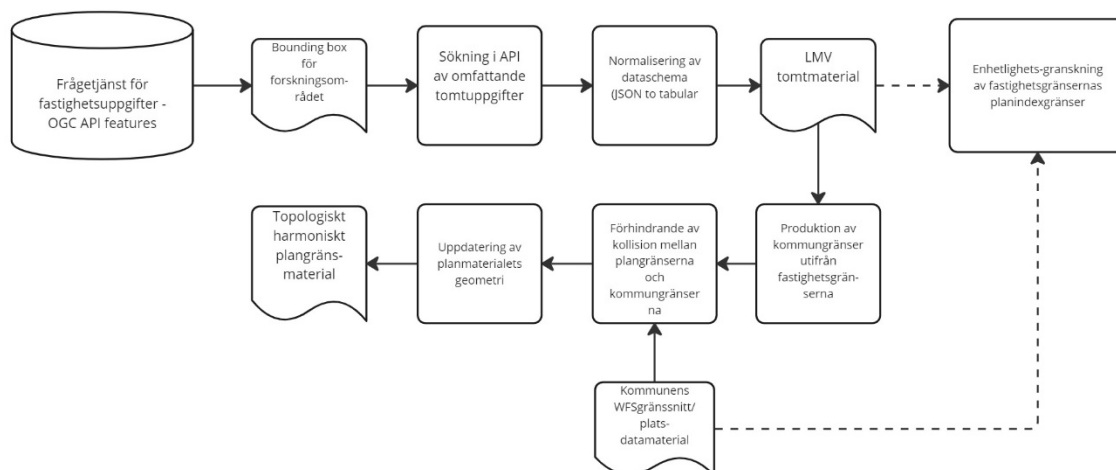


Bild 3. Processen för korrigering av kommungränskollision i generalplanerna. Det avsnitt som separerats med en streckad linje beskrivs närmare som en del av den geometrisk-topologiska jämförelsen.

Verksamhetslogiken för ETL-omvandlaren vid generalplanegränserna är följande:

1. För kommunen bildas en maskering av kommungränsen över fastighetsgränserna som fås från LMV:s tomtdata. Användaren kan om hen så önskar besluta om huruvida exklaverna ska beaktas eller inte (fastighetsgränser som ligger utanför den egentliga kommunavgränsningen).
2. Planindexen itereras individuellt och man kontrollerar om planindexet finns inom den bildade kommungränsmasken. Om indexet omfattas helt av maskeringen görs inget (det finns inget behov av gränsändringar).
3. Om planindexet inte omfattas av maskeringen helt, skär det fastighetsgränsen delvis eller är helt utanför maskeringen (det senare fallet förekom inte i kommunerna i Södra Savolax). Då finns det ett behov av att korrigera gränsen.
4. Man beräknar geometrin för de avsnitt där planindexet korsar maskeringen av kommungränsen (=intersection).
5. Man beräknar geometrin för de avsnitt där maskeringen av kommungränsen avviker från planindexet (=difference).
6. Man filtrerar difference-geometrier. De delar av difference-geometrin som ligger inom de yttre gränserna för det planindex som jämförs förkastas. Dessutom förkastas de difference-geometriandelar där arealen är exceptionellt stor. Varje kommun har ett tolerans-gränsvärde för arealen som finns listat i ETL-skriptets docstring och [Jupyter Notebook](#).
7. Till sist beräknas kopplingen mellan den bildade intersect-geometrin och den filtererade difference-geometrin (= union), som bildar en ny geometri för det planindex som behandlas.

3.5 Korrigeringar av egenskapsdata

Datumen (t.ex. datum för godkännande av planen, datum för fastställande) hade angetts i mycket varierande former och delvis bristfälligt i planmaterialet från kommunerna. Olika kommuner använde som datumformat bl.a. DD.MM.YYYY, YYYYMMDD och YYYY. Dessutom hade uppgiften ofta angetts som textfält i stället för som datum-uppgiftstyp.

Som en del av ETL-processen skapades en funktion som översätter tabulär data i textform till date-datatyp i formen DD.MM.YYYY. Om utgångsuppgiften endast är ett år, gavs som datum den första dagen i januari året i fråga.

4. Geometrisk-topologisk jämförelse

4.1 Jämförelse av fastighetstomt och planindex

I jämförelsen av fastighetstomter och planindex lyfter man fram hur väl fastighetsgränserna samstämmer med gränserna för planindexen. Den topologiska samstämmigheten uttrycks som ett procenttal som beräknas som förhållandet mellan arealen för ett enskilt fastighetsskifte och för en korsningspunkt i planindexet (intersection) och fastighetstomtens areal.

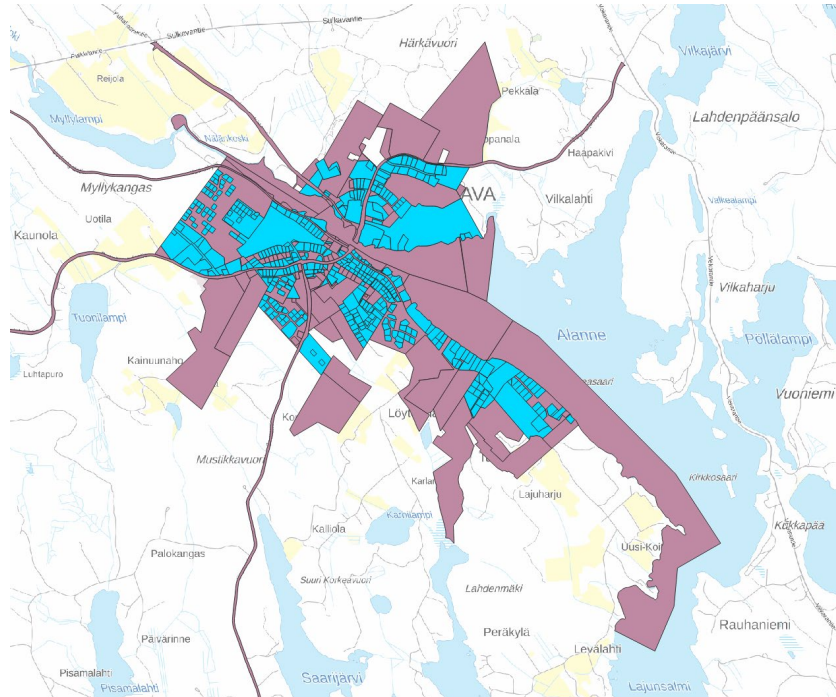


Bild 4. Exempel på jämförelse av tomter och planindex. Med blått de tomter som motsvarar planindexgränserna med minst 90 % noggrannhet. Med rödaktigt de tomter vars motsvarighet till planindexgränserna är under 90 %.

4.2 Jämförelse av FDS- och kommunmaterial

Geometrisk-topologisk jämförelse görs i ETL-verktyget enligt OGC:s DE-9IM-standard ([Dimensionally Extended 9-Intersection Model](#)), med hjälp av vilken man identifierar enhetliga, överlappande och olika geometrier. Dessutom jämför ETL-omvandlaren plangränsernas form och arealer automatiskt.

I DE-9IM-modellen anges relationen mellan två geometrier med en 9-siffrig modell (Tabell 2). I VOOKA-projektet var den överlägset vanligaste relationsformen mellan FDS och kommunmaterialet "212101212".

Tabell 2. Exempel på DE-9IM relationsmodell.

DE-9IM	Insida (Interior)	Gräns (Boundary)	Utsida (Exterior)
Insida (Interior)	2	1	2
Gräns (Boundary)	1	0	1
Utsida (Exterior)	2	1	2

Modellen läses enligt följande:

- 2: Korsningspunkten för geometrierna är ett område (Polygon) dvs. tvådimensionell geometri
- 1: Korsningspunkten för geometrierna är en linje (LineString) dvs. endimensionell geometri
- 0: Korsningspunkterna för geometrierna är en punkt (Point), dvs. nolldimensionell geometri

Andra värden i DE-9IM-modellen kan vara:

- F: False - korsningspunkten förekommer inte
- T: True - vilket icke-false värde som helst är tillåtet (0,1 eller 2)
- *: allt är tillåtet

Om relationsmodellen mellan två geometrier motsvarar formen "T*F**FF*", är geometrierna helt lika. I VOOKA-projektet uppfylldes inte detta villkor för en enda plan som motsvarade FDS och kommunmaterial. Detta beror sannolikt på att digitaliseringen av planmaterialet har skett på olika sätt och DE-9IM-modellen tillåter inte ens en bråkdel avvikelse. **Därför tolkades motsvarande planer i VOOKA-projektet som geometrisk-topologiskt enhetliga om deras topologiska noggrannhet var minst 98 %.**

Topologisk noggrannhet uttrycks i ETL-omvandlaren med ett iou-värde. Förkortningen kommer från orden intersection of union, korsningspunkten för geometriernas beröringspunkt. Om de två planerna är geometrisk-topologiskt exakt likadana, är deras korsningspunkt och beröringspunktens arealer exakt desamma. Förhållandet mellan dessa är således 100 %. Om det finns geometrisk-topologiska skillnader i planerna är deras beröringspunkt av en annan storlek än korsningspunkten. Då skiljer sig förhållandet mellan beröringspunktens och korsningspunktens arealer också från 100 %.

Geometrisk-topologisk integritet uttrycks i ETL-verktyget totalt i fälten i Tabell 3.

Tabell 3. Datafält som uttrycker geometrisk-topologisk integritet och som produceras i ETL-verktyget.

Egenskapsuppgifter	Förklaring
area_ha	Planens areal i hektar.

refe_area_ha	Areal i hektar för en motsvarande plan.
de9im_pattern	Relation enligt DE-9IM-modellen med modellen med 9-siffror.
topo_equal	Boolean-information. Är planerna som jämförs topologiskt enhetliga eller inte. Ja, om iou-värdet är minst 98.
iou	Intersection of union-procent. Anges som förhållandet mellan berörings- och korsningspunkten i de planer som jämförs.
refe_planbeteckning	Planbeteckning för motsvarande plan.
a_delta_%	Den relativa felprocenten för arealerna i planerna som jämförs.

Under VOOKA-projektet fanns det ingen identifierare mellan kommunernas data och FDS-data. På basis av egenskapsuppgifterna kunde man således inte veta vilken kommuns plan som motsvarade FDS-planen och tvärtom. I ETL-verktyget har problemet lösts så att motsvarigheten till planen i det andra datasetet är den plan vars iou-procent är störst.

En geometrisk-topologisk jämförelse kan genomföras med ett verktyg som producerats i VOOKA-pilotprojektet antingen kommun- eller FDS-materialbaserat. Det rekommenderas dock att FDS-materialet används som grund för jämförelsen eftersom materialet i princip är enhetligare i många kommuner i Finland!

5. Jämförelse av egenskapsuppgifter

Medan det är möjligt att göra en geometrisk-topologisk jämförelse även utifrån kommunmaterialet har jämförelsen av egenskapsuppgifterna i VOOKA-pilotprojektet endast genomförts på FDS-basis. Sammanfattningsvis uttrycks enhetligheten i FDS- och kommunmaterialets egenskapsuppgifter i ETL-verktyget i fälten i Tabell 4.

Tabell 4. Datafält som uttrycker egenskapsuppgifternas enhetlighet och som produceras i ETL-verktyget.

Egenskapsuppgifter	Förklaring
kl_equal	Boolean-information. Är uppgiften om planslag i FDS-materialet densamma som i kommunmaterialet.
hyv_equal	Anger om datumet för godkännande av FDS- och kommunmaterialet är detsamma.
voim_equal	Anger om FDS- och kommunmaterialets

ikraftträdandedatum är detsamma.

Det första egenskapsfältet "kl_equal" anger som boolean-information om de jämförda materialens planslag är desamma. I dessa finns det tidvis stora variationer i VOOKA-pilotprojektets olika kommuner, eftersom:

1. FDS ofta i detaljplanerna berättar om planen är en riktgivande tomtindelning eller inte. I pilotprojektets kommunmaterial saknas denna information.
2. FDS innehåller ingen information om planer utan rättsverkningar. Denna information är också sällsynt men förekommer tidvis i VOOKA-pilotprojektets kommunmaterial.

Det andra och tredje egenskapsfältet "hyv_equal" och "voim_equal" anger att datumen för godkännande och ikraftträdande är enhetliga samt eventuella brister i dem. I fälten kan förekomma:

1. "Godkdatum/ikrafttdatum saknas i både"
 - a. I både FDS- och kommunmaterialet saknas det datum som ska granskas för planens del.
2. "Godkdatum/ikrafttdatum saknas i FDS"
 - a. I både FDS-materialet saknas det datum som ska granskas för planens del.
3. "Godkdatum/ikrafttdatum saknas i kommun"
 - a. I både kommunmaterialet saknas det datum som ska granskas för planens del.
4. 1 - Datumen som granskas är desamma
5. 0 - Datumen som granskas skiljer sig från varandra

Datumen för fastställandet ingår inte i jämförelsen, eftersom de saknas helt i FDS-materialet. Som allmän princip är de egenskapsuppgifter som ingår i jämförelsen de enda som kunde jämföras i VOOKA-pilotprojektet.

6. PDF-länkningskonvertering

6.1 PDF-länkningstavla

Det är möjligt att länka PDF-bilagor i ett enhetligt format till planernas geodata om man vet vilken planindexbeteckning varje bilaga är kopplad till. I VOOKA-pilotprojektet saknades denna information nästan utan undantag i planernas geodatamaterial. Därför genomfördes identifieringen av planbilagor och motsvarande planindex som manuellt arbete.

För PDF-länkningen i ETL-verktyget sammanställdes en kopplingstavla (Tabell 5), där varje rad innehåller information om planens indexbeteckning (FDS, kommunens material eller båda) samt

planens dokumenttyp (t.ex. programmet för delaktighet och bedömning). Sammantaget var kopplingstavlans schema följande:

Tabell 5. PDF-länkningstavlans egenskapsdatafält och deras förklaringar.

Egenskapsuppgifter	Förklaring
Kommun	Kommunens officiella kommunkod.
Kommunens indexbeteckning	Planindexbeteckning för kommunens planmaterial i geodataformat.
FDS-indexbeteckning	Planindexbeteckning för FDS-materialet (planbeteckning_1).
Original filename	Det ursprungliga filnamnet på kommunens planbilaga.
Planslag	Koduppsättning för planslag på övre nivå.
Dokumenttyp	Koduppsättning för dokumenttyp, verbal.
Dokumenttyp2	Koduppsättning för dokumenttyp, numerisk.
Status	Koduppsättning som anger om planbilagan är giltig.
Gällande	Är bilagan giltig eller inte (boolean).
Multipage	Finns det fler än en sida i bilagan (boolean).

Koduppsättning för planslag:

- ak
- rak
- yk

Dokumenttyp -koduppsättning (verbal och numerisk):

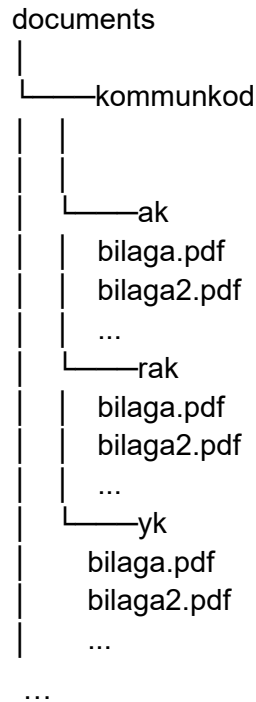
- 1 = plankarta (inkl. märkningar och bestämmelser)
- 2 = plankarta (inkl. inte märkningar och bestämmelser)
- 3 = märkningar och bestämmelser (separat)
- 4 = planbeskrivning
- 5 = program för deltagande och bedömning
- 6 = annat

Status-koduppsättning:

- ok
- inte ok

6.2 Katalogstruktur för planbilagor

Planbilagor från kommunerna sparades i en standardiserad katalogstruktur i resurshanteringen. Katalogstrukturens stomme hade formen:



ETL-verktyget utnyttjar katalogstrukturen i fråga för att namnge filer på nytt!

6.3 Nya namn på planbilagor

Principen för att ge bilagor nya namn är att skapa enhetliga bilagenamn. ETL-verktyget skapar standardiserade namn på bilagor i anslutningstabellen i det nya egenskapsfältet "new_name" enligt följande principer:

1. Bilagans namn beaktar den officiella kommunkoden
2. Bilagans namn beaktar plantypen enligt koduppsättningarna i datasystemet för den byggda miljön (Ryhti)
3. Bilagans namn beaktar handlingens art enligt koduppsättningarna i Ryhti
4. Bilagans namn beaktar planbeteckningen
5. Två eller flera bilagor får inte ha samma namn

Ett exempel på ett namn på en bilaga som ETL-verktyget producerar är "491-31-0304-13030-2.pdf", där:

- 491 är S:t Michels officiella kommunkod

- 31 är en detaljplan enligt Ryhtis koduppsättning för planslag
- 0304 hänvisar enligt Ryhti-handlingens art -koduppsättning till att bilagan täcker både plankartan (03) och bestämmelserna (04)
- 13030 är kommunens indexbeteckning
- 2 är ett glidande ordningsnummer (491-31-0304-13030-1.pdf finns redan, varmed två bilagor med samma namn har hittats för en gällande plan)

ETL-verktyget gör det möjligt att namnge bilagan på nytt antingen utifrån FDS-indexbeteckningen eller kommunens egen indexbeteckning.

OBS! Praxis för att ge planbilagor nya namn är endast ett förslag till namngivning av bilagor i Ryhti! Den standardiserade namngivningspraxisen kan ändras så länge Ryhti-genomförandeprojektet pågår!

6.4 Länkning av bilagor till geodatamaterial

ETL-verktyget länkar de filnamn som namngetts på nytt från länkningstabellen till geodatamaterialet på basis av plantyp, kommunkod samt planens indexkod (antingen FDS eller kommunens kod, beroende på vilkendera som bilagorna har namngetts på nytt). Planslag och kommunkod behövs eftersom:

- 1) Inom kommunen kan t.ex. stranddetaljplaner ha samma kommunens indexbeteckning som detaljplanerna
- 2) Olika kommuner kan ha exakt samma planbeteckning.

Varje bilaga finns efter körningen i det geodatamaterialet i ett eget (t.ex. i fältet planer för deltagande och bedömning "pdb").

6.5 PDF/A-konvertering

ETL-verktyget konverterar kommunernas bifogade PDF-filer automatiskt med Ghostscript-konverteraren till PDF/A-arkivformat. Konverteringen grundar sig i tillämpliga delar på den [Python-Ghostscript-konverterare](#) som är öppet tillgänglig vid GitHub. En konverterare som anpassats för VOOKA-projektets behov finns som [en egen version](#) på projektets GitHub-sidor.

Om det erhållna PDF-formatet har skapats med en version som inte kan fås i PDF/A-format, sparas det som sådant. Om vissa av kommunens PDF-filer av tekniska skäl som beror på kommunens material inte konverteras till PDF/A-format, sparas dessa i normalt PDF-format.

7. Implementering av plandatamodellen

I ETL-verktyget implementeras [plandatamodellen](#) genom att omvandla planmaterialet i GeoPandas GeoDataFrame-format till GeoJSON-format. För ETL-omvandlare anges som materialkälla antingen "kommun" eller "FDS" samt planbeteckningarna för båda

materialkällorna. Som en del av jämförelsen av materialet har man i punkt 4.2 skapat datafältet "refe_planbeteckning", som t.ex. i FDS-baserat material anger den kommunkod som motsvarar planen. Om motsvarande kod inte hittas (t.ex. om planobjektet endast finns i FDS:s material), lämnar ETL-omvandlaren uppgiften om kommunens indexbeteckning i GeoJSON tomt.

Med hjälp av ETL-verktyget är det dessutom möjligt att normalisera JSON-materialet enligt plandatamodellen tillbaka till geopackage-format så att materialet är lättare att läsa och granska för kommunerna i geodataprogram.

8. Behov av vidareutveckling

8.1 OGC API Features geografisk avgränsning

De geografiska gränserna för landskapet Södra Savolax är hårt kodade som bounding boxar i OGC API Features geer-skript. I fortsättningen ska bounding box inom andra geografiska områden ersättas med önskade koordinater.

8.2 Kommungränskollisioner på generalplaner

Korrigeringslogiken för kommungränskollision på generalplanerna fungerar bra, men inte fullständigt. Under VOOKA-projektet identifierades ett tydligt fall som ännu inte har besvarats med hjälp av logik (Bild 5). Planindexet blir "öppet" vid gränsen till det andra planindexet, varvid delgeometrin för difference som bildas i beräkningen inte filtreras bort med en arealjämförelse.

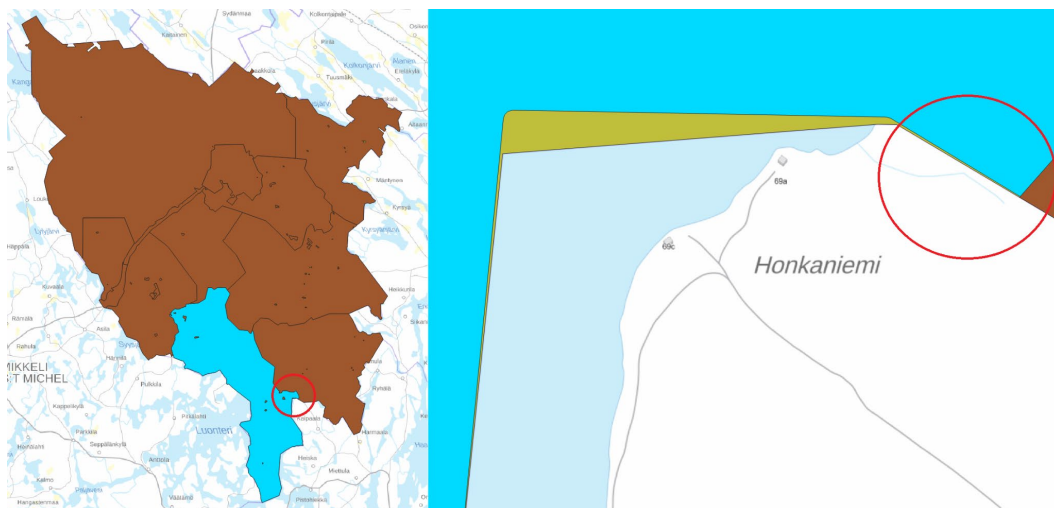


Bild 5. Exempel på en situation där korrigeringen av kommungränskollisionen i generalplansindexet inte fungerar fullständigt. Vid beräkningen av maskeringen av kommungränsen och planindexets difference blir indexet på gränsen till det andra planindexet "öppet" i förhållande till maskeringen, varvid skillnaderna inte filtreras bort med en jämförelse av arealen. Det blå området (planindexet) förenas inte med det gröna (fastighetskommungränsen), även om det borde.

Filtreringen av difference-geometrier som baserar sig på arealens storlek grundar sig på försök eftersom det finns många skillnader i de spatiala precisionerna i kommunmaterialet. För VOOKA-projektets del har varje kommuns fungerande gränsvärde testats separat och listats som referens för vidareutveckling. Även om kommunernas toleransvärden är enhetliga kan man inte blint lita på dem i olika geografiska områden, utan varje kommun ska hitta ett fungerande gränsvärde separat genom testning. Som idé till fortsatt utveckling för detta behov kunde man också utbilda AI/ML-funktioner för att bedöma det lämpligaste toleransvärdet för varje kommun.

8.3 Objektregisterbeteckningar från tomtuppgifterna

Lokaliseringsuppgifterna för fastighetsbeteckningarna finns förutom på poängnivå även på områdesnivå från LMV:s OGC API Features-gränssnitt som en del av tomternas omfattande lokaliseringsuppgifter. Att utnyttja den regionala referensnivån för att länka objektregisterbeteckningarna till planindexen är ett noggrannare sätt än att utnyttja poängnivån. I genomförandet av VOOKA-projektet kan en del av fastighetsbeteckningarna inte länkas till planindex som i verkligheten tangerar fastighetsbeteckningen, eftersom uppgifterna om fastighetsbeteckningens läge är bundna till endast en punkt i stället för till influensområdet.

Dessutom vill man i plandatamodellen för datasystemet för den byggda miljön i fråga om objektregisterbeteckningarna veta om de helt och hållet ingår i planen samt arealen för området som ingår i planen. Denna information kan inte fås endast genom att utnyttja poängnivån.

8.4 Jämförelse av fastighetstomt och planindex

ETL-verktyget som utvecklats inom VOOKA-projektet identifierar inte om fastighetsgränserna har bildats innan planen trätt i kraft eller efter det. Således kan den inte indikera om syftet med planindexets gräns är att sammanfalla med fastighetsgränsen eller inte.

8.5 DE-9IM-beräkning

Vid DE-9IM-beräkning beaktas inte om planen som motsvarar jämförelsen kan vara false positive. På grund av detta stiger till exempel arealernas relativa felprocent i värsta fall till tusentals och iou-procenten är minimal. Det vore bra att uppdatera ETL-omvandlaren med gränsvärden som filtrerar bort false positive -fall. Som motsvarande plan godkänns till exempel inte en plan med vilken arealernas relativa felprocent överstiger 100 %

8.6 Jämförelse av egenskapsuppgifter

I VOOKA-pilotprojektet är det möjligt att jämföra egenskapsuppgifterna endast för tre datafält (planslag, godkännandedatum och ikraftträdandedatum) på FDS-materialbasis. I andra geografiska områden kan egenskapsuppgifterna vara mer omfattande, varvid även andra datafält kan tas med i jämförelsen. Dessutom är det möjligt att utvidga den FDS-baserade jämförelseprincipen så att den blir kommunbaserad.