

Metsätehon raportti 252
22.5.2019

METSÄTIEDON PALVELUALUSTAN KONSEPTITESTAUS

Jarmo Hämäläinen
Tapio Räsänen
Kirsi Riekki
Juha-Antti Sorsa
Risto Ritala

ISSN 1796-2374 (Verkojulkaisu)

METSÄTEHO OY
Vernissakatu 1
01300 Vantaa

www.metsateho.fi

METSÄTIEDON PALVELUALUSTAN KONSEPTITESTAUS

Jarmo Hämäläinen
Tapio Räsänen
Kirsi Riekkö
Juha-Antti Sorsa
Risto Ritala

Metsätehon raportti 252
22.5.2019

ISSN 1796-2374 (Verkojulkaisu)

© Metsäteho Oy

SISÄLLYS

ESIPUHE	2
TIIVISTELMÄ	3
1 JOHDANTO	4
2 TULOKSET	6
2.1 Kokeilujärjestelmän rakenne	6
2.2 Datakyselyjen toteutus	6
2.3 Testatut aineistot.....	7
2.4 Aineistojen esiprosessointi, yhdistäminen ja tulostiedostojen luonti.....	8
2.4.1 Esiprosessointi.....	8
2.4.2 Yhdistäminen	9
2.4.3 Tulostiedostojen luonti	11
2.5 Esimerkkejä käyttötapauksista	11
2.6 Käyttäjätestaukset ja niiden palaute	12
2.7 Testialustan suorituskyky- ja kustannustarkastelut.....	14
2.8 Organisointi ja rahoitusmalli	14
3 JOHTOPÄÄTÖKSET	16
KIRJALLISUUS	17

ESIPUHE

Konseptitestaus toteutettiin Metsätehon ja maa- ja metsätalousministeriön yhteishankkeena. Työ kuului maa- ja metsätalousministeriön Metsätieto ja sähköiset palvelut -hankekokonaisuuteen, jonka tavoitteena oli tehostaa metsävaratiedon hyödyntämistä, parantaa tiedon laatua ja liikkuvuutta sekä kehittää sähköisiä palveluita.

Metsäteho vastasi kokeilujärjestelmän määrittelystä, rakentamisesta ja testauksesta. Maa- ja metsätalousministeriö toteutti organisointi- ja rahoitusmalliin liittyvät juridiset selvitykset. Tampereen teknillinen yliopisto (nykyisin Tampereen yliopisto) toimi asiantuntijana datakyselyihin ja -fuusioon liittyvissä kysymyksissä ja CGI vastasi kokeilujärjestelmän teknisestä rakentamisesta.

Projektityöhön osallistuivat Jarmo Hämäläinen, Tapio Räsänen, Kirsi Riekkö 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.

Metsätehon puolesta esitän parhaat kiitokset hankkeen rahoittajille ja työhön osallistuneille.

22.5.2019

Jarmo Hämäläinen

TIIVISTELMÄ

Hankkeessa rakennettiin ja testattiin metsätiedon palvelualustakonsepti, joka on tarkoitettu tehostamaan metsätiedon hallintaa ja helpottamaan sen hyödyntämistä sovelluksissa ja palveluissa. Sitä varten rakennettiin kokeilujärjestelmä Microsoft 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 14 eri tyyppistä tietolähdettä sen toiminnallisuuden rakentamista ja testausta varten. Sidosryhmien asiantuntijat testasivat palvelualustaa, minkä pohjalta koostettiin kehittämisehdotuksia 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 päätöstukisovellusten laadintaa, hylätasoisien 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 perustelluina 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 hylätiedot, kannattaa sijoittaa alustan sisälle.

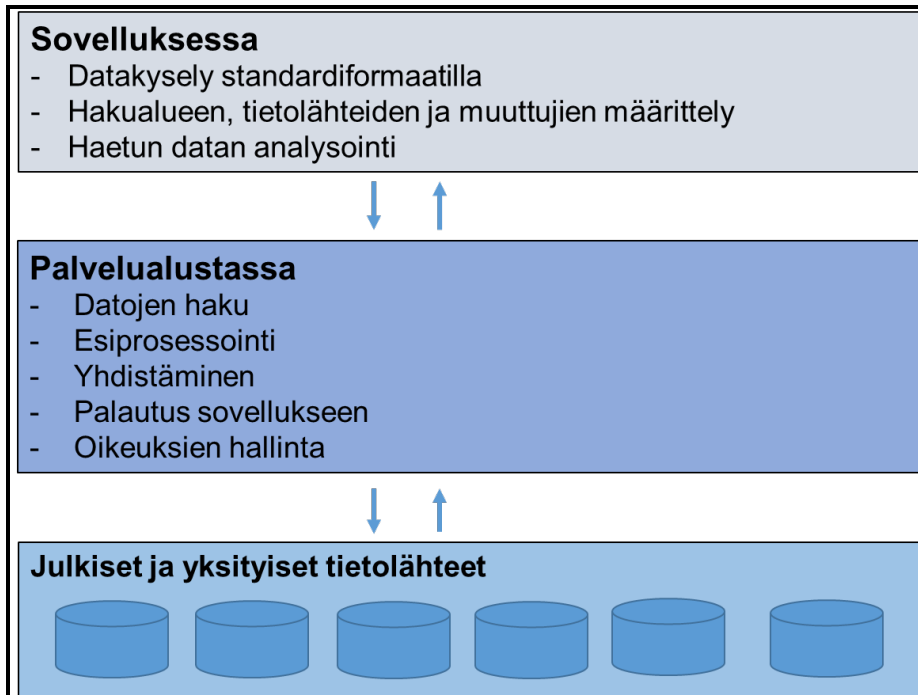
Pilvialusta tarjoaa tehokkaan ja skaalautuvan teknisen ympäristön palvelulle. Alustakonseptissa on paljon metsäalan innovaatioita ja uutta liiketoimintaa vauhdittavaa potentiaalia. Tulevaisuudessa se voisi olla datojen jakelualustaa laaja-alaisempi datojen, mallien, sovellusten ja palvelujen markkinapaikka.

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

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

1 JOHDANTO

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 & Ritola 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.

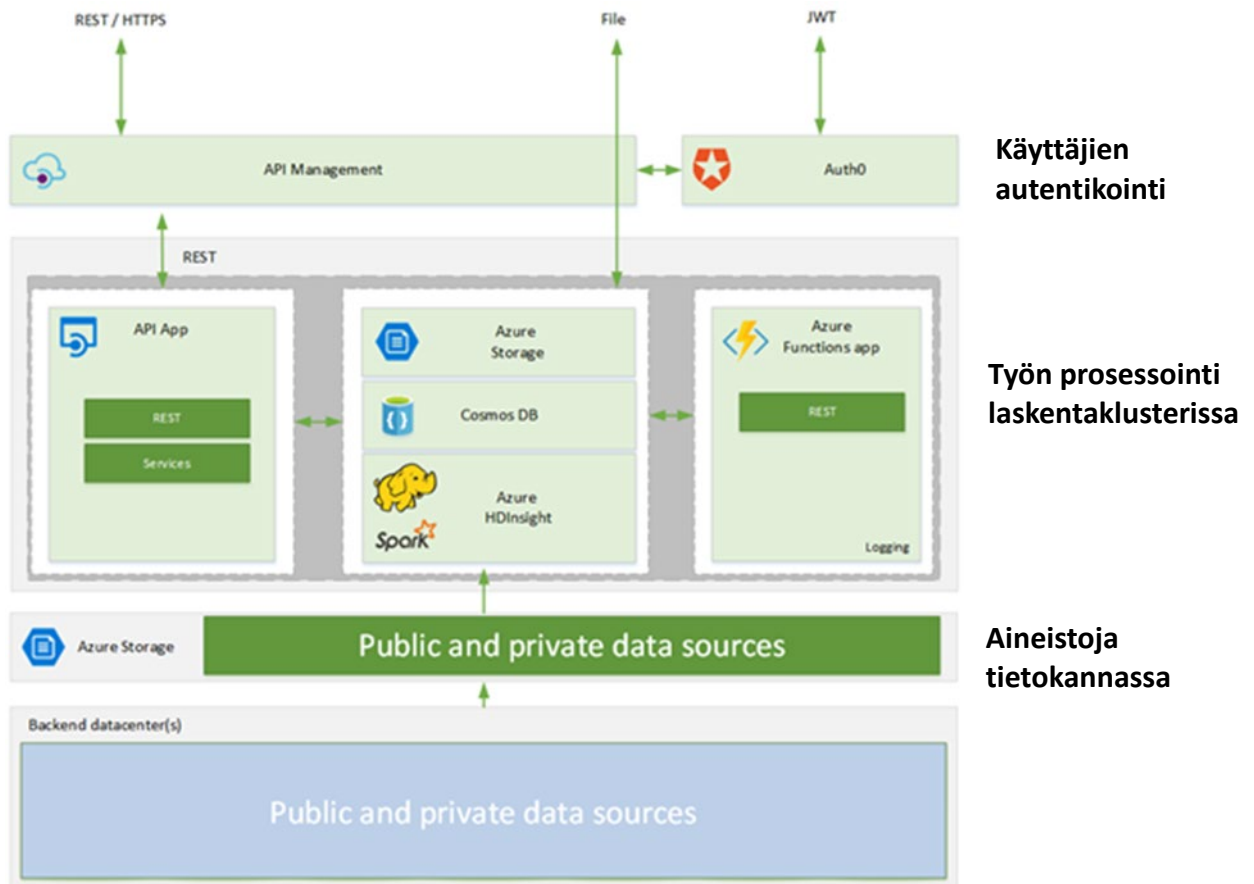
Hanke toteutettiin Metsätehon ja maa- ja metsätalousministeriön yhteishankkeena Metsätieto ja sähköiset palvelut -kärkihankkeen osana. Metsäteho vastasi hankkeen johtamisesta, järjestelmän toiminnallisesta määrittelystä, tietoaineistojen kokoamisesta, järjestelmän testauksesta, organisointi- ja rahoitusvaihtoehtojen tarkastelusta sekä raportoinnista. Maa- ja metsätalousministeriö vastasi organisointi- ja rahoitusmalliin liittyvistä juridisista selvityksistä. Tampereen teknillinen yliopisto toimi asiantuntijana datojen haun, esiprosessoinnin ja yhdistämisen menetelmien kehityksessä sekä datahakuihin tarvittavan kyselykielen määrittelyssä ja kyselykoodia tuottavan apuvälineen kehittämisessä. CGI vastasi kokeilujärjestelmän teknisestä määrittelystä sekä sen rakentamisesta. Metsäkeskus osallistui hankkeeseen asiantuntijajäsenenä.

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.

2 TULOKSET

2.1 Kokeilujärjestelmän rakenne

Projektin ensivaiheessa määriteltiin kokeilujärjestelmän toiminnalliset ja tekniset vaatimukset. Vaatimusmäärittelyn pohjalta rakennettiin järjestelmä Microsoft Azure -pilvialustalle (kuva 2).



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

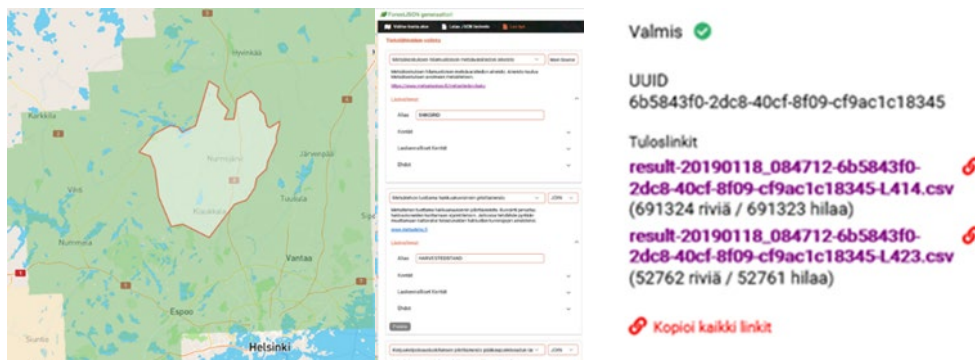
2.2 Datakyselyjen toteutus

Palvelualustan käyttö perustuu dataa käyttävään sovellukseen ohjelmitavaan, alustalle lähetettävään kyselyyn, jonka perusteella alusta hakee tarvittavat datat ja palauttaa ne sovellukselle. 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 kuntalistalta ja tekemällä em. datahaun valinnat alavetovalikoilla (kuva 3). Sovellus tuottaa automaattisesti ForestJSON-ohjelmakoodin, joka voidaan lähettää alustalle tai jota voidaan käyttää koodin laadinnan ja muokkauksen pohjana. Sovellus voi olla tarpeellinen myös tuotteistetussa palvelualueella etenkin sellaisia datan hyödyntäjiä varten, joilla ei ole käytössään varsinaista loppukäyttösovellusta. Muut sovellukset voivat kuitenkin käyttää kaikkia palvelualueen toiminnallisuuksia ohjelmallisesti ilman Generator-sovellusta.

Palvelualue 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 vastaus voidaan antaa csv- tai GeoPackage-formaatissa.



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

2.3 Testatut aineistot

Kokeilujärjestelmään liitettiin yhteensä 14 eri tyyppistä aineistoa, joiden avulla rakennettiin ja testattiin alustan toiminnallisuus (taulukko 1). Tietolähteiden tarkempi kuvaus lähteessä Riekkö & Räsänen (2019). 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.

Taulukko 1. Testialustaan liitetyt paikkatietoaineistot. MVMI tarkoittaa monilähteistä valtakunnan metsien inventointia.

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

2.4 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-50-karttalehtijakoa (24 x 48 km). Käytännössä tämä tarkoittaa sitä, että myös kyselyn hakualue jaetaan karttalehtiin, ja jokainen karttalehti käsitellään erillisenä alustassa kyselyä suoritettaessa.

2.4.1 Esiprosessointi

Palvelualustan spatiaalinen tiedonhallintayksikkö on valittu siten, että se vastaa Suomen metsäkeskuksen hilamuotoisen metsävaratiedon esittämisessä käytettyä hilajakoa. Metsäkeskuksen hilaruutu edustaa 16 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.

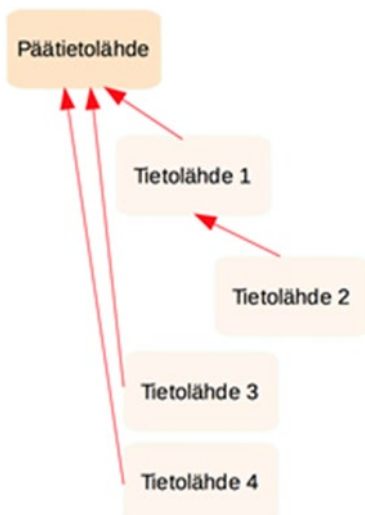
Osa testialustan aineistoista on rajapinnasta saatavia, dynaamisia tietoja, kuten esimerkiksi Ilmatieteen laitoksen tarjoamat aineistot. Kyselykohtainen rajapintahaku todettiin projektissa tällaisille aineistoille parhaaksi vaihtoehdoksi, koska päivitystahti on nopea ja siirrettävä tietomassa pieni. Rajapinnan kautta alustaan liitetyt tietolähteitä 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. Aluemuotoisilla 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.

2.4.2 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. (Mäkelä ym. 2019).

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 palvelualueella 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 vastauskentät 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 Mottimalliin. UPDATE tuottaa erillisinä lisäkenttinä kasvunlaskennan jälkeiset puustotunnukset kyseiselle hilalle.

2.4.3 Tulostiedostojen luonti

Palvelualusta palauttaa tulostiedostot UTM-50-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 palauttamat csv-tiedostot ovat sellaisenaan avattavissa paikkatieto-ohjelmilla, koska hilakohtaiset koordinaatit ovat mukana aineistossa.

Joillain testatuilla tietolähteillä, kuten FMI:n metsäpaloindeksilla ja Metsäkeskuksen metsävarakuviolla, 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 sille aggregoidut kenttien arvot. Myös kaikki nämä tiedostot tulevat karttalehdittäin.

2.5 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 hakkuukonedatata muodostetut hakatun alueen rajat, hakkuun ajankohta ja hakkuutapatieto. Nämä tiedot liitettiin alkuperäisen hilatiedon lisäkentiksi. Harvennettujen hilojen puustotietoja ei tässä vaiheessa päivitetty laskennallisesti. Siinä voitaisiin periaatteessa soveltaa lähtöpuustoon ja harvennismalleihin 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 alue-rajauksille lasketaan niihin kuuluvien, kyselyn ehdot 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 puukauppakohteiden 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 ja tulosten laskenta- ja esittämisperiaatteet. Palvelualusta hakee ja yhdistää datat määrittelyjen pohjalta. Sovellus laskee ja esittää hakkuumahdollisuudet graafeina ja tuottaa myös tiedostot jatkoanalyysijä varten.

2.6 Käyttäjätestaukset ja niiden palaute

Projektiryhmän testikäytön lisäksi palvelualusta ja hakkuumahdollisuussovellus avattiin sidosryhmille testattavaksi 20.12.2018–31.3.2019 väliseksi ajaksi. Testaajille toimitettiin testausohje, avattiin yhteys 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 minuuttiin. Yksi käyttäjä kokeili myös datan vientiä karttatasoksi organisaation omaan suunnittelujärjestelmään.

Käyttäjäpalautteen keskiarvo oli 3,5 asteikolla 1–5, jossa 5 oli paras arvosana. Yhteenveto testaajien havainnoista ja kehittämissuhteista oli seuraava:

Palvelualustalla nähtiin yleisesti olevan monenlaisia hyödyntämismahdollisuuksia vastaajien organisaatioissa, 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
- luonnonsuojelualueiden 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 ylipäätään 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 (toteutettiin hankkeen loppuvaiheessa)
- 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 Generator -sovelluksen käyttöliittymään, mm. aineistojen näkyvyys kartalla, kyselyajojen ja vastaustiedostojen hallinta, tietokenttien valinta datan olemassaolon mukaan.

2.7 Testialustan suorituskyky- ja kustannustarkastelut

Kokeilualustan suorituskykyä ja kustannuksia tarkasteltiin tekemällä eri laajuisia datakyselyjä (taulukko 2). Kyselyalueiden pinta-ala vaihteli 4,3 hehtaarista 13,5 miljoonaa hehtaariin ja tietolähteiden määrä kolmesta viiteen hilamuotoiseen tai aluemuotoiseen (polygoni) 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 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
"Maa-kunta"	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
"Maa-kunta"	542 850	21,2 milj.	hila, MVMI, kunta, kklk, htapa	1 h 13 min	1,53
"Suuralue"	13,5 milj.	527 milj.	hila, MVMI, kunta	3 h 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

2.8 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, päivitystä ja jakelua. Näin se pystyy vastaamaan entistä tehokkaammin asiakastarpeisiin.
- Keskeisimmät julkiset metsätiedon varastot ovat Metsäkeskuksen hallussa ja ylläpitävällä.

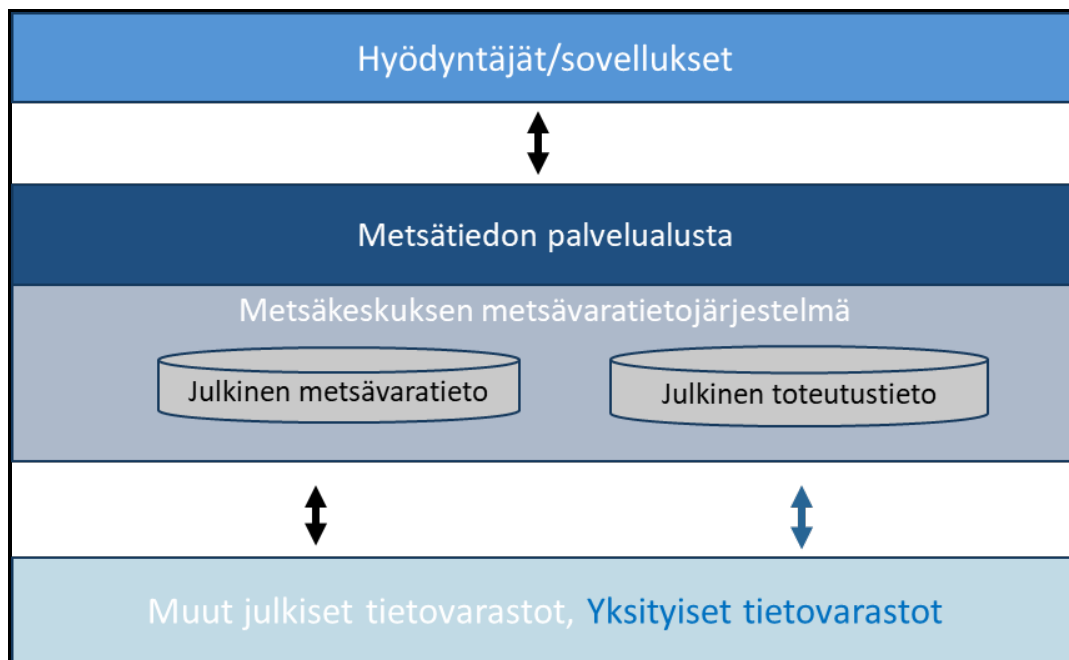
- Palvelualustan suorituskyvyn kannalta avaintietovarastot kannattaa sijoittaa alustan sisälle – eikä rajapintojen taakse – mikä on helpoimmin järjestettävissä, kun Metsäkeskus on palvelualustan omistaja.
- Näköpiirissä oleva operatiivisen toteutustiedon käyttö Metsäkeskuksen metsävaratiedon ajantasaistuksessa tulee heijastumaan päivitysprosessin vaatimuksiin. Tehokkaaseen ja skaalautuvaan pilviratkaisuun perustuva 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 valtiontukeen. Selvityksen tuloksista riippuu, missä laajuudessa palvelualusta voidaan tarjota yksityisten toimijoiden käyttöön ja voidaanko yksityisiä tietoaineistoja 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 JOHTOPÄÄTÖKSET

Konseptitestauksen tulokset edistävät seuraavan sukupolven metsätietoekosysteemin rakentamista. Palvelualustakonsepti 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. Pilvialusta tarjoaa tehokkaan ja skaalautuvan teknisen ympäristön palvelulle.

Palvelualustasta on hyötyä kaikille metsäalan toimijoille ja alan ulkopuolisille, jotka hyödyntävät metsään liittyvää paikkatietoa, sekä ICT-sovellusten ja palvelujen tuottajille. Alustakonseptissa on myös paljon metsäalan innovaatioita ja uutta liiketoimintaa vauhdittavaa potentiaalia. Tulevaisuudessa se voisi olla datojen jakelualustaa laaja-alaisempi datojen, mallien, sovellusten ja palvelujen markkinapaikka.

Suomen metsäkeskuksen metsävaratietojärjestelmän yhteydessä alusta on omiaan tehostamaan ajantasaisen metsävaratiedon tuottamista, päivitystä ja jakelua. Metsäkeskuksen laajan metsävaratietojärjestelmän vaatimukset ja erityispiirteet heijastuvat luonnollisesti rakennettavaan järjestelmäkokonaisuuteen. Niitä ei ole voitu ottaa kattavasti huomioon tässä hankkeessa. Myös käyttäjäpalautteissa saadut ehdotukset kannattaa hyödyntää tuotteistusvaiheessa.

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

Hilatiiedon merkitys päätöstukijärjestelmien perusaineistona on kasvamassa, ja sen tietosisällön kehittäminen on olennaista. Tärkeitä kehittämiskohteita ovat etenkin puulajeittaiset puusototunnukset kuten runkolukusarjat ja laatutiedot, maaperätiedot sekä ympäristöasioiden hallintaa tukevat tietokokonaisuudet.

Rajala & Ritala (2016) kuvasivat metsävaratiedon palvelualustat teoreettisesti bayesilaisena paikkatietojärjestelmänä. Nyt pilotoitu palvelualusta on onnistuneesti toteuttanut monet konseptin yleisistä toiminnallisuuksista. Käytännön tarkasteluissa on rajoitettu esimerkiksi datafuusiossa valitsemaan data luotettavimpana pidetystä lähteestä, jos samaa suuretta koskeva tieto on saatavana monesta eri tietolähteestä. Palvelualusta kuitenkin tukee myös bayesilaisien yhdistettyjen estimaattien laskentaa, jos sama suure on saatavissa eri tietolähteestä ja kukin tietolähde pystyy tuottamaan oman mittaustietonsa epävarmuusarvion. Tämä antaa mahdollisuuden tutkia, miten muodostetaan yhä täsmällisempää tietoa metsän käsittelyn tueksi.

Palvelualustaan liittyviä jatkotutkimuksen aiheita ovat

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

KIRJALLISUUS

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

Mäkelä, J., Koljonen, M., Vorne, H., Rieki, K. & Räsänen, T. 2019. ForestJSON-ohje. Metsätiedon palvelualustan konseptitestaus. Metsätehon raportti 252. LIITE. Saatavissa: <http://www.metsateho.fi/metsatiedon-palvelualustan-konseptitestaus-raportti>. [Viitattu 24.6.2019].

Rajala, M. & Ritala, R. 2016. Data platform promoting forest data utilization through uniform access to heterogeneous data. Metsätehon raportti 240. Saatavissa: <http://www.metsateho.fi/data-platform-promoting-forest-data-utilization/>. [Viitattu 24.6.2019].

Rieki, K. & Räsänen, T. 2019. Tietolähteiden tekninen kuvaus. Metsätiedon palvelualustan konseptitestaus. Metsätehon raportti 252. LIITE. Saatavissa: <http://www.metsateho.fi/metsatiedon-palvelualustan-konseptitestaus-raportti>. [Viitattu 24.6.2019].

Ritala, R. & Juntunen, T. 2019. Uusien tietolähteiden sovitus metsätiedon palvelualustan hi-laresoluutioon. Metsätiedon palvelualustan konseptitestaus. Metsätehon raportti 252. LIITE. Saatavissa: <http://www.metsateho.fi/metsatiedon-palvelualustan-konseptitestaus-raportti>. [Viitattu 24.6.2019].