

LOPPURAPORTTI
25.4.2019

SEURAAVAN SUKUPOLVEN PALVELUALUSTAN KEHITTÄMINEN METSÄTIEDON JAKELUUN-
KONSEPTITESTAUS

Tiivistelmä

Hankkeessa rakennettiin ja testattiin metsätiedon palvelualustakonsepti, joka on tarkoitettu tehostamaan metsätiedon hallintaa ja helpottamaan sen hyödyntämistä sovelluksissa ja palveluissa. Sitä varten rakennettiin kokeilujärjestelmä MS Azure pilvialustalle, määriteltiin ForestJSON-kyselykieli ja rajapinta datakyselyjä varten sekä kehitettiin datojen esiprosessointi-, ajantasaistus- ja yhdistämismenetelmät. Lisäksi laadittiin ForestJSON Generator- sovellus alustan testausta ja erillisten datakyselyjen tekoa varten. Alustaan liitettiin 12 eri tyyppistä tietolähdettä sen toiminnallisuuden rakentamista ja testausta varten. Sidosryhmien asiantuntijat testasivat palvelualustaa, minkä pohjalta koostettiin kehittämissuunnitelmia alustan tuotteistamisen tueksi. Lisäksi laadittiin ehdotus alustan organisointi- ja rahoitusmallista.

Projekti vahvisti käsitystä, että palvelualustan toiminnallisuudella voidaan tehostaa metsätiedon hallintaa ja hyödyntämistä. Konsepti tukee erityisesti data-analyysiin perustuvien kehittyneiden päätöksisovellusten laadintaa, hilatasoisen datan käsittelyä ja sen ajantasaistusta sekä uusien julkisten ja yksityisten tietolähteiden käyttöönottoa ja yhdistämistä. Se tarjoaa kiinnostavia mahdollisuuksia myös tutkimusaineistojen hakuun ja yhdistämiseen.

Alusta on perustelluinta sijoittaa Suomen metsäkeskuksen metsävaratietojärjestelmän yhteyteen, jolloin metsäkeskus olisi palvelun omistaja. Palvelualustan toiminnallisuus tehostaa metsäkeskuksen perustehtävää eli ajantasaisen metsävaratiedon tuottamista, päivitystä ja jakelua. Palvelualustan suorituskyvyn kannalta avaintietovarastot, esimerkiksi ajantasaisena pidettävät hilatiedot, kannattaa sijoittaa alustan sisälle.

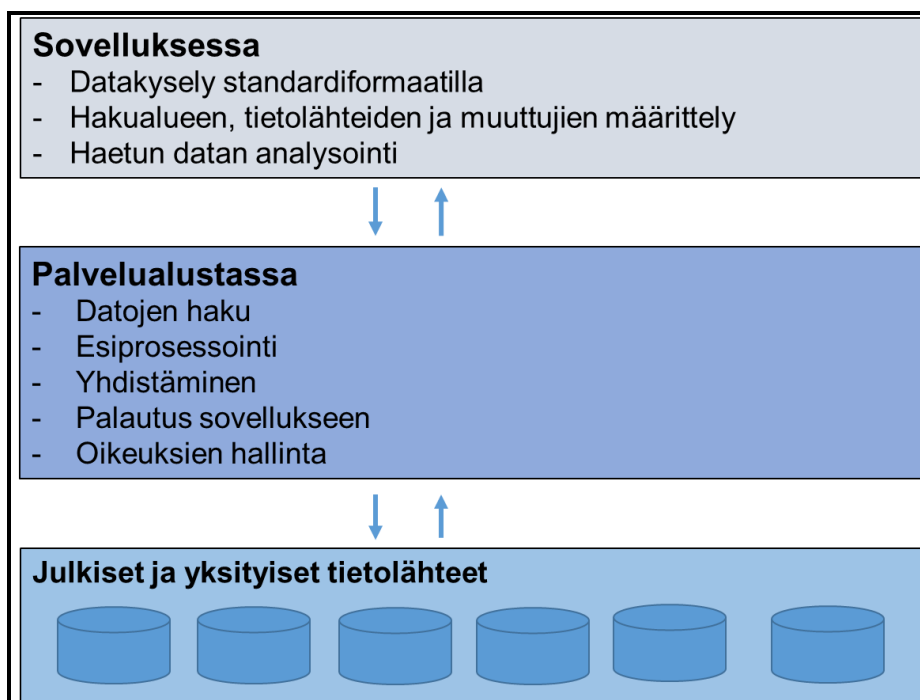
Pilvialusta tarjoaa tehokkaan ja skaalautuvan teknisen ympäristön palvelulle. Alustakonseptissa on paljon metsäalan innovaatioita ja uutta liiketoimintaa vauhdittavaa potentiaalia. Tulevaisuudessa se voisi olla datojen jakelualustaa laaja-alaisempi datojen, mallien, sovellusten ja palvelujen markkinapaikka.

Alustakonseptin tuotteistamisen ja soveltamisen kannalta eri tietolähteiden rajapintojen parantaminen on tärkeää. Lisäksi datojen luotettavuusestimaatit ja esimerkiksi joukkoistetusti kerätyn tiedon laadunhallinta korostuvat jatkossa.

Hilatiedon merkitys on kasvamassa ja sen tietosisällön kehittäminen on olennaista. Tärkeitä kehittämiskohteita ovat etenkin puulajeittaiset puustotunnukset, maaperätiedot sekä ympäristöasioiden hallintaa tukevat tietokokonaisuudet.

1. Tutkimuksen tausta ja tavoitteet

Metsätiedon palvelualustan tarkoituksena on välittää ja yhdistää dataa sovelluksille (kuva 1). Se tarjoaa palveluja datojen fuusiointiin, jakeluun ja käyttöoikeuksien hallintaan. Pyrkimyksenä on madaltaa kynnystä kehittää uusia, metsätietoa entistä monipuolisemmin hyödyntäviä sovelluksia ja palveluita. Tavoitteena on sovellus- ja palvelukehityksen kustannustehokkuuden paraneminen sekä joustavuuden lisääminen, kun uusia tietotarpeita ja -lähteitä ilmaantuu (Rajala ja Ritala 2016). Erityisesti pyritään edistämään datalähtöisten päätöstukijärjestelmien kehittämistä eri toimijatasoille.



Kuva 1. Metsätiedon palvelualustan tarkoituksena on tehostaa eri tietolähteiden käyttöä sovelluksissa.

Metsätiedon palvelualustasta tehtiin vuonna 2016 esitutkimus (Hämäläinen ym. 2017), ja tämä konseptitestaus käynnistettiin sen jatkoksi. Tämän hankkeen tavoitteena oli määrittellä ja rakentaa kokeilujärjestelmä ja testata sen avulla alustakonseptin toiminta käytännön datalähteiden ja niitä hyödyntävien esimerkkisovellusten avulla, tarkastella alustan organisointi- ja rahoitusvaihtoehtoja sekä tehdä ehdotus alustan tuotteistamisesta.

2. Tutkimusosapuolet ja yhteistyö

Hanke toteutettiin Metsätehon ja Maa- ja metsätalousministeriön yhteishankkeena. Metsäteho vastasi hankkeen johtamisesta, järjestelmän toiminnallisesta määrittelystä, tietoaineistojen kokoamisesta, järjestelmän testauksesta, organisointi- ja rahoitusvaihtoehtojen tarkastelusta sekä raportoinnista. Maa- ja metsätalousministeriö vastasi Metsäkeskus-vaihtoehtoa koskevan organisointi- ja rahoitusehdotuksen juridisista selvityksistä. Tampereen teknilliseltä yliopistolta ostettiin asiantuntija-apua datojen haun, esiprosessoinnin ja yhdistämisen menetelmien kehitykseen sekä datahakuihin tarvittavan kyselykielen määrittelyyn ja kyselykoodia tuottavan apuvälineen kehittämiseen. CGI vastasi kokeilujärjestelmän teknisestä määrittelystä sekä sen rakentamisesta. Metsäkeskus osallistui hankkeeseen asiantuntijajäsenenä. Yhteistyö toimi hyvin ja eri toimijoiden osaaminen ja roolitus muodostivat tuloksellisen kokonaisuuden.

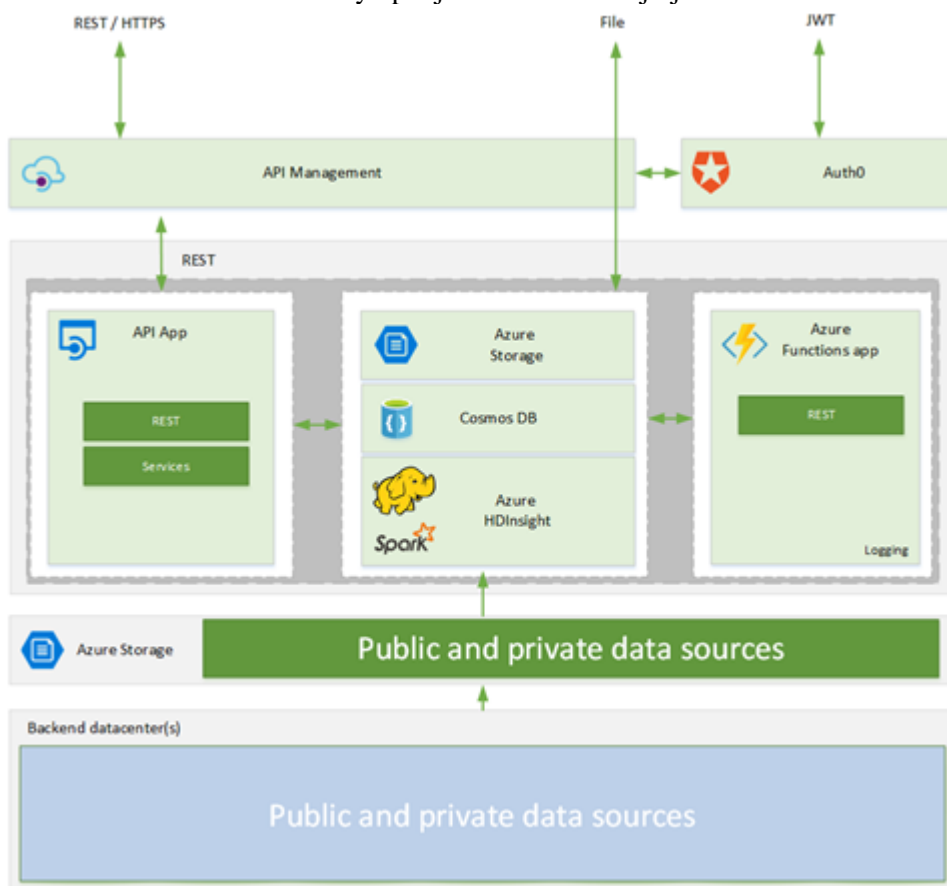
Projektityöhön osallistuivat Jarmo Hämäläinen, Tapio Räsänen, Kirsi Rieki ja Juha-Antti Sorsa Metsäteho Oy:stä, Risto Ritala, David Hästbacka ja Teijo Juntunen Tampereen yliopistosta, Juha Peltoniemi, Heikki Vorne, Matti Koljonen ja Jukka Mäkelä CGI Suomi Oy:stä sekä Niina Riissanen Maa- ja metsätalousministeriöstä.

Projektin ohjausryhmänä toimi Metsätieto ja sähköiset palvelut –hankkeen ohjausryhmä. Projektin operatiivisessa johtoryhmässä oli edustajat päärahoittajatahoista eli Maa- ja metsätalousministeriöstä, Metsätehosta, Metsä Groupista, Metsähallituksesta, Stora Ensosta ja UPM:stä sekä asiantuntijajäsen Suomen metsäkeskuksesta.

3. Tutkimuksen tulokset

3.1 Kokeilujärjestelmän rakenne

Projektin ensivaiheessa määriteltiin kokeilujärjestelmän toiminnalliset ja tekniset vaatimukset. Vaatimusmäärittelyssä tuotetut dokumentit on esitetty liiteluettelossa ja ne on saatavissa Metsätehon verkkosivuilta. Vaatimusmäärittelyn pohjalta rakennettiin järjestelmä Microsoft Azure -pilvialustalle (kuva 2).



Kuva 2. Kokeilujärjestelmän tekninen rakenne MS Azure-alustalla. Kuva: CGI

3.2 Datakyselyjen toteutus

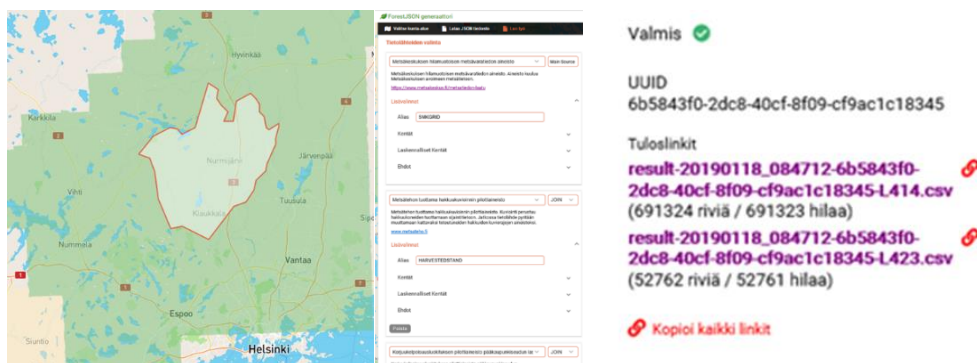
Palvelualustan käyttö perustuu dataa käyttävään sovellukseen ohjelmoitavaan, alustalle lähetettävään kyselyyn, jonka perusteella alusta hakee tarvittavat datat sovellusta varten. Sitä varten kehitettiin ja kuvattiin standardimuotoinen ForestJSON –kyselykieli, joka pohjautuu paikkatietoalalla käytettävään GeoJSON-formaattiin. Kyselyssä määritellään datahaun sisältö seuraavasti:

- Hakualue
- Tietolähteet
- Tietolähteiden yhdistämistapa
- Tietolähteiden kentät (muuttujat)
- Haun ehdot (suodatus)
- Mahdolliset laskennalliset kentät

Alustan testauksen helpottamiseksi laadittiin ForestJSON Generator -sovellus, jolla datakyselyt voidaan tehdä rajaamalla alue kartalle graafisessa käyttöliittymässä tai valitsemalla hakualueeseen kunnat kuntalialta ja tekemällä em. valinnat alasetoivalikoilla (kuva 3). Sovellus tuottaa automaattisesti ForestJSON -ohjelmakoodin,

joka voidaan lähettää alustalle tai käyttää koodin laadinnan ja muokkauksen pohjana. Sovellus voi olla tarpeellinen myös tuotteistetussa palvelualustassa etenkin sellaisia datan hyödyntäjiä varten, joilla ei ole käytössään varsinaista loppukäyttösovellusta. Muut sovellukset voivat kuitenkin käyttää kaikkia palvelualusta toiminnallisuuksia ohjelmallisesti ilman Generator –sovellusta.

Palvelualusta palauttaa kysytyt datat csv-muodossa loppukäyttösovellukselle niiden analysointia ja muuta jatkokäsittelyä varten. Kyselyn tulos voidaan myös palauttaa hakualueelle aggregoituna, jolloin hakuehdot täyttävistä hiloista lasketaan halutuille muuttujille koko aluetta koskevat arvot. Tällainen palaute voidaan antaa csv- tai GeoPackage -formaattissa.



Kuva 3. ForestJSON Generator -sovelluksella voidaan tehdä datakyselyjä. Kyselyalue rajataan karttakäyttöliittymässä, datalähteet ja hakuehdot valitaan alusvetovalikoilla ja kysely lähetetään alustalle. Alusta palauttaa kyselyn tulokset tiedostoina.

3.3 Testatut aineistot

Kokeilujärjestelmään liitettiin yhteensä 12 eri tyyppistä aineistoa, joiden avulla rakennettiin ja testattiin alustan toiminnallisuus (taulukko 1).

Taulukko 1. Testialustaan liitetyt paikkatietoaineistot.

Aineisto	Tuottaja	Luonne
Hilamuotoinen metsävaratieto	Metsäkeskus	Julkinen
Hilamuotoinen metsävaratieto (MVMI)	Luke	Julkinen
Metsävarakuviot	Metsäkeskus	Julkinen
Kemera-aineistot	Metsäkeskus	Julkinen
Metsänkäyttöilmoitukset	Metsäkeskus	Julkinen
Korjuukelpoisuusluokitus	Arbonaut & Metsäkeskus	Julkinen
Maastotietokanta	Maanmittauslaitos	Julkinen
Suomen kuntajako 2017	Maanmittauslaitos	Julkinen
Säädata (rajapintapalvelusta)	Ilmatieteen laitos	Julkinen (ilmainen palvelu)
Metsäpaloindeksi (rajapintapalvelusta)	Ilmatieteen laitos	Julkinen (maksullinen palvelu)
Hakkuukuviot (hakkuukonedata)	Metsäteho	Yksityinen (koeaineisto)
Muutostulkinta-alueet (satelliittidata)	Terramonitor	Yksityinen (koeaineisto)

Näiden lisäksi kokeiltiin muita yksityisiä aineistoja, joita voidaan joustavasti yhdistää alustaan kertaluonteista käyttöä varten.

Suomen metsäkeskuksen hila-aineistoa käsitellään alustassa keskeisenä tietolähteenä, koska se tulee olemaan tulevaisuuden täsmämetsätalouden informaation hallinnan perusyksikkö. Alustakonseptissa on sen takia paneuduttu erityisesti tehokkaiden menetelmien kehittämiseen hiladatan käsittelyyn, ajantasaistukseen ja jakeluun. Lisäksi hila on luontevin datafuusion perusyksikkö, koska sillä on yksiselitteinen sijainti ja dimensiot.

3.4 Aineistojen esiprosessointi, yhdistäminen ja tulostiedostojen luonti

Projektissa kehitettiin menetelmät eri tyyppisten aineistojen liittämiseen alustaan, niiden esiprosessointiin ja yhdistämiseen sekä tulostiedostojen tuottamiseen. Projektin aikana havaittiin, että alustan suorituskyvyn kannalta on hyvä jakaa käsiteltävät aineistot kiinteään, ennalta määrättyyn aluejakoon, jota noudattaen esiprosessointi, yhdistäminen ja tulosten tuottaminen tehdään. Tässä projektissa käytettiin Maanmittauslaitoksen UTM-100 - karttalehtijakoa (48 km x 96 km). Käytännössä tämä tarkoittaa sitä, että myös kyselyn hakualue jaetaan karttalehtiin, ja jokainen karttalehti käsitellään erillisenä alustassa kyselyä suoritettaessa.

Esiprosessointi

Palvelualustan spatiaalinen tiedonhallintayksikkö on valittu siten, että se vastaa Suomen metsäkeskuksen hilamuotoisen metsävaratiedon esittämiseen käytettyä hilajakoa. Metsäkeskuksen hilaruutu edustaa 16 m x 16 m kiinteätä, neliön muotoista aluerajausta, joka on sidottu Suomessa yleisesti käytettyyn koordinaatistoon ETRS89 / TM35FIN(E,N) (EPSG:3067) ja on esitetty ns. vektorimuodossa. Palvelualustalla hiloja käsitellään kuitenkin pistemäisinä kohteina alueiden sijaan suorituskykyistä. Palvelualustan hilalla on yksilöllinen tunniste ja hilaruudun keskipistettä vastaavat koordinaatit, sekä joukko tietokenttiä.

Testialustaan liitetyt hila- ja vektorimuotoiset paikkatietoaineistot (taulukko 1) vietiin alustalle esiprosessoimalla ne alustalle ennen kyselyjen tekoa. Esiprosessointiin päädyttiin, koska testiprojektissa havaittiin, että suurten ja harvoin päivittyvien tietomassojen haku rajapinnoista palvelualustaan toistuvasti ei ole vasteaikojen kannalta järkevää. Esiprosessoinnin yhteydessä tietolähteen sisältö muutetaan taulukoksi, joka yhdistää alustan hilapisteen tunnisteeseen ja tietolähteestä saadun tietosisällön kyseiselle hilalle.

Osa testialustan aineistoista on rajapinnasta saatavia, dynaamisia tietoja, kuten esimerkiksi Ilmatieteen laitoksen tarjoamat aineistot. Kyselykohtainen rajapintahaku todettiin projektissa tällaisille aineistoille parhaaksi vaihtoehdoksi, koska päivitystahti on nopea ja siirrettävä tietomassa pieni. Rajapinnan kautta alustaan liitetyt tietolähteet ei esiprosessoida palvelualustaan, vaan niille määritellään säännöt, kuinka tietolähteen sisällöt viedään hiloille osana kyselyvastauksen muodostamista. Esimerkiksi tässä testauksessa hilalle asetettiin lähimmän säähavainnon tai -ennusteen arvo. Harvaresoluutioisilla pistemäisillä havainnoilla on myös mahdollista interpoloida hiloille arvoja havaintopisteiden välisille alueille (Ritala & Juntunen 2019).

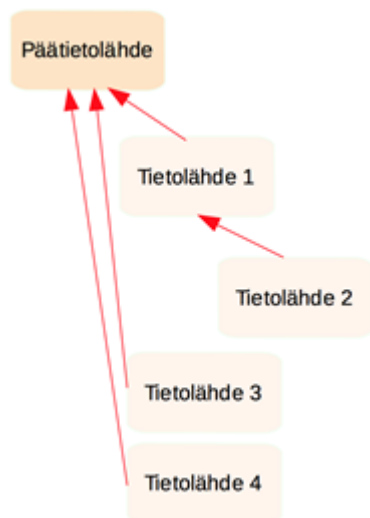
Palvelualustan toiminnan perustuessa hilojen käsittelyyn, oli Metsäkeskuksen hilan ja vastaavassa esitysmuodossa olevien hilatietolähteiden vienti alustaan periaatteeltaan suoraviivaista. Muiden hilamuotoisten, mutta eri resoluutiolla esitettyjen tietolähteiden esiprosessointi voidaan tehdä resoluutiosta riippuen joko keskiarvoistamalla arvot hilalle, tai asettamalla lähin arvo alustan hilan arvoksi (tarkempi kuvaus Ritala & Juntunen 2019).

Vektorimuotoiset aineistot esiprosessoitiin määrittelemällä säännöt, kuinka tiedot viedään hiloille. Alumuotoisilla tietolähteillä esiprosessoinnissa asetettiin hilakenttiin ne arvot, jotka löytyvät hilan keskipisteen kohdalla olevien aluerajausten tietosisällöistä. Sen sijaan viivamaiset ja pistemäiset tiedot, joita testiprojektissa sisälsi Maanmittauslaitoksen maastotietokanta, asetettiin niille hiloille, joilla ne sijaitsevat. Testissä oli mukana myös yksi vektorimuotoinen relaatiotietokanta, Metsäkeskuksen metsävarakuviot, jonka taulurakenne purettiin auki esiprosessoinnissa.

Yhdistäminen

Palvelualustan päätoiminnallisuus on tarjota monipuolisia vaihtoehtoja ja mahdollisuuksia yhdistää eri tietolähteiden sisältöjä toisiinsa hilasijaintikohtaisesti (ns. spatiaalinen yhdistäminen, engl. spatial join). Tämän vuoksi kyselyiden teko ja kyselykieli ForestJSON suunniteltiin niin, että käyttäjä voi itse valita haluamansa yhdistelmän ja/tai rajauksen eri aineistojen välillä. Kyselyä tehdessään käyttäjä valitsee yhden palvelualustan tietolähteistä ns. päätietolähteeksi, johon muiden tietolähteiden tietosisältöjä yhdistetään. Päätietolähde on yleensä Metsäkeskuksen hila-aineisto, mutta se voi olla yhtä hyvin joku muu tietolähde. Tietolähteitä voidaan

yhdistää myös hierarkkisesti (kuva 4). Kysely suoritetaan palvelualustassa edeten hierarkkisesti tietolähteiden järjestyksessä.



Kuva 4. Esimerkki tietolähteiden yhdistämisen hierarkiasta. (ForestJSON ohje, CGI)

Käyttäjä määrittelee kyselyyn tietolähdekohtaisesti:

- aluerajauksen, jolta tietosisällöt halutaan (mm. ForestJSON Generator -sovelluksen käyttöliittymän avulla tai käyttämällä hilatunnisteita)
- tietolähteen yhdistämisvaihtoehdon päätielähteeseen/ylemmän hierarkiatason tietolähteeseen
- vastauksen koostamiseen käytettävät ehdot kenttien arvoille
- vastauksessa palautettavat kentät
- vastaukseen tuotettavat laskennalliset kentät tietolähteen kentistä.

Tietolähteiden yhdistäminen tapahtuu palvelualustassa siten, että kyselyn annettu aluerajaus muutetaan ensin hiloiksi samaan tapaan kuin tietolähteiden esiprosessoinnissa. Tämän jälkeen tiedot yhdistetään hilatunnisteiden avulla eri tietolähteistä.

Yhdistämisvaihtoehtoja tietolähteiden välillä on neljä erilaista. Voidaan sisällyttää toinen tietolähde kokonaan, ja toisesta vain päällekkäiset osat, sekä sama toisinpäin. Voidaan sisällyttää molempien tietolähteiden kaikki arvot kyselyalueelta. Neljäs vaihtoehto on, että liitettävä tietolähde rajaa toisen tietolähteen hiloja pois vastauksesta.

Tietolähteiden kenttien arvoille voidaan antaa ehtoja käyttäen loogisia vertailuoperaattoreita, jotka määrittävät kuinka tietolähteen tietosisällöllä rajataan yhdistämistä. Kyselyvastauksessa palautettavien kenttien joukko voi olla eri kuin millä yhdistäminen on tehty, ja palautukseen voidaan määritellä ylimääräisiä, laskennallisia kenttiä, jotka lasketaan tietolähteen kenttien perusteella. Kyselyn ylimmällä hierarkiatasolla on vielä mahdollista lisätä ehtoja ja laskennallisia kenttiä, jolloin kaikki alempien tasojen tietolähdekohtaiset vastaukset ovat käytettävissä ehtojen määrittelyyn.

Alustaan on lisäksi toteutettu erilliset funktiot IFNULL ja UPDATE. Funktio IFNULL korvaa tietyn kentän tyhjät arvot jollakin toisen kentän arvolla, tai toisen tietolähteen kentän arvolla, riippuen hierarkiatasosta. Kasvunlaskentafunktio UPDATE on käytössä Metsäkeskuksen hila-aineistolle puuston määrän ajantasaistamiseen, ja se perustuu Luonnonvarakeskuksen Motti-malliin. UPDATE tuottaa erillisinä lisäkenttinä kasvunlaskennan jälkeiset puustotunnukset kyseiselle hilalle.

Tulostiedostojen luonti

Palvelualusta palauttaa tulostiedostot UTM-100 -karttalehdittäin, kun koko kysely on valmistunut. Vastaukset tulevat csv-tiedostona, jossa yhdelle hilalle tulee vastaukseen yksi rivi. Se sisältää hilan tunnisteen ja keskipisteen

koordinaatit (EPSG:3067), sekä kaikki kyselyssä olevien tietolähteiden kentät sekä mahdolliset laskennalliset - ym. lisäkentät koostettuna yhteen. Palvelualustan palauttavat csv-tiedostot ovat sellaisenaan avattavissa paikkatieto-ohjelmilla, koska hilakohtaiset koordinaatit ovat mukana aineistossa.

Joillain testatuilla tietolähteillä, kuten FMI:n metsäpaloindeksilla ja metsäkeskuksen metsävarakuvioilla, kyselyvastaukset muodostavat relaatiotietokanta-tyyppisen rakenteen. Tällöin palvelualusta palauttaa erillisiä lisätiedostoja csv-formaatissa, joihin eri relaatiotietokannan tauluja vastaavat tiedot on tallennettu siten, että varsinainen hilakohtainen vastaustiedosto sisältää yhtenä kenttänä viiteavaimen toisen tiedoston sisältöihin. Mikäli kysely sisältää hilamuotoiset metsävaratiedot pää tietolähteenä ja tulosten aggregoinnin kyselyn alueelle, voi käyttäjä valita relaatiotietokanta-tyyppisen palautuksen GeoPackage-tiedostoformaattissa, jolloin tiedosto sisältää alkuperäisen kyselyn aluerajauksen ja aggregoidut kenttien arvot. Myös kaikki nämä tiedostot tulevat karttalehdittäin.

3.5 Esimerkkejä käyttötapauksista

Projektissa kokeiltiin palvelualustaa seuraaviin käyttötarkoituksiin:

- Hilatietojen päivitys hakkuutiedoilla
- Hilan puustotietojen päivitys Motti-kasvumalleilla
- Kuviotietojen aggregointi hilatiedoista
- Hakkuumahdollisuuksien analysointi eri tietolähteistä

Hilatietojen ajantasaistusta hakkuukonedatan pohjalta testattiin siten, että metsäkeskuksen hiladataan fuusioitiin hakkuukonedatasta muodostetut hakatun alueen rajat, hakkuun ajankohta ja hakkuutapatie. Nämä tiedot liitettiin alkuperäisen hilatiedon lisäkentiksi. Harvennettujen hilojen puustotietoja ei tässä vaiheessa päivitetty laskennallisesti. Siinä voitaisiin periaatteessa soveltaa lähtöpuustoon ja harvennushiloihin perustuvia laskennallisia menetelmiä. Kehitysvaiheessa olevat jäävän puuston automaattiset mittausmenetelmät antanevat siihen tulevaisuudessa tarkempia mahdollisuuksia.

Hilan puustotietojen päivitystä varten Luken Motti-kasvumallikirjaston Linux-versio implementoitiin palvelualustaan. Alustaan toteutettiin funktio hilojen kasvunlaskentaa varten ja testattiin funktion toiminnallisuus. Samalla koottiin havaintoja mallien suorituskyvystä sekä niiden tuottamista tuloksista. Pilvialusta tarjoaa periaatteessa tehokkaan ympäristön hilatasoiselle kasvun laskennalle, mutta Motti-kasvumalleja on tarpeen kehittää, jotta ne pystyvät paremmin hyödyntämään alustan laskentakapasiteetin. Kasvumallien tuottamien hilatason tulosten tarkempaan analysointiin ei ollut tässä hankkeessa mahdollisuuksia. Lähtöaineistot ja kasvatuksen tulokset toimitetaan Lukelle mallien jatkokehitystä varten.

Kuviotietojen aggregointi hilatiedoista voidaan nähdä alustan lisäarvopalveluna, joka edistää hiladatan hyödyntämistä. Sitä varten kehitettiin palvelualustalle algoritmit, joilla halutuille polygoneille lasketaan niihin kuuluvien kyselyn ehtojen täyttävien hilojen attribuuteista sekä puulajikohtaiset että koko puustoa koskevat yhdistelmätulokset. Attribuuttien aggregointitavat (moodi, keskiarvo, summa) määritettiin tietolähdekohtaisesti. Aggregoidut tiedot tulostetaan csv- tai GeoPackage-formaatissa. Jälkimmäisessä formaatissa tiedot voidaan viedä metsäkeskuksen kuviotietokantaan asianomaisiin tietotauluihin.

Alustan loppukäyttöperiaatetta testattiin Metsätehon pilottisovelluksella, joka on tarkoitettu puukauppa-kohteiden hakuun ja hakkuumahdollisuuksien laskentaan. Sovellus rakennettiin palvelualustan ulkopuolisena erillishankkeena yhteistyössä CGI:n kanssa. Sovellus käyttää aineistoina:

- Suomen metsäkeskuksen hila- ja kuviomuotoista metsävaratietoa,
- Suomen metsäkeskuksen korjuukelpoisuusluokitusta,
- Luken MVMI-tietoa (hila),
- Metsätehon hakkuukuviointiaineistoa sekä
- Maanmittauslaitoksen maastotietokantaa

Sovelluksen käyttäjä rajaa liittymässä tarkasteltavan alueen sekä valitsee hakuehdot sekä tulosten laskenta- ja esittämisperiaatteet. Palvelualusta hakee ja yhdistää datat määrittelyjen pohjalta. Sovellus laskee ja esittää hakuumahdollisuudet graafeina ja tuottaa myös tiedostot jatkoanalyysijä varten.

3.6 Käyttäjätestaukset ja niiden palaute

Projektiryhmän testikäytön lisäksi palvelualusta ja hakuumahdollisuussovellus avattiin sidosryhmille testattavaksi 20.12.2018 – 31.3.2019 väliseksi ajaksi. Testaajille toimitettiin testausohje, avattiin linkki kokeilujärjestelmään, järjestettiin Skype-opastustilaisuus ja neuvottiin tarvittaessa järjestelmän käytössä. Kokeilukäyttäjiltä koottiin testauspalaute web-kyselyllä.

Yhteensä 12 henkilöä viidestä organisaatiosta kokeili järjestelmää. Testiajoja tehtiin viidestä muutamaa kymmeneen ajoon testaajaa kohti. Ajot olivat yleensä datakyselyjä erikokoisille alueille ja eri tietolähteitä yhdistäen. Ajojen kesto vaihteli yhdestä 10:een minuuttiin. Yksi käyttäjä kokeili myös datan vientiä karttatasoksi organisaation omaan suunnittelujärjestelmään.

Käyttäjäpalautteen keskiarvo oli 3.5 asteikolla 1 – 5, jossa 5 oli paras arvosana. Yhteenvedo testaaajien havainnoista ja kehittämisehdotuksista oli seuraava:

Palvelualustalla nähtiin yleisesti olevan monenlaisia hyödyntämismahdollisuuksia vastaajien organisaatiossa, mutta käyttöalueiden arviointi on vielä alustavaa. Käyttömahdollisuuksia mainittiin olevan mm. seuraaviin tarpeisiin:

- leimikon tai korjuulohkon tarkennetun tiedon muodostus hakemalla hilatietoja ja yhdistämällä niitä muihin tietolähteisiin sekä tiedot ajantasaistamalla
- organisaatioiden omien metsävara-aineistojen ajantasaisuuden hallinta ja vertailut
- luonnonsuojelun alueiden metsävaratietojen ajantasaistus
- hakuumahdollisuuksien haku eri kriteereillä
- korjuuolosuhteiden reaaliaikainen ennustaminen
- laajat tietohaut yleisesti erilaisilla suodatuksilla
- tietojen yhdistäminen ja jalostaminen uudeksi korkeamman tason tiedoksi, esim. mahdollisimman tarkan tiedon tuottaminen dynaamisesti
- sovelluskehityksessä ja sovellusten ylläpidossa integrointityön vähentäminen hankkimalla kaikki tarvittava data yhden rajapinnan kautta.

Tietolähteinä palvelualustalla katsottiin tarvittavan lähtökohtaisesti niitä julkisin varoin tuotettavia paikkatietoaineistoja, joita voidaan käyttää metsävaratietojen tuottamiseen, ajantasaistukseen, hakuumahdollisuuksien arviointiin (mm. metsän käytön rajoitteet ja kaavatiedot) sekä erilaisia olosuhteita kuvaavia aineistoja. Aineistojen saatavuus ylipäättään koettiin tärkeäksi: palvelualustan nähtiin voivan toimia kanavana hankkia aineistoja, joihin yksittäisillä organisaatioilla ei ole pääsyä tai se on hankalaa.

Testauksessa ei ollut vielä mahdollisuutta palauttaa kyselyn tuloksia kuviomuotoisina aggregoituina tietoina, johon useampikin vastaaja toivoi muutosta. ForestJSON -kyselykieleen testaajat eivät nähneet tarvittavan laajennuksia tai uusia ominaisuuksia. Testauskäyttö oli lähinnä kyselyjen muodostamista käyttöliittymän avulla, jolloin mahdolliset sovelluskehittäjien erityistarpeet eivät tulleet testauksessa esille muutoin kuin, että käytettävistä funktioista olisi tarvittu enemmän tietoa. ForestJSON -dokumentaation arvioitiin olevan riittävää.

Kehittämistarpeina ehdotettiin mm. seuraavia:

- kuviomuotoisen palautetiedon muodostaminen ja käsittely
- vektorimuotoisten aineistojen käsittely alustassa (ei konvertointia hilamuotoon)
- puustotietojen nopea kasvunlaskenta tai aineistojen päivitys alustalla valmiiksi nykyhetkeen
- runkolukusarjojen laskenta hilatietoihin mukaan
- mahdollisuus hakea vain edellisen haun jälkeen muuttuneet tiedot (tietolähteiden tietosisältömuutosten hallinta)
- parannuksia ForestJSON Generaattorin käyttöliittymään, mm. aineistojen näkyvyys kartalla, kyselyajojen ja vastaustiedostojen hallinta, tietokenttien valinta datan olemassaolon mukaan.

3.7 Testialustan suorituskyky- ja kustannustarkastelut

Kokeilualustan suorituskykyä ja kustannuksia tarkasteltiin tekemällä eri laajuisia datakyselyjä (taulukko 2). Kyselyalueiden pinta-ala vaihteli 4.3:sta hehtaarista 13.5 miljoonaan hehtaariin ja tietolähteiden määrä kolmesta viiteen hila- tai polygonitasoiseen tietolähteeseen. Kyselyt tehtiin pilvialustan peruskokoonpanolla eli käytössä oli yksi palvelinkone, jonka ajankäytön ja tuntihinnan perusteella laskettiin kyselyn suorat kustannukset. Kyselyt tehtiin siten, että prosessoitavana oli vain yksi kysely kerrallaan.

Kyselyn kesto riippui odotetusti sen laajuudesta ja vastausaika vaihteli puolestatoista minuutista vajaaseen neljään tuntiin. Palvelualustan käynnistyminen vie minimissään minuutin, joten sitä nopeammin ei vastausta saatu pieniinkään kyselyihin. Kustannusten vaihteluväli oli kolmesta sentistä 4.8:aan euroon. Taulukosta voi nähdä, että pinta-alan lisääntyminen kuusinkertaiseksi suunnilleen kolminkertaisti ajanmenekin ja kustannukset. Tuloksista saa suuntaa antavan käsityksen alustan suorituskyvystä. Todellinen tuotteistetun palvelun tehokkuus riippuu pilvialustan kokoonpanosta ja yhtäaikaisten kyselyjen määrästä.

Taulukko 2. Testialustan suorituskyky ja kustannukset eri laajuisilla datakyselyillä.

Alue	Pinta-ala, ha	Hiloja, kpl	Datat	Vastausaika	Kustannukset, €/kysely
”Työmaa”	4,3	167	hila, mvmi, kunta	1 min 30 s	0,03
”Kunta”	93 400	3,6 milj.	hila, mvmi, kunta	6 min	0,12
”Maakunta”	542 850	21,2 milj.	hila, mvmi, kunta	17 min	0,36
”Kunta”	93 400	3,6 milj.	hila, mvmi, kunta, kklk, htapa	24 min	0,51
”Maakunta”	542 850	21,2 milj.	hila, mvmi, kunta, kklk, htapa	1h 13 min	1,53
”Suuralue”	13,5 milj.	527 milj.	hila, mvmi, kunta	3h 50 min	4,83

”Kunta” = Vihti + Nurmijärvi
 ”Maakunta” = Espoo, Hausjärvi, Vantaa, Hyvinkää, Järvenpää, Kerava, Kirkkonummi, Kärkölä, Mäntsälä, Nurmijärvi, Riihimäki, Sipoo, Tuusula, Vihti
 ”Suuralue” = Noin 1/3 Suomen pinta-alasta, Etelä-Suomi

3.8 Organisointi ja rahoitusmalli

Projektissa valmisteltiin ehdotus alustan organisoinnista ja rahoituksesta. Ehdotuksena on, että palvelualusta tuotteistetaan Suomen metsäkeskuksen metsävaratietojärjestelmän yhteyteen (kuva 5) ja metsäkeskus on palvelun omistaja seuraavin perustein:

- Palvelualustan toiminnallisuus tehostaa metsäkeskuksen perustehtävää eli ajantasaisen metsävaratiedon tuottamista, päivytystä ja jakelua. Näin se pystyy vastaamaan entistä tehokkaammin asiakatarpeisiin.
- Keskeisimmät julkiset metsätiedon varastot ovat metsäkeskuksen hallussa ja ylläpitovastuulla.
- Palvelualustan suorituskyvyn kannalta avaintietovarastot kannattaa sijoittaa alustan sisälle – eikä rajapintojen taakse – mikä on helpoimmin järjestettävissä, kun metsäkeskus on palvelualustan omistaja.
- Näköpiirissä oleva operatiivisen toteutustiedon käyttö metsäkeskuksen metsävaratiedon ajantasaistuksessa tulee heijastumaan päivitysprosessin vaatimuksiin. Tehokkaaseen ja skaalautuvaan pilviratkaisuun perustuvan alusta on omiaan helpottamaan ajantasaistuksen toteutusta.

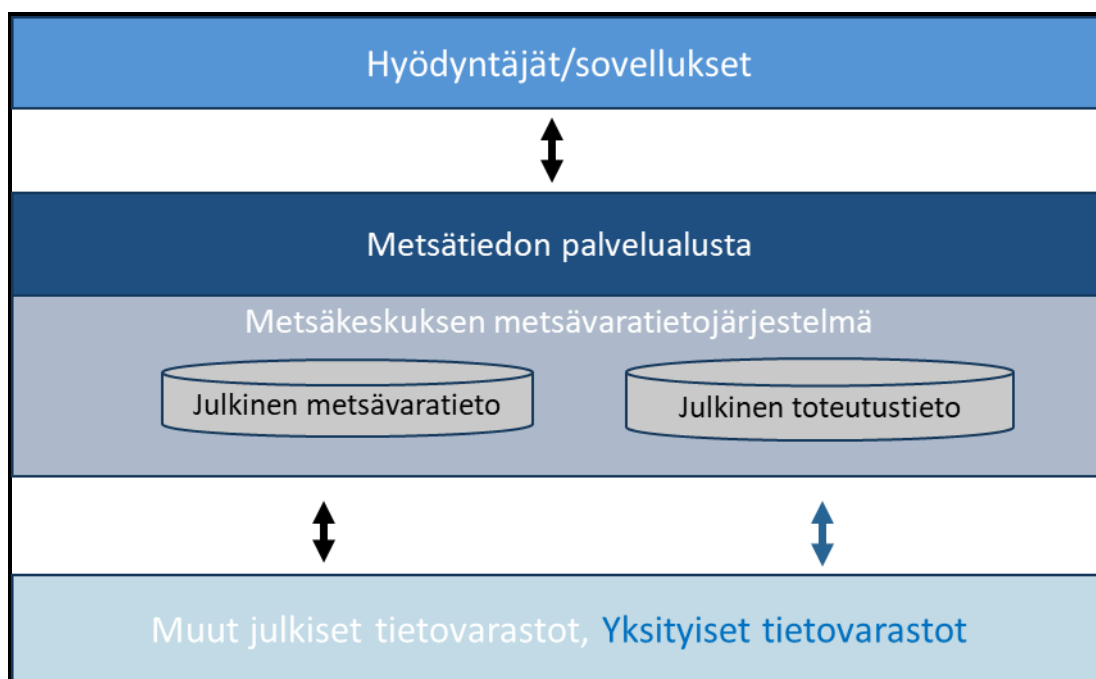
Ajatuksena on, että palvelualustan kautta voidaan hyödyntää myös muiden kuin metsäkeskuksen tuottamia julkisia ja yksityisiä tietolähteitä. Datalähteen omistaja päättää datan käyttöoikeuksista ja maksullisuudesta. Käyttöoikeudet ja laskutusperusteet voidaan hallita palvelualustassa.

Rahoituksen lähtökohtana on se, että Metsäkeskus rahoittaa omien tehtäviensä ja tietopalveluidensa hoitoon tarvittavan alustan rakentamisen, ylläpidon ja kehittämisen. Alustan käytöstä eli datojen hausta ja yhdistämisestä peritään omakustannushinta.

Lisäksi ehdotetaan, että alan toimijoista kootaan neuvottelukunta ohjaamaan alustan kehittämistä.

Maa- ja metsätalousministeriö on selvittänyt organisointi- ja rahoitusehdotuksen toteutusedellytyksiä suhteessa metsäkeskuksen tehtäviin ja asiaan liittyviin säädöksiin. Selvitykset liittyvät erityisesti valtioneuvoston. Selvityksen tuloksista riippuu missä laajuudessa palvelualusta voidaan tarjota yksityisten toimijoiden käyttöön ja voidaanko

yksityisiä tietoaaineistoja jakaa ja yhdistää palvelualustalla yksityisiin tarpeisiin. Tätä kirjoitettaessa selvitykset ovat vielä osittain kesken.



Kuva 5. Metsätietoekosysteemi 2020 -. Metsätiedon palvelualusta ehdotetaan liitettäväksi osaksi metsäkeskuksen metsävaratietojärjestelmää. Ne muodostavat yhdessä ”metsätietoalustan”, joka varastoi, yhdistää ja välittää dataa eri tietolähteistä loppukäyttäjien sovelluksille.

3.9 Toteutusvaiheen arviointi

Projekti eteni valtaosin suunnitelman mukaisesti ja päätavoitteet saavutettiin. Projektin loppuvaiheessa resursseja suunnattiin erityisesti tehtäviin, jotka tukevat alustan tuotteistamista metsäkeskuksen metsävaratietojärjestelmän yhteyteen.

Kokeilualustan testauksen käynnistäminen sidosryhmien piirissä myöhästyi noin kolmella kuukaudella aiotusta, ja projektin kokonaisaikataulua jouduttiin venyttämään sen takia. Viivästys aiheutui joidenkin järjestelmäosioiden rakentamisen ja aineistojen esiprosessoinnin suunniteltua pidemmästä kestoista. Alustan testausvaihe jäi hiukan suunniteltua suppeammaksi osin aikataulusyistä ja osaksi sen takia, että hilatietoa hyödyntäviä sovelluksia ei vielä ollut käytettävissä testaukseen siinä laajuudessa kuin oli kaavailtu.

MS Azure -pilvialusta tarjosi erittäin toimivat työkalut kokeilujärjestelmän rakentamiseen. Alustan toiminnallisuus saatiin nopeasti rakennetuksi valmiita ohjelmistomoduuleja hyödyntäen.

Projektin operatiivisessa toteutuksessa käytetty Scrum-menetelmä osoittautui toimivaksi ja tulokselliseksi malliksi tämäntapaiseen hankkeeseen.

3.10 Julkaisut

Pääjulkaisut ovat viimeistelyvaiheessa (valmistuvat 5/2019):

- Metsätehon raportti ja tuloskalvosarja
- ForestJSON-ohje (moniste)
- Metsätiedon palvelualustan tietolähteet ja tietomallit (moniste).
- Data interface services to access and fuse heterogeneous forest data (tieteellinen artikkeli)

Hämäläinen ym. 2019. Metsätiedon palvelualusta. Esitys MMM:n Metsätieto ja sähköiset palvelut -hankkeen loppuseminaarissa 22.1.2019.

Saatavissa: <https://mmm.fi/documents/1410837/11872529/Metsätiedon+palvelualusta.pdf/caa0209d-57d3-ae1d-4b5d-db15c1e24a17/Metsätiedon+palvelualusta.pdf.pdf>

Juntunen, T. & Ritala, R. 2019. Uusien tietolähteiden sovitus metsätiedon palvelualueen hilaresoluutioon. Tampereen yliopisto. Moniste. Saatavissa (6/2019): www.metsateho.fi

Hämäläinen, J., Räsänen, T., Ritala, R., Häme, T. & Tergujeff, R. 2017. Seuraavan sukupolven palvelualueen kehittäminen – esiselvitys. Metsätehon raportti 241. Vantaa.
Saatavissa: <http://www.metsateho.fi/seuraavan-sukupolven-palvelualueen-kehittaminen-esiselvitys/>

Rajala, M. & Ritala, R. 2016. Data platform promoting forest data utilization through uniform access to heterogeneous data. Metsätehon raportti 240. Vantaa.
Saatavissa: <http://www.metsateho.fi/data-platform-promoting-forest-data-utilization/>.

4. Tulosten arviointi

4.1 Tulosten soveltamiskelpoisuus

Konseptitestauksen tulokset edistävät seuraavan sukupolven metsätietoekosysteemin rakentamista. Palvelualuekonsepti osoittautui toimivaksi ja se on omiaan tehostamaan heterogeenisten metsätietolähteiden käyttöä uusien sovellusten ja palvelujen pohjana. Lisäksi alusta helpottaa datojen hankintaa ja yhdistämistä esimerkiksi viranomaiskäyttöä tai tutkimusta varten.

Konseptista on hyötyä kaikille metsäalan toimijoille ja alan ulkopuolisille, jotka hyödyntävät metsään liittyvää paikkatietoa sekä ICT-sovellusten ja palvelujen tuottajille. Alusta tarjoaa myös potentiaalia metsätiedon, sovellusten ja palvelujen markkinapaikan ja niihin pohjautuvan uuden liiketoiminnan kehittämiseksi.

Rakennettu kokeilujärjestelmä on sellaisenaan siirrettävissä pilvialustaan tuotteistamista ja jatkokehittämistä varten. Suomen metsäkeskuksen laajan metsävaratietojärjestelmän vaatimukset ja erityispiirteet heijastuvat luonnollisesti rakennettavaan järjestelmäkokonaisuuteen. Niitä ei ole voitu ottaa kattavasti huomioon tässä hankkeessa. Myös käyttäjäläpäläuteissa saadut ehdotukset kannattaa hyödyntää tuotteistusvaiheessa.

4.2 Tulosten tieteellinen merkitys

Raportissa (Rajala & Ritala 2016) kuvattiin metsävaratiedon palvelualueesta teoreettisesti bayesilaisena paikkatietojärjestelmänä. Nyt pilotoitu palvelualue on onnistuneesti toteuttanut monet raportin yleisistä toiminnallisuuksista. Käytännön tarkasteluissa on rajoitettu esimerkiksi datafuusiossa valitsemaan data luotettavimpana pidetystä lähteestä, jos samaa suuretta koskeva tieto on saatavana monesta eri tietolähteestä. Palvelualueesta kuitenkin tukee myös bayesilaisten yhdistettyjen estimaattien laskentaa, jos tietolähde pystyy tuottamaan oman mittaustietonsa epävarmuusarvion. Tämä antaa mahdollisuuden tutkia, miten muodostetaan yhä täsmällisempää tietoa metsän käsittelyä tueksi.

Mikael Filppula teki Tampereen Teknillisen Yliopiston diplomityönään palvelualueen avulla tehtävän tyyppileimikoinnin feasibility study'n.

Tutkimuksen näkökulmasta palvelualueesta liittyviä jatkotutkimuksen aiheita ovat:

- epävarmuustiedon kerääminen ja hyödyntäminen monilähde-estimaatteja laskettaessa
- ajantasaistamisen epävarmuustarkastelut
- koneoppimismenetelmiä soveltaen kehitettävät laskennalliset nopeat kasvumallit ajantasaistamiseen
- usean lähekkäisen hilapisteen arvoja yhdistelevät laskennalliset kentät (konvoluutiosuotimet) metsädatan tulkinnan tukena

Palvelualueesta on kansainvälisesti merkittävä monilähteistä paikkatietoa yhdistävä rajapintapalvelu.