

6.2.2018

Selvitys tutkimuslaitosten datanhallinnan ja tieteellisen laskennan tarpeista –

Loppuraportti

Turo Kilpeläinen

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva lomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

DL-kokonaisuus

- Opetus- ja kulttuuriministeriö toteuttaa yhdessä tutkimus- ja innovaatiotoimijoiden kanssa vuosien 2017–2021 aikana datanhallinnan ja laskennan tutkimusinfrastruktuurien, palveluiden ja osaamisen kehittämisohjelman.
- Kehittämisohjelman yhtenä tavoitteena on tarjota korkeakoulujen lisäksi CSC – Tieteen tietotekniikan keskus Oy:n datanhallinnan ja laskennan tutkimusinfrastruktuuri ja siihen liittyvät palvelut jatkossa myös valtion tutkimuslaitosten käyttöön.
- Kehittämisohjelman toteutuksen tueksi on tehty selvitys, jossa kuvataan valtion 12 tutkimuslaitoksen datanhallinnan ja laskennan sekä tietoliikenneyhteyksien rooli/tarve ja tulevaisuuden tavoitetilä. Selvityksessä on kartoitettu myös laskentakapasiteetin yksilöityä tarvetta. Lopuksi esitetään toimenpide-ehdotuksia, jotka sisältävät myös esityksen kansallisiksi datanhallinnan ja laskennan piloteiksi.
- Selvityksen on toteuttanut Turo Kilpeläinen (Rehtori/Toimitusjohtaja, KAMK) kehittämisohjelman ohjaamiseen ja koordinointiin osallistuvan Suomen Akatemian toimeksiannosta.
- Lisätietoja Datanhallinnan ja laskennan kehittämisohjelmasta: www.minedu.fi/DL2021

Tutkimuslaitosten datanhallinnan ja tieteellisen laskennan rooli ja tarpeet

Selvityksen tavoitteena on kuvata valtion 12 tutkimuslaitoksen (Tilastokeskus) datanhallinnan ja laskennan sekä tietoliikenneyhteyksien rooli/tarve ja tulevaisuuden tavoitetila



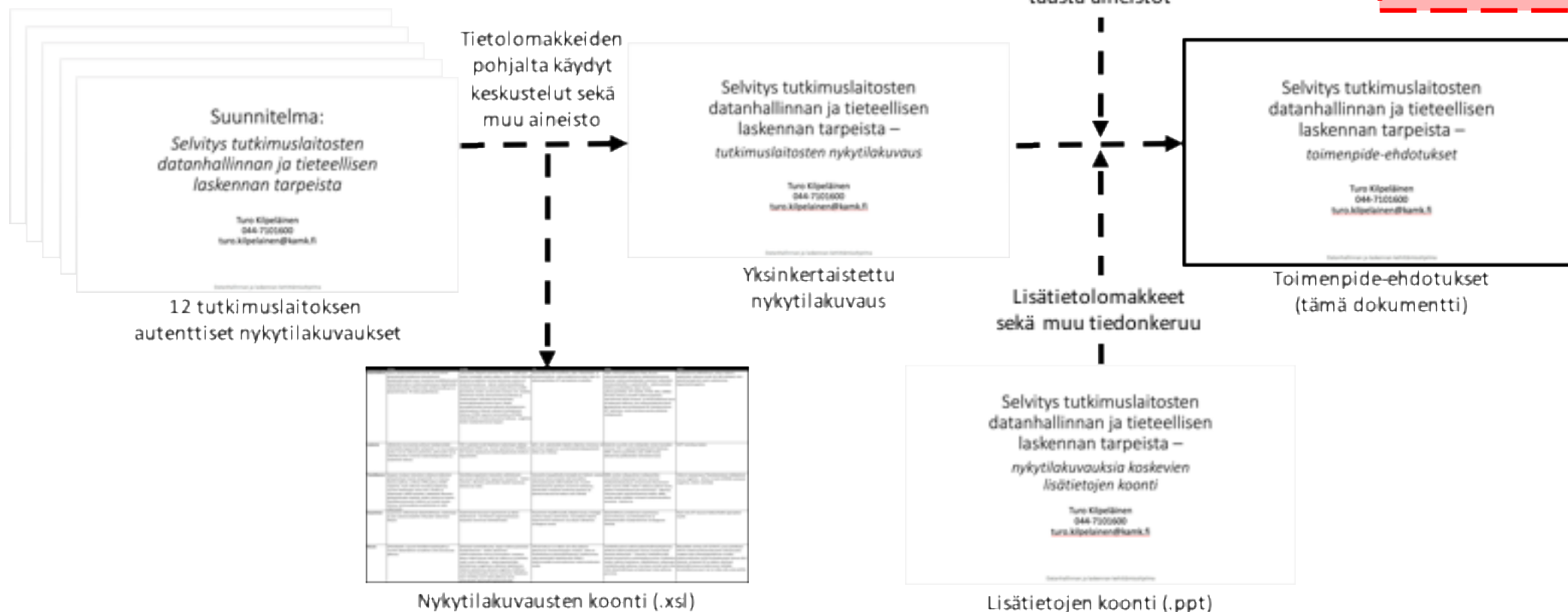
Selvityksen tuloksia hyödynnetään valtion tutkimusinfrastruktuuriympäristön ja palveluiden kehittämisresurssien priorisoinnin tukena

Selvitys toteutetaan pääasiallisesti haastatteleamalla ylintä johtoa sekä heidän osoittamia henkilöitä syksyn 2017 aikana siten, että ylätasen linjaukset ovat valmiit joulukuussa 2017

Toimenpide-ehdotusten tietopohja

HUOM!!

Dokumentissa esitetyt toimenpide-ehdotukset ovat selvitysmiehen omia näkemyksiä siitä, miten DL-toimintaa tulisi tutkimuslaitosten ja koko korkeakoulusektorin osalta edistää tulevaisuudessa



Yhteenveto tutkimuslaitosten DL-kehittämiseen vaikuttavista osa-alueista

1

Tutkimuksen DL-intensiteetti kasvaa

- **DL-intensiivisen tutkimuksen** osuuden **ennakoidaan** kokonaisuudessaan **kasvavan**
- Hankkeiden ja niissä syntyvän **monimuotoisen datan määrä kasvaa** jatkuvasti
- **Dataintensiivisen laskennan** osuuden **kasvu huomioitava DL-investoinnissa**
- Tutkimuslaitosten **DL-kapasiteetti** on monin paikoin **elinkaarensa päässä**

2

DL-osaamista lisättävä kaikilla tasoilla

- **DL-osaamista lisättävä** tutkimustyön edellytysten parantamiseksi kaikkialla
- DL-intensiivistä **tutkimusta** ja **DL-palveluyhteistyötä johdettava** kansallisesti **yhtenä kokonaisuutena**
- DL-kokonaisuuden kestävä hallinta vaatii **erikoistumista**
- Avoimen tieteen ja datan mahdollisuuksien **opetus-integrointia vauhditettava**

3

Budjetti- ja hankerahoituksen tasapaino ratkaistava

- DL-kehittämiseen kohdennettava kansallisia ja organisatorisia **resursseja** vaikuttavuuden lisäämiseksi
- Budjettirahoituksen laskun ja **hankerahoituksen kasvun tasapaino ratkaistava** strategisen DL-kehittämisen mahdollistamiseksi
- Selkeä **tarve toimivalle DL-kokonaisratkaisulle** sekä strategiselle johtamiselle

4

Tietohallintopalveluita kehitettävä kansallisesti

- Tutkimuslaitosten **pirstaloutuneet ICT-palvelut ja -osaaminen** selkeä kilpailukyvyyn kasvun este
- **Valtorin ja CSC:n** roolit tukevat toisiaan **yhteistyön ja työnjaon** syventämiseksi
- Tutkimuslaitosten ja korkeakoulujen **DL-palvelut saatettava tasavertaiseen asemaan** yhteistyöedellytysten parantamiseksi

Yhteenveto DL:n keskeisimmistä haasteista ja kehityskohteista

1	2	DL-kokonaisuuden kehittäminen	3	4
Tietohallintopalveluiden yhteistyö ja työnjako	DL-peruspalveluiden kansallinen kehittäminen	Ala- ja organisaatio-kohtaiset DL-ratkaisut	Tulevaisuuden kansalliset DL-palvelut	
- Tutkimuslaitosten ICT-palvelut ja -osaaminen pirstaloituneet, DL:n strategisessa johtamisessa kehitettävää - Valtion omistamien Valtorin ja CSC:n yhteistyötä ja työnjakoa on syvennettävä tutkimuslaitosten ja korkeakoulujen DL-palveluiden sekä tutkimuksellisten yhteistyöedellytysten parantamiseksi	- DL-palveluiden osalta korkeakoulut ovat tutkimuslaitoksia paremmassa asemassa - CSC:lle organisatorisesti, kansallisesti ja kansainvälisesti pirstaloituneen tutkimus- ja tutkimusta koskevan tiedon kokonaiskoordinaatiovastuu - Tuetaan kansallisten DL-peruspalveluiden kehittämistä pilottien avulla	- Tutkimuslaitosten strategista DL-kehittämiskykyä on vahvistettava - Omien DL-ympäristöjen liisäarvo perustuu strategisten kilpailutekijöiden vahvistumiseen, johon resurssit ovat löydettävissä - kansallisesta perus- ja tulevaisuuden DL-ratkaisuissa tehtävästä yhteistyöstä	- CSC:n kyky ja mahdollisuus pysyä teknologian aallonharjalla varmistettava - Visionääristä DL-johtamiskyvykkyyttä lisättävä toiminnan kaikilla tasoilla - DL-intensiivisen huippututkimuksen palvelutarpeisiin vastattava toimintaedellytysten, kilpailukyvyn ja vaikuttavuuden parantamiseksi	

Yhteenveto tutkimuslaitosten tarpeita tukevasta DL-investoinnista

1

DL-intensiivinen huippu- tutkimus Suomessa

- Tutkimuslaitosten nykyinen DL-intensiivisen tutkimusrahoituksen määrä varsin pieni
- **Tutkimuslaitosten** laskenta- ja erityisesti dataintensiivisen **laskentakapasiteetin tarve** kokonaisuutena **kasvussa**
- DL-laiteinvestoinnissa painottettava myös **toimintamallien ja osaamisen** kehittämistä
- **DL-osaamisella tärkeä rooli lähes kaikessa tutkimuksessa**

2

Tutkimuslaitosten näkemys DL-investoinnista

- Budjettirahoituksen laskun ja hankerahoituksen kasvun tasapaino ratkaistava osana strategista DL-kehitystä
- Tutkimusorganisaatioiden **avoimen datan** ympärillä tehtävää **yhteistyötä kehitettävä kokonaisuutena**
- **Osaamisinvestoinnit** tärkeitä
- Tavoitteena **keskitetty, joustava ja kehittynyt DL-kokonaisratkaisu**

3

Varsinainen
DL-investointi

DL-kapasiteetti-, -työvälineet ja -palvelut

- **DL-investoinnin palveltava laskenta-** (ml. GPU) **sekä** kasvavan ja monimuotoistuvan **dataintensiivisen tieteen tarpeita**
- DL-palvelualustateknologian **tuettava datan hyödynnettävyyttä tutkimuksen raja-pinnoilla** (ml. PPP)
- **KV-yhteistyön** edellytykset on **turvettava** (ml. data-liikenteen tietoturva)

4

Toimintamallit ja osaaminen

- DL-intensiivistä tutkimus-yhteistyötä tuettava mm. tarkoituksenmukaisella **datanhallintapolitiikalla**
- **Hyviä käytänteitä** edelleen vahvistettava ja levitettävä
- DL-intensiivinen tutkimus ja siitä syntyvä data on **integroitava** korkeakoulujen **opetukseen** tehokkaasti
- DL:n **strategisen merkityksen tietoisuutta lisättävä**

Selvityksessä käytettävä terminologia

(ATT-käsitteistö)

- *Data*: digitaalinen aineisto, johon ei liity tulkintaa
- *Datanhallinta*: datankeruu, organisointi ja tallentaminen siten, että se on helposti löydettävissä ja käytettävissä
- *(tieteellinen) laskenta*: tietokoneella suoritettava, suuriin data-aineistoihin kohdentuva ja suurta laskentatehoa edellyttävä laskenta, jolla pyritään matemaattisten mallien avulla ratkaisemaan tieteellisiä ongelmia
- *Tutkimus*: systemaattista ja luovaa toimintaa, jonka tavoitteena on olemassa olevan tiedon lisääminen
- *Tutkimusaineisto*: tutkimuksessa tuotettava ja/tai käytettävä materiaali, jota analysoimalla pyritään vastaamaan tutkimusongelmaan
- *Tutkimusta koskeva tieto*: tutkimuksen toistettavuuteen ja todennettavuuteen liittyvä tieto
- *Tutkimusinfrastruktuuri*: tarkoittaa keskitetysti, hajautetusti tai virtuaalisesti saatavilla olevia välineitä, laitteistoja, tietoverkkoja, tietokantoja, aineistoja ja tutkimuksen palveluita, jotka mahdollistavat tutkimuksen toteuttamisen sen eri vaiheissa
- *Tietoliikenneyhteys*: tietoliikennettä välittävä yhteys esimerkiksi tutkimusinfrastruktuurien välillä

Sisältö

- Tietopohja ja yhteenveto
- **Nykytila-analyysi**
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Nykytilakuvauksessa hyödynnetyn tietolomakkeen ja haastattelun runko

1. Datanhallinnan tarpeet
 - Nyky- ja tavoitetilakuvaus yksinkertaiselle SWOT-pohjalle
 - Kuvaus tutkimuslaitoksen tutkimusaineistoihin ja tutkimukseen liittyvän tiedon avoimuuden politiikoista
2. Laskennan tarpeet
 - Nyky- ja tavoitetilakuvaus yksinkertaiselle SWOT-pohjalle
 - Kuvaus tutkimuslaitoksen laskentaan liittyvistä politiikoista
3. Tutkimuksen tietoliikenteen tarpeet
 - Tutkimuksen datanhallinnan ja laskennan tietoliikenteeseen liittyvät tarpeet ja yhteistyömahdollisuudet
4. Osaamistarpeet
 - Datanhallinnan ja laskennan osaamisen tarpeet
5. Muuta huomioitavaa
 - Vapaa sana aiheesta datanhallinnan ja laskennan kehittäminen



Sisältö

- Tietopohja ja yhteenveto
- **Nykytila-analyysi**
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Tutkimuslaitoskohtaiset nykytilakuvaukset

Luke											
- Datanhallinta - Luken data auki tutkimuskäyttöä varten (ATT); jatkossa raakadatasta tehdään tuotteita - Strateginen linjaus: Valtorin palvelut otetaan 100 % käyttöön (tulossopimus) - Yksittäiset datasetit eivät kovin suuria (pitkäaikaiseurannat kasvattavat) ja luokiteltua dataa vähän – hankedata, erityisesti genetiikka ja laserkeilaus kasvavat nopeasti sekä seurannoissa ja tutkimuksessa - Laskenta - Geneettisten menetelmien ja tutkimuksen kehitys sekä massadatan kasvava käyttö lisäävät tehokas laskentakapasiteetin tarvetta - Laskenta tehokannettavilla, Luken laskentapalvelimilla ja/tai Valtorin tarjoamissa pilvipalveluissa sekä CSC:n tehokas laskentakapasiteetissa - liittymisen yliopistojen laskentaklusteriin? - Omaa laskentakapasiteettia on, mutta elinkaarensa päässä – tavoitteena hankkia useamman koneen klusteri tai siirtä pilvipalveluihin - Tietoliikenne - Meneillään Valtorin selvitys ratkaista tietoliikenteen kapasiteettiongelma - Luken toimipaikkarakenteen voi edellyttää Funetin käyttöä ainakin yliopistopaikkakunnilla ja Jokioisilla - Tietoliikenneyhteydet voivat aiheuttaa pullonkauloja CSC:n palveluiden käytölle - Osaaminen - Osaamisvajetta keskeinen ongelma – tutkijavetoista, käyttö- tai menetelmätukea ei ole - Käynnissä Luken tutkijoista koostuvan, vertaistukena toimivan laskentaverkoston rakentaminen - CSC:n palveluiden hyödyntäminen (mm. kurssit) - Muuta - Laskentakapasiteetin saatavuus hidastaa projektien toteuttamista ja heikentää kilpailukykyä - Valtorin ja CSC:n palvelut saatava yhdestä luukusta – nyt eivät täytä vaatimuksia - Oletus on, että CSC:n palveluihin ei ryntäystä, piloteilla liikenteeseen: genetiikka, laserkeilaus ja oGIIR/paikkatieto - Ohjelmistolisenssit selvittettävä: akateeminen/viranomaiskäyttö/kaupallinen	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

	SYKE										
Luke	- Datanhallinta - Data: tutkimushankkeet ja seurannat - Lähes kaikki data ATT-periaatteiden mukaisia - Keskitettyt järjestelmät ja datanhallinta mitoitettu hyvin omiin tarpeisiin - Järjestelmissä avoimet rajapinnat – aineistot helposti siirrettävissä - Keskitetyn järjestelmän ylläpito vaativaa, muutokset hitaasti toteutettavia - Datan tallennustilan kasvava puute ja toiminnan hankerahoittuminen keskeisimpiä uhkia - Laskenta - Laskennan tarve/rooli tulee kasvamaan jatkossa - Riittävän tehokkaiden laskentaympäristöjen sekä datan tallennustilan puute - Tietoliikenne - Kansallinen satelliittidatakeskus hyvä esimerkki keskitetystä, yhteistyössä toteutetusta palvelusta - Rajapintoja mm. Luke, Metsähallitus, IL ja Luomus/HY sekä TULA-net - Valtorin rooli epäselvä - Osaaminen - Tällä hetkellä osaaminen riittävä omiin tarpeisiin, vahva osaaminen sovelluskehitystyössä ja paikkatietoinfrastruktuurissa - Strategisen kehittämisosaamisen puute - Substanssiosaamisen priorisoinnilla vaikutus laajasti ICT-osaamiseen/-palveluihin - Muuta - SYKE on käynnistänyt sisäisen ympäristötiedon hallinnan strategisen kehittämisohjelman - Toimivien käytänteiden edelleen vahvistaminen - Kansallisesti horisontaalinen yhteistyö luo uusia mahdollisuuksia	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

		VTT									
Luke	SYKE	- Datanhallinta - Data pääosin tutkimusprojektikohtaista eikä siten julkista - mittaus- ja laskentadatan määrä kasvaa - Oma laskentakapasiteetti käytössä (keskitetty laskentaklusteri, tutkimusalueiden erilliset laskentayksiköt sekä tehotyöasemapohjainen kapasiteetti) - Avoimen toiminnan ohjeistus olemassa - Datanhallinnan tärkeys EU- ja muissa julkisesti rahoitetuissa hauissa, sisäinen kehitysryhmä käynnistynyt - Kehitteillä avoimen datan metadatan keräämistä ja julkaisemista tukeva tutkimustietojärjestelmä (CRIS) - Laskenta - Oma kapasiteetti tärkeä tietoturvallisuuden, laskentadatan ja laskentatuloksien tiedonsiirron sekä ohjelmistolisenssien näkökulmasta - Ei organisatorista laskennallisen tutkimuksen strategiaa (ulkoisten laskentapalveluiden käyttö) - ulkoisten laskentapalveluiden käyttö ja tuntemus vähäistä, kuten myös laskennan skaalaaminen HPC-luokkaan - Oman kapasiteetin ajantasaisena pitäminen riippuu VTT:n investointikyvystä - Tietoliikenne - Kahdennettu 10G Funet-liittymä (Internet), toimipisteiden väliset yhteydet eri operaattoreiden toteuttamia - Tietoliikennetkaisu toimivia Funet/Valtari-toimijoiden välillä - Tarve tietoturvallesiin ja tehokkaisiin tiedonsiirtoyhteyksiin kasvaa (pilvipalvelut) - Osaaminen - Pitkät perinteet teknisessä teollisen tason laskennassa – kuitenkin pirstaloitunut - Laskennallisen tutkimuksen osaamisen kehittäminen ja ylläpito jäykkää - ATT-vaatimukset tunnetaan vielä puutteellisesti – vaatii koulutusta ja viestintää - Muuta - Laskennallisen tutkimuksen, siihen liittyvien menetelmien ja ohjelmistojen kehitystä edistetään lähinnä tutkimus- ja kehityshankkeissa (strategia tarvitaan) - Laskennallinen tutkimus riippuvainen laskentakapasiteetista (joustamattomuus uhka) - Laskentakapasiteettia on Suomessa saatavilla, mutta sen joustava käyttö organisaatiorajojen yli on haastavaa	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
			- Datanhallinta - Datapolitiikka johtamisen ja käytännön työskentelyn väline: "tietoaaineistot tehokäyttöön" - Datanhallintaan liittyvien tietojärjestelmien kehittämisestä vastaa Tietojärjestelmäyksikkö - Suuri osa datanhallinnasta tapahtuu substanssiyksiköissä yhteistyössä asiantuntijoiden kanssa - Aineistot Aineisto- ja järjestelmärekisterissä – luokitellulle tiedolle oma rekisteriverkko - Tutkimusdatan hallinnan järjestelmät vielä tällä hetkellä omassa konesalissa - Tietoaaineistojen käyttöä tehostetaan sisäisesti, kansallisesti (Tulanet) ja kansainvälisesti - Datan käsittely siiloutunut hallinnollisesti, menetelmällisesti, järjestelmissä sekä erilaisiin käyttöoikeuskäytäntöihin (yhdistely vaikeaa tai jopa mahdotonta mm. lupaprosessien takia) - Aineistojen metadata puutteellista, monet aineistot yksittäisten henkilöiden varassa - Laskenta - Tieteellisen laskennan asiantuntijaryhmä koordinoi laskentaan liittyviä asioita - CSC:n Pouta-palvelut käytössä - kohtalaisen kypsät laskennan ympäristöt - Erittäin paljon osaamista tieteellisen laskennan eri osa-alueilta - Laskentakapasiteettia riittämättömästi tulevaisuutta ajatellen, lisäkapasiteetin hankinta hidasta - Tietoliikenne - Valtori hallinnoi pääsääntöisesti tietoliikennepalveluita – Valtorin tarjoamat pilvipalvelut tulevaisuuden näkökulmasta periaatteessa hyvältä, mutta käyttöön saaminen hidasta ja kallista - Tietoliikenneneratkaisut pääsääntöisesti tukee THL:n tarpeita, mutta pilvipalveluiden (e- ja cPouta) hyödyntäminen lisääntyy tulevaisuudessa - CSC:n palveluissa paljon potentiaalia - Osaaminen - Osaamista on riittävästi, mutta se ei ole kaikilta osin optimaalisessa käytössä - kentän muutos asettaa jatkuvia paineita osaamisen kehittämiseen - Ei laskentaan erikoistunutta ylläpito-osaamista, erityisesti Valtorin osaaminen on heikkoa - Muuta - Hyvä yhteistyö laskennan, tietohallinnon ja substanssin välillä - Iso osa asiantuntijoiden työajasta menee aineistojen hallintaan - Laskentakapasiteetin niukkuus, jäykät rakenteet ja tekniikoiden käyttöönoton hitaus uhkia								
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

				GTK							
Luke	SYKE	VTT	THL	- Datanhallinta - GTK-tasoinen tietovarannon hallintamalli sekä ATT-perustaiset politiikat tutkimukselle ja datan hallinnalle olemassa - Koko tiedonhallinnan ketju hallussa, itse toteutettuna, kustannukset - Yhteistyökumppanit pääsevät hyödyntämään GTK:n dataa (esim. geologinen tietovaranto) verkkopalveluiden ja rajapintojen kautta - Uudet datalähteet (etenkin massamainen ja rakenteeton data) ja projektimuotoinen tieto kasvaa – dataa tuottavat prosessit muuttuneet ja osin poistuneet - Geotietokeskus, ulkoiset tietovirrat, verkostoituminen ja kumppanuudet sekä uudet teknologiat mahdollisuuksia - Laskenta - Laskentaa suhteellisen vähän, CSC:n palveluita hyödynnetty – aineistojen prosessointi haastavaa, moniulotteinen mallinnus lisääntyy - Uusi, mahdollistava infrastruktuuri on monipuolinen ja tehokas - Verkostojen ja kumppanuuksien kautta tukea löytyy - Valtion keskitetty tuotanto sekä kaupallisten lisenssien jäykkyys riski - Tietoliikenne - Valtorin palvelut käytössä, palvelu ei vastaa tulevaisuuden tarpeita - Selkeä tarve toimivalle kokonaisratkaisulle (tuottajana CSC/Valtori) - Osaaminen - Datanhallinnan osaaminen ja tekijät selkeä haaste, nopeat organisaatiomuutokset - Omaa laskentaosaamista on, mutta ei kattavasti - Matemaattisen laskennan ja algoritmien kehittämisosaamista tarvitaan - Strategisesti moniulotteisuus, monikanavaisuus, monimuotoisuus, yhteentoimivuus, automatisointi ja optimointi haastavat datanhallinnan ja laskennan kehittämisen - Muuta - Jatkossa suuret tietomassat (skannerit) ja erilainen kuvamateriaali kasvaa - Jatkossa Suomen kattavan laserkeilauksen tietoaineisto hallittava yhdessä paikassa - GTK olettaa saavan palvelulupauksen, jota se itse on määrittelemässä, testaamassa ja toteuttamassa	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
emassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

					IL						
Luke	SYKE	VTT	THL	GTK	- Datanhallinta - Datanhallinta hajautunut organisaation tasolla asiakaslähtöisesti (tutkimusprojekteille), datan saavutettavuuden haasteet - Eudat.eu-palvelupilotti CSC:n kanssa käynnissä - Oma INSPIRE-direktiivin mukainen avoimen datan palvelu käytössä - FMI-tutkimusdatapolitiikka työn alla, tutkimuksen avoin data määritelty sekä KV-politiikat tutkimusprojektien mukaan – strategisessa ATT-ajattelussa kehitettävää - Laskenta - Tutkimushankkeissa ulkopuolisten laskentaresurssien käyttö mahdollista (ECMWF, Prace, CSC...) - Kapasiteetti joustavasti käytettävissä, nopeasti priorisoitavissa eikä laskenta-aikaa tarvitse anoa - Yhden säikeen työt (seriaalityöt) syövät laskentanoodin, pitäisi kohdentaa kokonaan omaan kapasiteettiin - Oma strateginen laskenta-ajan priorisointi säilytettävä - Tietoliikenne - Tutkimuksen kannalta olennaiset korkean kapasiteetin laajaverkkoyhteydet Funetissa – muut laajaverkkoyhteydet ja esim. mittalaiteyhteydet Valtorilta (selkeä roolitus CSC:n ja Valtorin tuottamien tietoliikenneyhteyksien osalta); - Tietoliikennearkatkaisut palvelee FMI:n tarpeita; - Funetin kansainväliset ulottuvuudet Nordunet ja Geant ym. ovat keskeisessä roolissa, kansallisten ratkaisujen tulee huomioida kv. kehitys - Osaaminen - Pääsääntöisesti osaava henkilökunta, mutta käyttäjätukea tarvitaan enemmän vertaistuen lisäksi - Optimoinnin tukia puutteellista, laskenta-data-käyttäjät-ketju saatava kuntoon - Muuta - Miten resurssien allokatio hoidetaan tarve- eikä politiikkalähtöisesti? - Teknisille standardiratkaisuille tilausta esim. tutkimuksen avoin data - Operatiivisen toiminnan ja tutkimuksen raja datan osalta voi olla häilyvä, mahdollisuus liittää tutkimusdata osaksi operatiivista toimintaa tulee säilyä	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

						STUK					
Luke	SYKE	VTT	THL	GTK	IL	- Datanhallinta - Dataa säteilymittauksista (omat, talonsisäiset järjestelmät) hankkeissa toteutettaviin kyselytutkimuksiin (vain muutama henkilötyövuosi) - Datanhallinnalle ei STUK-tason käytänteitä - Säteilyvalvontaan liittyvä data verkkosivuilla ja IL:n järjestelmässä, TKI-data pyydettyäessä - Laskenta - Vähäisistä resursseista johtuen hyödynnetään yhteistyökumppaneiden (yliopistot, IL) resursseja ja jonkun verran ulkoisia palveluja - Jatkossakin tarve säilyttää jonkun asteinen laskentakapasiteetti ja osaaminen talossa - Tietoliikenne - Tarpeen mukaan toteutetut ratkaisut: yliopistoille ulkoisten pilvipalveluiden kautta sekä IL:lle tietoturallinen palvelu, myös ulkoinen kovalevy käytössä - Suurten tiedostojen siirto esim. STUKin ja yliopistojen välillä hankalaa, edellyttää ulkoisten pilvipalveluiden käyttöä, joiden tietoturva haaste - Tietoliikennemuutos Valtorin ja Funetin kautta tulossa, toimivuudesta ja palveluista ei vielä näkemystä - Osaaminen - Osaaminen vähentynyt datanhallintaan, laskentaan ja jopa substanssityöhön liittyvään laskentaan liittyen - Muuta - Yhteiskäyttö, sujuvat tietoliikenneyhteydet ja suurten datamäärien turvallinen siirto korostuvat jatkossa	EVIRA	TTL	MML	VATT	FIIA
emassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
							- Datanhallinta - Valvontaan ja tutkimukseen liittyviä aineistoja useassa eri tietojärjestelmässä (esim. KUTI, ELVI ja Elmo), joihin muilla kuin tiedon tuottajilla pääsääntöisesti rajattu pääsy - Riskiarvioihin liittyvät aineistot projektien omissa kansioissa useissa eri tiedostomuodoissa - Tiedon avoimuutta tuetaan mm. tiedolla johtamisen kautta ja julkaisemalla tuloksia Evira.fi-nettisivuilla - Oman ydintoiminnan näkökulmasta keskeiset rekisterit ja järjestelmät toimivat - Tiedonhallinta ei vielä kokonaisuutena täysin hallussa – ongelmia tiedon hyödyntämisessä - Raportointivelvoite EU-komissiolle, EFSA:lle ja OIE:lle - Datan analyysiä, saatavuutta ja julkaisemista parannetaan Qlik-analyysijärjestelmällä - Laskenta - CSC:n palvelut eivät laajasti käytössä – laskenta- ja tiedontallennuspalvelu ostettu kokogenomisekvenssointidatan osalta - Laskentojen säilytys paikallisesti RAID:ssa, muisti rajoitteena - EVIRALLA ei ole tarjota skaalautuvia laskentapalveluita (Valtorin kapasiteetti) - Tietoliikenne - Tietoliikennepalvelut toteutettu vaihtelevasti, perustuen yksittäisten tapausten tarpeisiin – Valtori-vetoista - Ulkoisten palveluiden käyttö haastavaa tietoturvan takia - Osaaminen - Vaatimukset kasvavat raportoinnin ja datan lisääntyessä – tarvittaisiin organisaatiotason linjauksia toiminnan kehittämiseksi - Muuta - Yhteistyö luo tehokkuutta, laajan tutkimusaineiston hyödyntäminen - tiedon jakaminen tutkimuslaitosten kanssa joustavaksi! - Tuotetun datan määrä kasvaa sekä sen tallennus ja käsittely vaatii uusia ratkaisuja - laskentapalveluiden kehittämisen ongelmana vähäinen käyttötarve - Valtorin palveluissa yleisesti ongelmia - paikallisen tuen merkitys suuri myös jatkossa - Tarve yhtenäiselle datanhallintajärjestelmälle				
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
								- Datanhallinta - Datanhallinta ISO-sertifioitu - Data: tilastotiede- ja aineistonhallinta- sekä projektiresursseja - Data- ja julkaisupolitiikka ATT-periaatteita mukaillen - Laskenta - Pilvi- yms. palveluiden käytön ohjeistus menossa - On-premises-tyyppinen suurteholaskentakapasiteetti tähän asti riittänyt - Tietoliikenne - Toteutettu kaupalliselta toimittajalta (ei Valtori), vastaa olemassa olevia tarpeita eikä kansallisia yhteistyötarpeita tällä hetkellä ole - Funetin palvelutarjonta ajoittain arvioinnin kohteena, palveluiden osittaista hankintaa (palvelu tai palvelumoduuli) kerrallaan tulisi edistää - Osaaminen - Osaaminen hyvällä tasolla, kilpailu kovaa - Strategia asettaa laajoja vaatimuksia, massadatan käytön laajeneminen keskeinen osa lyhyen tähtäimen strategista visiota - Muuta - Infrastrukturi on tähän asti ollut pääosin jakautunut ministeriörajojen mukaan, mikä on hankaloittanut palvelukehitykseen osallistumista sekä palveluiden käyttöönottoa OKM:n alalle kuulumattomien tutkimuslaitosten osalta			
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

									MML		
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	- Datanhallinta - MML tutkimuspolitiikka ja ATT-pohjainen Open Access -julkaisupolitiikka olemassa - Paikkatietoaineistot avoimia - Tutkimushankkeiden aineistot tutkijoiden tietokoneilla/MML:n palvelimilla – tutkimustiedon hallinta puutteellista - Operatiiviset datat (maasto- ja kiinteistötietokanta) hyvin ja kattavasti hallussa - Osa mittaustuloksista (esim. geodeettiset perusmittaukset) KV-tietokannoissa - ATT aktiivista, mutta tarvitsee pontta johdosta tutkijatasolle - Laskenta - Laskenta suurelta osin tutkijoiden omien koneiden varassa, CSC:n superkonekapasiteetti käytössä - MML tutkimuspolitiikka sekä oGIIR-hanke esimerkki paikkatiedon teholaskennasta - Tietoliikenne - MML-verkon ulkopuolinen tutkijaverkko - Neuvottelut yliopistojen kanssa menossa Paikkatietokeskuksen siirtymisestä Otaniemeen ehkä vuonna 2020 - Valtori-ratkaisut paikoin hyviä, paikoin hankaloittavat perustoimintaa – näkemys tulevaisuuden palvelutarpeesta heikko - MML-verkko pitää sisällään runsaasti luottamuksellista aineistoa - tietoturva - Osaaminen - Datanhallinnan ja laskennan osaamisessa parannettavaa - Suurteholaskennan ja pilvipalveluiden hyödyntäminen strategisesti tärkeää - Muuta - Mahdollisuutena tutkimusdatanhallintaohjelmisto yhdessä tutkimuslaitosten kanssa - Kustannukset haastaa kehitystyön – Tulanetin mahdollisuudet tarjota koulutusta ja yhteistyöfoorumia - Luokitellun tiedon hallinta haasteena - Räätälöitävien ratkaisujen mahdollisuudet jatkossa - Tarvitaan ymmärrystä siitä, miksi datanhallintaan ja laskentaan tulee jatkossa panostaa	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
										- Datanhallinta - Pirstaloitunut arkkitehtuuri, johon Valtorin paketoitua ratkaisua eivät istu (40 tutkijaa) - Oma palvelinympäristö vaatii uudistamista, kapasiteettiongelma - Laskenta N. 30 tutkijaa tekee raskasta tilastollista ja numeerista laskentaa vaativaa tutkimusta - Tilastollisen laskennan ympäristö aineistoriippuvainen (FIONA/VATTin omat laitteistot) - Laskentaresurssit (FIONA + muu toiminta) eivät riittäviä, pohdinnassa omien resurssien hankinta tehokkaan työn varmistamiseksi - Tietoliikenne - Valtorin tarjoamassa Tilastokeskuksen etäkäytössä suuria ongelmia – yhteys ei toimi (SYKE:llä vastaavia ongelmia, Valtori selvittää) - Osaaminen - Hyvin ohut ICT-resurssit Valtori/Valtti-operaation myötä - Muuta - Meneillään sisäinen selvitys (dl 12/2017), jossa pohditaan VATT:n tutkimusinfrastruktuurien tulevaisuutta – sisältää myös yhteistyöpohdinnan muiden tutkimuslaitosten ja/tai korkeakoulujen kanssa (GLC Helsinki, erityisesti HY ja Aalto) - Yhteistyö datanhallinnassa ja laskennassa nähdään tervetulleena ja juuri nyt on oikea aika asiaa pohtia	
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset nykytilakuvaukset

											FIIA
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	• Datanhallinta • Pohjaa julkiseen tietoon • Laskenta • Ei laskentatarvetta • Tietoliikenne • Ei erityisiä toiveita • Osaaminen • Valtion ICT-resurssit riittäviä
Olemassa olevaa CSC-yhteistyötä											

Sisältö

- Tietopohja ja yhteenveto
- **Nykytila-analyysi**
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - **Substanssikohtainen nykytilakuvaus**
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Keskeisimmät havainnot - datanhallinta

Datanhallinta	• Hallittava data voidaan pääsääntöisesti jakaa kahteen osaan: pitkäaikaisseurantoihin/valvontaan sekä hankkeissa muodostuvaan dataan – hankemuotoinen data kasvaa koko ajan haastaen organisatoriset datanhallinnan periaatteet ja datan avoimuuden • ATT-periaatteet ja -politiikat on pääsääntöisesti omaksuttu datanhallinnan kehittämisen ja käytön suuntaviivoiksi, kattavuus edellyttää edelleen organisatorista huomiota ja strategista johtamista • Valtori-muutos on aiheuttanut tutkimuslaitoksille merkittäviä muutoksia ICT-ympäristön ylläpidossa ja kehittämisessä • Valtorin tuottaman palvelutason ei nähdä vastaavan käytännön tarpeita kustannustehokkaasti • Suurilla tutkimuslaitoksilla mahdollisuus ja tarve säilyttää DL-osaamista kilpailukyvyn kehittämiseksi • Valtorin tarjoaman palvelinkapasiteetin lisäksi tutkimuslaitoksissa hallitaan merkittävä määrä omaa, niin pitkäaikaisseurantoihin liittyvää kuin hankkeissa syntyvää dataa • Niin kapasiteetista luopumisesta (elinkaari) kuin ATT-periaatteiden mukaisesta datanhallinnan käytänteiden kehittämisestä keskustellaan tutkimuslaitoksissa runsaasti • Avoimuus ja yhteistyö nähdään mahdollisuutena, johon on kuitenkin pystyttävä kohdentamaan resursseja (perusrahoituksen kaventuminen, hankerahan avulla ei tehdä strategista kehittämistä) • Tutkimuslaitosten ja/tai korkeakoulujen yhteiseen dataan liittyvä yhteistyö muutamia poikkeuksia (esim. geologinen tietovaranto, satelliittidatakeskus ja paikkatieto) lukuun ottamatta suhteellisen harvinaista
Laskenta	
	Tietoliikenne
Osaaminen	

Keskeisimmät havainnot - laskenta

Datanhallinta	
• Laskennan tarpeen ennakoidaan kasvavan jatkossa • Huomioitava kuitenkin, tarkoitetaanko laskennalla tehokannettavilla tehtävää, Valtorin kautta kolmansilta osapuolilta hankittavaa vai CSC:n tarjoamia laskentapalveluita • Raskasta laskentaa edellyttävien tarpeiden osuus koko laskennasta näyttäisi olevan suhteellisen pieni • Tapauskohtaiset kustannukset kylläkin merkittäviä (tutkimuslaitokset vs. yliopistot) • Tutkimuslaitokset käyttävät CSC:n laskentakapasiteettia kiertoteitse, yliopistojen kautta • Näyttäisi siltä, että tutkimuslaitosten vaikea osoittaa nykyinen laskentakapasiteetin käyttö saatikka määritellä tulevaisuuden tarpeet (ulkoisen laskentakapasiteetin käytön linjausta harvoin olemassa) • Laskennan tuki pääasiallisesti kollegiaalista • Valtori ei pysty vastaamaan tutkimuslaitosten laskentapalveluiden tarpeeseen kuin tarjoamalla kolmansien osapuolten valmiiksi kilpailutettuja palveluita – tietoturvasta huolehdittava • Työnjako CSC:n kanssa mahdollinen ja toivottava – Valtorin ja CSC:n selvitettävä ja kehitettävä palveluprosessit kuntoon yhden luukun -palveluperiaatteella • Lisenssiongelmien (akateeminen/viranomaiskäyttö/kaupallinen) ratkaistava CSC:n palveluiden osalta, koska tutkimuslaitoksella merkittävä määrä kaupallisessa tarkoituksessa tehtävää laskentaa • Yksittäisten datasettien hallinnasta pitäisi päästä yhteisiin datavarantoihin – data-analyysin, tekoälyn ja koneoppimisen aito hyödyntäminen voisi mahdollistua rajapinnoilla • Elinkeinoelämäyhteistyö jollakin tapaa nostettava esille!!	Laskenta
Tietoliikenne	
Osaaminen	

Keskeisimmät havainnot - tietoliikenne

	Datanhallinta	
	Laskenta	
Tietoliikenne	• Valtorin tarjoamat tietoliikenneyhteydet eivät palvele kasvavaa dataintensiivisen laskennan tulevaisuuden tarpeita • Tutkimustoiminnan tietoturvaan kohdistuva paine nousee esille suhteessa nykyiseen palveluun • Tietoliikenneyhteydet nousevat kysymykseen erityisesti maantieteellisesti hajautuneiden tutkimuslaitosten osalta • Funetin palveluihin ollaan tyytyväisiä (ne jotka käyttävät) • Globaaleihin tutkimusinfrastruktuureihin ja -aineistoihin pääsy nähdään paikoin kriittisenä • Olemassa olevissa datan hallintaan pohjautuvissa yhteistyörakenteissa myös tietoliikenteeseen liittyvät haasteet (kapasiteetti, tietoturva) on ratkaistu tapauskohtaisesti • Uusien avausten rakentaminen on monessa mielessä haastavaa • Pääsääntöisesti tietoliikenneyhteyksiä ei koeta ongelmaksi, jos organisaatiossa on oma vahva ICT-kehittämistoiminto • Valtorin myötä ICT-resurssit ovat pääsääntöisesti vähentyneet minimiin	
	Osaaminen	

Keskeisimmät havainnot - osaaminen

	Datanhallinta	
	Laskenta	
	Tietoliikenne	
• Valtorin myötä mahdollisuudet datanhallinnan ja laskennan osaamisen kehittämiseen ovat pirstaloituneet ja vähentyneet huomattavasti sekä kehittämisen vastuu siirtynyt käyttäjien kontolle • Käyttö- ja menetelmätukea ei monin paikoin saatavilla ollenkaan • Datanhallintaa ja laskentaa strategisesti linjanneet tutkimuslaitokset kokevat tilanteen siedettävänä, muttei kestäväenä • Substanssiosaaminen resursoitu kautta linjan datanhallinnan ja laskennan osaamista tärkeämmäksi • Datanhallinnan ja laskennan osaamista voidaan kehittää tutkijoille tai osaaminen/tuki voidaan hankkia (joko talon sisältä tai hankittuna ulkoa) – talokohtaiset linjaukset tarvitaan • Kilpailu osaajista kovaa samalla kun kehitysvauhti huimaa • Datanhallinnan ja laskennan strategisen tason kehittämisosaamisen puute voi johtaa datanhallinnan pirstaloitumiseen	Osaaminen	

Yleisiä huomioita toiminnan eri tasoilla



Muita huomioita ja selvitettäviä kokonaisuuksia

- CSC:n ja Valtorin yhteistyön ja työnjaon kehittäminen – tavoitteena palveluiden tarjonta yhden luukun -periaatteella
 - **Datanhallinta:** tutkimuskäytön osalta ei päällekkäisyyksiä – työnjako
 - **Laskenta:** Valtorilla ei ole (tarjoaa kylläkin kolmansien osapuolten (pilvi)palveluita) – työnjako
 - **Tietoliikenne:** VY- ja Funet-verkon integraatio toimii – yhteistyö tehokkaiden tieteellisen laskennan edellyttämien yhteyksien toteuttamiseksi
 - **Tuki/osaaminen:** ei päällekkäisyyksiä – työnjako
- Mitä CSC:n palveluiden avaaminen tutkimuslaitoksilla tarkoittaa käytännössä OKM:n, muiden ministeriöiden sekä tutkimuslaitosten osalta?
- Yleisen tietoisuuden lisäämiseksi on hyvä harkita, miten
 - tutkimuslaitosten datanhallinnan ja laskennan strategista johtamista voidaan tukea?
 - lisään tietoisuutta olemassa olevista palveluista?
 - Tietohallintojohtajien verkoston muodostaminen? Yhdessä korkeakoulujen kanssa?
 - lisään tutkimuslaitosten ja korkeakoulujen tietoisuutta olemassa olevasta datasta ja sen hyödyntämismahdollisuuksista?
- HUOM! Tässä dokumentissa esitetty nykytilakuvaus, olemassa olevat taustamateriaalit ja -tiedot sekä erikseen sovittavat lisäselvitykset toimivat pohjatietona datanhallinnan ja laskennan kehittämiseksi tehtäville kehittämisohjelmille

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- **Laskentakapasiteetin yksilöity tarve**
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

1) Laskentakapasiteetin yksilöity tarve

- Alla olevissa lisätietokysymyksissä käytetään seuraavaa jaottelua:
 - Tehokannettavat
 - Omat laskentapalvelut ja/tai kolmansien osapuolten pilvipalvelut
 - Suurtehotelaskentakapasiteetti (CSC tms.)
- Nimeä (henkilön nimi, organisaatio, tutkimuksen aihe) oman tutkimuslaitoksesi merkittävimmät vastuulliset tutkijat/hankejohtajat, jotka ovat suurteholaskentakapasiteetin merkittävimpiä käyttäjiä
 - VASTAUS:
- Nimeä (tutkimuksen aihe) keskeisimmät tutkimuslaitoksen oman tai kolmansilta osapuolilta hankittavien laskentapalveluiden hyödyntäjät
 - VASTAUS:
- Arvioi tutkimustoiminnassa käytettävän suurteholaskennan ja muiden laskentapalveluiden keskinäistä suhdetta
 - Suurteholaskennan osuus on ____ %
 - Muiden laskentapalveluiden osuus on ____ %
- Miten suurteholaskennan ja muiden laskentapalveluiden suhde tulee kehittymään tulevaisuudessa?
 - VASTAUS:

2) Datanhallintapilottien esitykset

- Nykytilaselvityksen aikana ilmeni tarve tiivistää tutkimuslaitosten ja korkeakoulujen yhteistyötä datanhallintaan hallintaan ja laskentaan liittyen pilotoinnin keinoin
 - Pilottien avulla voidaan päästä kiinni niihin haasteisiin, jotka tulee ratkaista datanhallintaan ja laskentaan liittyvien palveluiden käytössä
 - Taustalla Suomen valtion keväällä 2017 tekemä päätös investoida n. 35 MEUR datanhallinnan ja laskennan kehittämiseen
 - Pilotit toteutettaisiin mahdollisesti jo vuoden 2018 aikana
- Tutkimuslaitoksilta pyydetään tässä yhteydessä ehdotusta piloteiksi, jotka täyttävät seuraavat periaatteet
 - Datan näkökulmasta pilotit muodostavat temaattisia kokonaisuuksia, joiden ympärillä tutkimuslaitokset, korkeakoulut ja mahdollisesti myös yritykset ja elinkeinoelämä voisivat rakentaa uudenlaista yhteistyötä
 - Esimerkkeinä genetiikka, opintotietovaranto, paikkatieto
 - Piloteissa tulisi olla valmius tuoda nykyiset organisaatioperustaiset datasetit yhteen datavarantoon laajemman hyödyntämismahdollisuuksien aikaansaamiseksi
- Pilottien tavoitteena on toimintalähtöinen kehittäminen, jonka avulla
 - mahdollistetaan akuuttien tarpeiden täyttäminen
 - sitoutetaan toimijat (myös ministeriöt) yhteistyön positiivisten vaikutusten edistämiseen
 - kohdennetaan toimet nopeasti todellisiin tarpeisiin, niiden ratkaisuvaihtoehtoihin sekä esiin nouseviin ongelmiin

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- **Laskentakapasiteetin yksilöity tarve**
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Tutkimuslaitoskohtaiset lisätiedot

Luke											
- Merkittävimmät suurteholaskennan käyttäjät - Genetiikka (tallennustila: 50 TB, laskentayksikkö: 15.000.000 BU) ja kaukokartoitus (lähinnä keila-aineistot) - Mallinnus (spatiaaliset mallit) - Laskentapalveluiden hyödyntäjät - Suurimmat tarpeet genetiikassa, kaukokartoituksessa ja mallinnuksessa - Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta - Laskentakapasiteetin arviointi erittäin vaikeaa, hanketoiminnan kasvu aiheuttanut pirstaloitumista - Laskentakapasiteetti on tällä hetkellä riittämätöntä - Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta - Laskentakapasiteetin tarve kasvussa, suurta ryntäystä ei kuitenkaan odotettavissa - Pilottiesitykset - Tutkimuslaitosten tuottamien aineistojen (pitkäaikaisseurannat ja rekisteriaineistot) hallinta ja säilytys - Valtorin ja CSC:n tuottamien palveluiden yhteensovittaminen - Genetiikan laskennan pilotointi esim. seuraavissa projekteissa (sitien lisäresurssitarvetta ei ole) : Rumen metagonome study, comparative aligner study, arcArk transcriptome study	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

	SYKE										
Luke	- Merkittävimmät suurteholaskennan käyttäjät - Merialuemallinnus (CSC/Taito), bioinformatiikkalaskenta (CSC/Taito), Coherens – 3D-virtauslaskenta (CSC/Taito) - Laskentapalveluiden hyödyntäjät - geoinformatiikka/paikkatieto (IL-yhteistyö), vesistömallinnus (omat laskentapalvelimet), paikkatieto (omat laskentapalvelimet), päästöjen skenaariomallinnus (oma kannettava), monilähteinen seuranta-aineistojen hyödyntäminen ja datafuusio (hyödynnettävä infra auki) - Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta 20/80 - Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta - Suurteholaskennan tarpeen odotetaan lisääntyvän jonkin verran aineistomäärien ja mallinnuksen kehityksen myötä - Pilottiesitykset - Lähtöaineistojen ja datavirtojen toimintamallin määrittely ja testaus - Federoidut tutkimustietovarannot ja laskentapalvelut: olemassa olevan datan laajempi hyödyntäminen - Tutkimuslaitosten käyttötarpeet CSC:n ympäristölle	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

		VTT									
Luke	SYKE	• Merkittävimmät suurteholaskennan käyttäjät • structural dynamics and vibroacoustics, clean air and acoustic solutions, electrical powertrains and storage, systems engineering and simulation, structural integrity, reactor analysis, CFD modelling, material modelling and ecodesign, production host engineering, big data industrial application, smart health • Laskentapalveluiden hyödyntäjät • Energia,- teknologia-, biotekniikka- ja terveysteknologiaan liittyvä tutkimus, esim. laskennallinen virtausdynamiikka, rakenneanalyysi, multifysiikka-analyysit, genomi aineenvaihdunnalliset laskennat, molekyyliidynamiikka, matalan taajuuden sähkömagnetiikka, big data -teknologioiden hyödyntäminen, data-analytiikka ja koneoppiminen • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • CSC:n palveluiden käyttö olematonta, sopimus valmisteilla • 40/60 (VTT:n oman klusterin käyttö on suurteholaskentaa) • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Suurteholaskennan osuus kasvaa merkittävästi • Pilottiesitykset: Ei pilottihanke-esityksiä	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
emassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

			THL								
Luke	SYKE	VTT	• Merkittävimmät suurteholaskennan käyttäjät • Humaani genetiikka (FIMM-yhteistyö) • Muita aloja pilotointiasteella • Laskentapalveluiden hyödyntäjät • Soten ja maakuntauudistuksen ohjaus tuotteistetulla tiedolla, Whole genome sequencing in microbiology, PneuvoCarr kantajuustutkimukset • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • Iso osa raskaasta laskennasta omilla kannettavilla ja tehotyöasemilla, joista jälkimmäisistä pääosa on koottu löyhäksi verkostoksi • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Kasvaa, muttei rajatta • Pilottiesitykset • Isaacus – terveysalan aineistojen yhteiskäytettävyyden kehittäminen (aineistokatalogi.fi)	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

				GTK							
Luke	SYKE	VTT	THL	• Merkittävimmät suurteholaskennan käyttäjät • Rakoverkkomallinnus ja 3D-mallinnus, malmipotentialimallinnus, EM-mallinnus • Laskentapalveluiden hyödyntäjät • Kallioperä- ja malmimallinnus • Arvio suurteholaskennan ja nykyisestä laskentapalveluiden nykyisestä suhteesta • 2/98 • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Kasvaa arviolta 4 %:iin (GTK tulee virtualisoimaan työasemalaskentaa, laskennan siirto palvelimille) • Pilottiesitykset • Hyperspektriskannauksen ja muun materiaaliskannauksen kehittäminen ja aineistoon liittyvän tutkimustyön mahdollistaminen: tavoitteena toteuttaa skannaus-, analysointi- ja tallennus- sekä jakeluratkaisun kokeilu korkeakoulujen kanssa	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

					IL						
Luke	SYKE	VTT	THL	GTK	- Merkittävimmät suurteholaskennan käyttäjät - HUOM! Vastaukset eivät sisällä operatiivista laskentaa (ml. säämallit) - Ilmastotutkimus, ilmanlaadun mallinnus, kaukokartoitus, avaruustutkimus, merentutkimus - Laskentapalveluiden hyödyntäjät - Ilmastotutkimus: Ilmastojärjestelmä, globaali ja alueellinen ilmasto, hiilenkierto, ilmakehä, satelliittihavaintojen prosessointi - Ilmanlaadun mallinnus: paikallinen, alueellinen ja globaali ilmanlaatu, kaupunkimittakaavan ilmanlaatuun liittyvät ilmiöt, SILAM-ilmanlaatumalli - Kaukokartoitus: lukuisat eri kaukokartoitusmenetelmiä kehittävät tai kaukokartoitusmittauksia hyödyntävät kymmenet tutkimusprojektit (dataa teroja datanhallinta- ja tallennustarpeineen) - Avaruussää: avaruussään mallinnus, avaruustutkimus - Merentutkimus: merten uusanalyysit, arviot ilmastovaikutuksista merten tilaan - Arvio suurteholaskennan ja laskentapalveluiden suhteesta - Ilmastotutkimus: 80/20 (säilyy samana jatkossa) - Ilmanlaadun mallinnus: 70/30 (suurteholaskentatarve kasvaa) - Kaukokartoitus: 20/80 (säilyy samana) - Avaruussää: 80/20 (säilyy samana) - Merentutkimus: 80/20 (säilyy samana) - Pilottiesitykset - Suomalainen Satelliitti Data Cube - Suurten SAR-kuvamäärien matemaattinen analyysi (IL, MML, Aalto) - LES-mallin ja alueellisen ilmanlaatumallin yhdistäminen kaupunkimittakaavassa	STUK	EVIRA	TTL	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

Luke	SYKE	VTT	THL	GTK	IL	STUK					EVIRA	TTL	MML	VATT	FIIA
						• Merkittävimmät suurteholaskennan käyttäjät • Potilasdosimetrian ja mittanormaalitoiminnan kehittäminen • Laskentapalveluiden hyödyntäjät • Leviämislaskenta (yhteistyö IL:n kanssa), potilasdosimetrian ja mittanormaalitoiminnan kehittäminen (yhteistyö HY:n ja UTA:n kanssa) • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • Tietoja arvioinnin pohjaksi ei ole saatavilla • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Vaikea arvioida • Pilottiesitykset • Ei pilottihanke-esityksiä									
Olemassa olevaa CSC-yhteistyötä															

Tutkimuslaitoskohtaiset lisätiedot

Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	FIIA
							- Merkittävimmät suurteholaskennan käyttäjät - Suurteholaskentaa (CSC-taso) ei tällä hetkellä käytetä kuin kahdessa käyttöönotto-vaiheessa olevassa genomidatan (enimmäkseen bakteeri- ja virusgenomien) analysointiin keskittyvässä hankkeessa CSC-ympäristöissä Anniina Jaakkosen toimesta - Laskentapalveluiden hyödyntäjät - Tutkimuslaitoksen omat laskentapalvelut: Tehokannettavatase (riskiarvioinnin tutkimusyksikkö, kaikki data-analyysia hyödyntävät riskinarviointiprojektit): bayesilaiset tilastolliset mallit, MCMC-laskenta ja simulointi (elintarvikkeet ja eläimet), Monte Carlo -simulointi kannettavalla tietokoneella (kasvinterveyden riskinarviointi)), elintarvike- ja eläintoimialueet, tautivapauslaskelmat ja otoskokolaskenta, Eviran muut yksiköt, muu tilastollinen konsultointi - Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta 0/100 - Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta - Tulevaisuudessa genomidatan sekvensoitava näytemäärä kasvaa ja sen myötä kasvaa myös suurteholaskentakapasiteetin tarve - Suurteholaskennalle on tulevaisuudessa tarvetta, mutta tarpeen sijoittuminen talon sisällä on vielä jäsentymättä - Tulevaisuudessa WGS-aineistoja on analysoitava jossain laskentapalvelussa esim. CSC:llä, mutta toiveena olisi, että bioinformatikko käsittelisi raakadatan ennen kuin se tulisi riskinarvioinnin käsittelyyn - Pilottiesitykset - kokogenomisekvensointi datan analyysi - Kansallisen datapankin (EFSA:n Data Warehousea vastaavan kansallisen systeemin) perustaminen ja kehittäminen - Laskurin kehittäminen saanninarviointien tekemiseksi kuluttajan nettilomakkeeseen täyttämien ruoankäyttötietojen pohjalta - Ihmisille tautia aiheuttavien bakteerien genomidatan hallinta ja jakaminen Eviran ja THL:n kesken sekä avaaminen tutkimuskäyttöön				
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

								TTL			
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	• Merkittävimmät suurteholaskennan käyttäjät • CSC:n palvelut eivät ole käytössä • Laskentapalveluiden hyödyntäjät • Työaikatutkimus, Kunta- ja sairaalatyöntekijöiden hyvinvoinnin seuranta, Human Guided Data Analysis, Kansallisten kohorttiaineistojen analysointi (useita), Beyond Zen, Työntekijöiden fysiologiset mittaukset useissa projekteissa • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • 0/100 • Pidempi laskenta tapahtuu TTL:n omalla suurteholaskentainfralla sekä tehokannettavilla lokaalisti • Paineita työntää osa pienlaskennasta ja esikäsittelystä lähelle dataa (pilveen) • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Tällä vuosikymmenellä TTL:n laskentatarpeet pystytään täyttämään omalla infralla • Laskentatarpeet tulevat kasvamaan hallitusti • Pilottiesitykset • Ohjelmistoinfraskruktuurin rakentaminen henkilökohtaisten IoT-mittalaitteiden datan keräämisen keskittämiseen ja datan yhdistämiseen tietoturvalisessa ympäristössä (osana useampaa tutkimusprojektia)	MML	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

									MML		
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	- Merkittävimmät suurteholaskennan käyttäjät - Avoin paikkatiedon tutkimusinfra oGIIR - Laskentapalveluiden hyödyntäjät - Muutostulkinta ESA:n Sentinel-2-satelliittikuvilta - Paikkatietojen käsittely ja monien www-pohjaisten paikkatietopalveluiden pilotointi - Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta (<i>henkilökohtaiset koneet 55 %, omat laskentapalvelut 10 %, kolmannen osapuolen pilvipalvelut 0 %, CSC:n pilvipalvelut 25 %, CSC:n suurteholaskentapalvelut 5 %</i>) 5/95 (suurteholaskenta eräajoja) - Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta - Tulevaisuudessa pilvipalveluita ja suurteholaskentakapasiteettia tullaan käyttämään yhä enemmän etenkin suurten paikkatietoaineistojen käsittelyyn. Henkilökohtaiset koneet tulevat kuitenkin olemaan myös erittäin tärkeitä tutkijoiden töissä, joissa kehitetään menetelmiä ja käytetään pienempiä aineistoja. Lisäksi on monia tutkimusalueita, jotka vaaditaan reaaliaikaista datojen käsittelyä ja dedikoituja ratkaisuja (esimerkkeinä FinREF –GNSS laskenta, älyliikennetutkimus) - Pilottiesitykset - Paikkatietoaineistojen yhtenäistäminen, GeoCubes (MML, Luke., SYKE, GTK...)	VATT	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

										VATT	
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	• Merkittävimmät suurteholaskennan käyttäjät • Ei saatavilla • Laskentapalveluiden hyödyntäjät • Ei saatavilla • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • Ei saatavilla • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Ei saatavilla • Pilottiesitykset • Ei saatavilla	FIIA
Olemassa olevaa CSC-yhteistyötä											

Tutkimuslaitoskohtaiset lisätiedot

											FIIA
Luke	SYKE	VTT	THL	GTK	IL	STUK	EVIRA	TTL	MML	VATT	• Merkittävimmät suurteholaskennan käyttäjät • Ei ole • Laskentapalveluiden hyödyntäjät • Ei ole • Arvio suurteholaskennan ja laskentapalveluiden nykyisestä suhteesta • Ei ole • Arvio suurteholaskennan ja laskentapalveluiden tulevaisuuden suhteesta • Ei ole • Pilottiesitykset • Ei ole
Olemassa olevaa CSC-yhteistyötä											

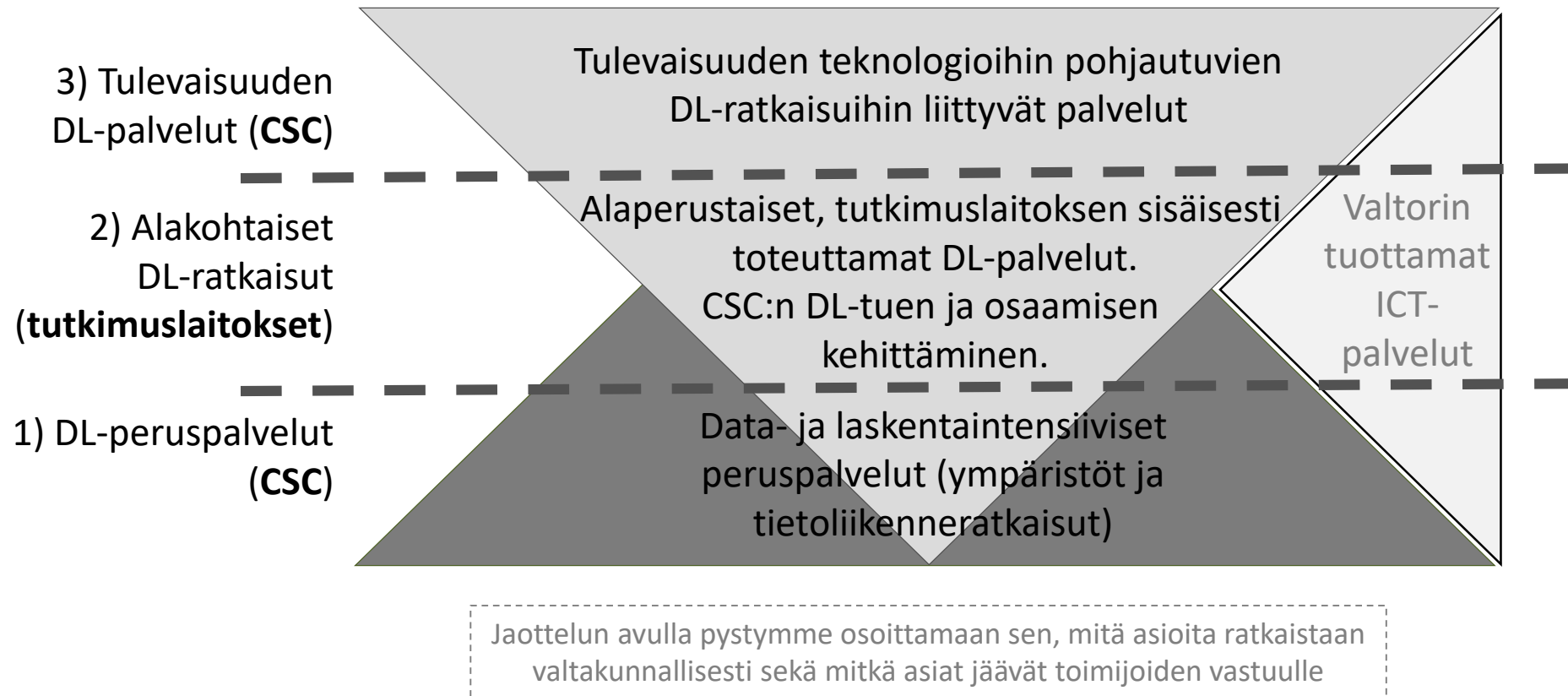
Alustavat johtopäätökset

- Tutkimuslaitosten oman arvion mukaan
 - laskenta- ja dataintensiivisen kapasiteetin tarve on kokonaisuutena kasvussa
 - kapasiteettitarpeen tulevaisuuden tarpeen ja jakautumisen (laskenta- vs. dataintensiivinen tutkimus) yksityiskohtainen arviointi on haasteellista
 - suurteholaskennan tarpeet ovat hienoisessa kasvussa, mutta kokonaisuutena varsin maltillisia
 - kansallisiin ja kansainvälisiin suurteholaskentaresurssiin tukeutuvaa toimintaa on ja sitä tulee olemaan myös jatkossa erityisesti IL:ssa, VTT:ssa, SYKE:ssä ja LUKE:ssa
 - yhteistyön lisäämiseen on tarvetta, joka näkyy mm. suurena monitoimijaisten datanhallinta ja laskentapilotti-ehdotusten määränä
 - DL-palveluissa tehtävän yhteistyön avulla voidaan tukea suomalaisen tieteen kehittymistä kokonaisuutena
- On huomattava, että tutkimuslaitosten DL-toiminnan täysin luotettavan kokonaiskuvan muodostaminen on haastava tehtävä
 - Selvityksen puitteissa on luotu yleiskuvaus, jonka tutkimuslaitokset ovat kahdessa selvityksen vaiheessa (nykytilakuvaus ja lisätietopyynnöt) vahvistaneet omien tietojen osalta

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- **Toimenpide-ehdotukset**
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Tavoitteena DL-intensiivisten palveluiden kansallinen kehittäminen



1) DL-peruspalvelut

Peruspalvelut		
• Tilanne/haaste • Valtorin tarjoamat kolmansien osapuolten DL-palvelut eivät palvele kansallisesti tutkimuslaitosten tarpeita (esim. julkisten pilvipalveluiden tietoturva) • DL-palveluiden osalta korkeakoulut ovat tutkimuslaitoksia paremmassa asemassa • Suomalainen tutkimustieto ja tutkimusta koskeva tieto on pirstaloitunut organisatorisiin, kansallisiin ja kansainvälisiin tietovarastoihin • Tavoitetilannekuva • Tutkimuslaitokset ovat tasavertaisessa asemassa korkeakoulujen kanssa kansallisten DL-peruspalveluiden saatavuudessa ilman lisäkustannuksia • Ratkaisuehdotus • Käynnistetään OKM/Juhta-vetoisesti Valtorin ja CSC:n yhteistyöneuvottelut tutkimukseen liittyvien DL-palveluiden järjestämiseksi yhden luukun periaatteella • Keskitetään kansallinen tutkimustiedon ja tutkimusta koskevan tiedon kokonaisvaltainen koordinaatio/hallinta CSC:lle • Käynnistetään monitoimijaiset, substanssiperustaiset DL-pilotit (esitys tämän kalvosarjan lopussa) yhteistyön edistämiseksi • Viestitään olemassa olevista palveluista nykyistä tehokkaammin • Yhteistyön pääasiallinen kohdentuminen • Tasavertaisten yhteistyöedellytysten aikaansaaminen	Alakohtaiset ratkaisut	Tulevaisuuden palvelut

2) Organisaatiospesifit ratkaisut

	Alakohtaiset ratkaisut	
Peruspalvelut	• Tilanne/haaste • Tutkimuslaitosten kyky strategisen DL-kehittämiseen on heikentynyt (ICT-palveluiden ulkoistus Valtorille sekä budjettirahoituksen väheneminen ja hankerahoituksen kasvu) • Tutkimuslaitosten resurssien pienentyminen on tarkoittanut priorisointia (tutkimuksellisen) substanssiosaamiseen, jolloin ICT-resurssit ja osaaminen eivät riitä oman kattavan ja ajantasaisen DL-ympäristön ylläpitämiseen ja kehittämiseen • Tavoitetilannekuva • Tutkimuslaitosten DL-kehittäminen on strategisesti johdettua, asianmukaisesti resursoitua sekä perustuu tarkoituksenmukaisin osin kansallisiin ja omiin DL-ratkaisuihin • Ratkaisuehdotus • Vapautetaan tutkimuslaitoskohtaisten strategisten valintojen kautta kehitysresursseja tutkimuslaitosten omien, alakohtaisten DL-ratkaisujen kehittämiseen käynnistämällä kansallinen yhteistyö perus- ja tulevaisuuden DL-palveluissa • Kehitetään CSC:n roolia DL-intensiivisen, alakohtaisen tutkimuksen osaamisen ja tuen toteuttamisessa • Luodaan yhteistyöfoorumi tutkimuslaitosten ICT-johtajille (ml. DL) sekä saatetaan heidät yhteen korkeakoulujen vastaavien verkostojen kanssa (FUCIO/AAPA) • Yhteistyön pääasiallinen kohdentuminen • Tietohallinnon strategisen johtamisen tuki • DL- ja substanssiosaamisen sekä niihin liittyvän tuen tasapainon varmistaminen tutkimuslaitoksissa, korkeakouluissa ja CSC:llä	Tulevaisuuden palvelut

3) Tulevaisuuden palvelut

		Tulevaisuuden palvelut
Peruspalvelut	Alakohtaiset ratkaisut	• Tilanne/haaste • Yksittäisillä tutkimuslaitoksilla ei ole mahdollisuutta investoida ja siten hyötyä (tai ylipäättään olla tietoisia) uusimmasta DL-teknologiasta • Tutkimuslaitosten DL-kyvykkyys vaikuttaa olevan hyvin operatiivista, jolloin DL-toimintaan liittyvää visionääristä johtamisosaamista on kehitettävä • Suomalaisen tieteen ja tutkimuksen kokonaisvaltainen johtaminen edistäisi myös kansallisen DL-palveluiden kehittämistä (ja toisinpäin) • Tavoitetilannekuva • CSC tarjoaa tulevaisuuden kokonaisvaltaiset ja joustavat DL-palvelut kansallisesti johdetun suomalaisen tieteen ja tutkimuksen tarpeisiin • Ratkaisuehdotus • CSC:lle varmistetaan kyky kehittää ja tarjota Aallon harjalla -palveluita korkeakouluille ja tutkimuslaitoksille • Selvitetään keinot koordinoida ja tukea erityisesti DL-intensiivisen huippututkimuksen ja niiden palvelutarpeiden kehittymistä (kansallisen koordinaation ja yksittäisten organisaatioiden autonomian haaste) • Yhteistyön pääasiallinen kohdentuminen • Tulevaisuuden toimintaedellytysten ja kilpailukyvyn varmistaminen

Valtorin ja CSC:n yhteistyö ja työnjako – tavoitteena yhden luukun palveluperiaate

- **Valtori ja CSC** ovat Suomen valtion omistamia ICT-yhtiöitä, joiden **perustehtävät ovat erilaisia, mutta toisiaan tukevia**
 - Valtorin tehtävänä on tuottaa valtionhallinnon toimialariippumattomat ICT-palvelut mm. tutkimuslaitoksille
 - CSC tuottaa ICT-asiantuntijapalveluita pääasiassa korkeakouluille ja tutkimukselle
- **Tutkimuslaitosten ja korkeakoulujen saamat DL-palvelut eivät ole tällä hetkellä tasavertaisia** (esim. hinnoittelu)
 - Valtorin ja CSC:n ICT-palveluiden yhteistyötä ja työnjakoa edistämällä voidaan tukea suomalaisen huippututkimuksen yhteistyön ja vaikuttavuuden kasvua

Valtori	Toiminnan osa-alue	CSC
Muun kuin tutkimusdatan hallinta	Datanhallinta – työnjako	Tutkimusdatan hallinta - Kansallisesti hallittavan datan merkitys (tietoturva)
Valtorilla ei laskentakapasiteettia - Ensisijaisesti CSC:n laskentapalveluiden tarjoaminen ja toissijaisesti kolmansien osapuolten tai on-premise-kapasiteetti	Laskenta – työnjako	Laskennallisen tieteen kansalliset ympäristöt ja -palvelut
VY-verkko - liityntäpalvelut valtion yhteisiin palveluihin, palveluntarjoajiin ja internetiin	Tietoliikenne – yhteistyö	FUNET-verkko - VY/Funet-integraatio toimii; tehokkaat DL:n edellyttämät (kaista, tietoturva...) yhteydet
Perinteiset, toimialariippumattomat tietohallinto-/ICT-palvelut	Tietohallintopalvelut – työnjako	DL:n tuki- ja osaamispalvelut - Tarve ja mahdollisuus vahvistaa roolia myös alakohtaisessa DL-osaamisessa ja -tuessa

- **Eri ministeriöiden alaisten tutkimuslaitosten mahdollisuutta** hyödyntää OKM:n alaisen korkeakoululaitoksen (yliopistot ja AMKit), pääasiassa **CSC:n tuottamia DL-palveluita voidaan tukea Valtorin ja CSC:n yhteistyön ja työnjaon tiivistämisen kautta** datanhallinnassa, laskennassa, tietoliikenteessä ja tietohallintopalveluissa (kuvio)
 - Yhteistyön tiivistäminen edellyttää jatkoselvitystä ja käsittelyä Juhta:ssa vuoden 2018 aikana
 - Yhteistyö- ja työnjakolinjaus olisi hyvä saada aikaiseksi ennen hankittavan laskentaympäristön käyttöönottoa

Yhteenveto DL:n keskeisimmistä haasteista ja kehityskohteista

1	2	3	4
Tietohallintopalveluiden yhteistyö ja työnjako	DL-peruspalveluiden kansallinen kehittäminen	Ala- ja organisaatio-kohtaiset DL-ratkaisut	Tulevaisuuden kansalliset DL-palvelut
- Tutkimuslaitosten ICT-palvelut ja -osaaminen pirstaloituneet, DL:n strategisessa johtamisessa kehitettävää - Valtion omistamien Valtorin ja CSC:n yhteistyötä ja työnjakoa on syvennettävä tutkimuslaitosten ja korkeakoulujen DL-palveluiden sekä tutkimuksellisten yhteistyöedellytysten parantamiseksi	- DL-palveluiden osalta korkeakoulut ovat tutkimuslaitoksia paremmassa asemassa - CSC:lle organisatorisesti, kansallisesti ja kansainvälisesti pirstaloituneen tutkimus- ja tutkimusta koskevan tiedon kokonaiskoordinaatiovastuu - Tuetaan kansallisten DL-peruspalveluiden kehittämistä pilottien avulla	- Tutkimuslaitosten strategista DL-kehittämiskykyä on vahvistettava - Omien DL-ympäristöjen lisäarvo perustuu strategisten kilpailutekijöiden vahvistumiseen, johon resurssit ovat löydettävissä kansallisesta perus- ja tulevaisuuden DL-ratkaisuissa tehtävästä yhteistyöstä	- CSC:n kyky ja mahdollisuus pysyä teknologian aallonharjalla varmistettava - Visionääristä DL-johtamiskyvykkyyttä lisättävä toiminnan kaikilla tasoilla - DL-intensiivisen huippututkimuksen palvelutarpeisiin vastattava toimintaedellytysten, kilpailukyvyn ja vaikuttavuuden parantamiseksi

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- **Toimenpide-ehdotukset**
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Tutkimuslaitosten merkittävimmät laskenta- ja dataintensiivisten palveluiden tutkimusalueet

Laskentaintensiivinen tutkimus

- **Luke:** Genetiikka, kaukokartoitus, mallinnus
- **SYKE:** Merimallinnus, bioinformatiikka, 3D-virtauslaskenta
- **VTT:** Structural dynamics and vibroacoustics, clean air and acoustic solutions, electrical powertrains and storage, systems engineering and simulation, structural integrity, reactor analysis, CFD modelling, material modelling and ecodesign, production host engineering, big data industrial application, smart health
- **THL:** Humaani genetiikka
- **GTK:** Rakoverkkomallinnus ja 3D-mallinnus, malmipotentialimallinnus, EM-mallinnus
- **IL:** Ilmastotutkimus, ilmanlaadun mallinnus, kaukokartoitus, avaruustutkimus, merentutkimus
- **STUK:** Potilasdosimetrian ja mittanormaalitoiminnan kehittäminen
- **Evira:** Genomidatan analysointi
- **MML:** Avoin paikkatiedon tutkimusinfra oGIIR

Dataintensiivinen tutkimus

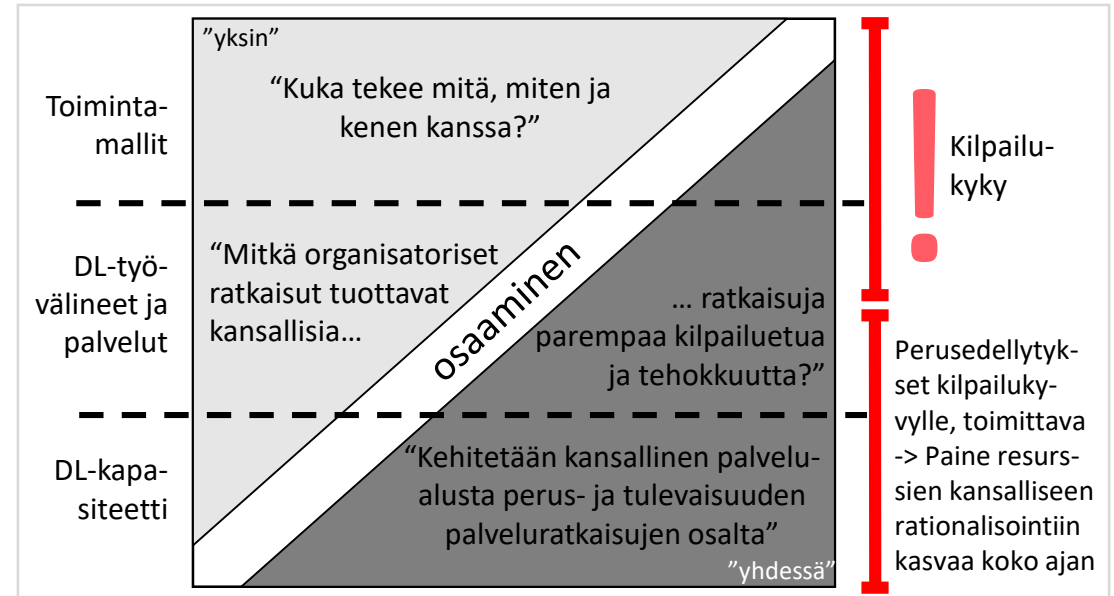
- **Luke:** Genetiikka, kaukokartoitus, mallinnus
- **SYKE:** Paikkatieto, vesistömallinnus, päästöjen skenaariomallinnus, monilähteinen seuranta-aineistojen hyödyntäminen ja datafuusio
- **VTT:** Energia-, teknologia-, biotekniikka- ja terveysteknologiaan liittyvä tutkimus, esim. laskennallinen virtausdynamiikka, rakenneanalyysi, multifysiikka-analyysit, genomi aineenvaihdunnalliset laskennat, molekyyliidynamiikka, matalan taajuuden sähkömagnetiikka, big data -teknologioiden hyödyntäminen, data-analytiikka ja koneoppiminen
- **THL:** Soten ja maakuntauudistuksen ohjaus tuotteistetulla tiedolla, whole genome sequencing in microbiology, PneuvoCarr kantajuustutkimukset
- **GTK:** Kallioperä- ja malmimallinnus
- **IL:** Ilmastotutkimus, ilmanlaadun mallinnus, kaukokartoitus, avaruustutkimus, merentutkimus
- **STUK:** Leviämislaskenta, potilasdosimetria ja mittanormaalitoiminnan kehittäminen
- **Evira:** Bayesilaiset tilastolliset mallit, MCMC-laskenta ja simulointi (elintarvikkeet ja eläimet), Monte Carlo -simulointi (kasviterveyden riskiarviointi), elintarvike- ja eläintoimialueet (tautivapauslaskelmat ja otoskokolaskenta)
- **TTL:** Työaikatutkimus, kunta- ja sairaalatyöntekijöiden hyvinvoinnin seuranta, human guided data analysis, kansallisten kohorttiaineistojen analysointi, beyond Zen, työntekijöiden fysiologiset mittaukset
- **MML:** Muutostulkinta ESA:n Sentinel-2-satelliittikuvilta, paikkatietojen käsittely ja monien WWW-pohjaisten paikkatietopalveluiden pilotointi

Tutkimuslaitosten oma näkemys DL-intensiivisen tutkimuksen tulevaisuudesta

- Laskenta- ja dataintensiivisen **kapasiteetin tarve on kokonaisuutena kasvussa**
 - **Suurteholaskennan tarpeet** ovat **hienoisessa kasvussa**, mutta **kokonaisuutena varsin maltillisia**
 - kansallisiin ja kansainvälisiin suurteholaskentaresursseihin tukeutuvaa toimintaa on ja sitä tulee olemaan myös jatkossa erityisesti IL:ssa, VTT:ssa, SYKE:ssä ja LUKE:ssa
 - **Hankkeissa syntyvän datan määrä kasvaa jatkuvasti**, joka tarkoittanee dataintensiivisen tutkimuksen määrän kasvua suhteessa laskentaintensiiviseen tutkimukseen
 - Vastaavasti **tutkimuslaitosten, korkeakoulujen sekä elinkeinoelämän tiivistyvän yhteistyön** kautta heterogeenisen datan määrä tulee kasvamaan ja siten aiheuttamaan lisätarpeita niin kapasiteetille kuin osaamiselle
 - **Tutkimuslaitosten DL-kapasiteetti on monin paikoin elinkaarensa päässä**
- **DL-palveluissa tehtävän kansallisen yhteistyön avulla voidaan tukea suomalaisen tieteen kehittymistä kokonaisuutena**
 - Selkeä **tarve toimivalle DL-kokonaisratkaisulle sekä strategiselle johtamiselle** (ml. avoin tiede) kansallisella ja organisatorisella tasolla
 - **Tapauskohtaisista DL-yhteistyörakenteista ja -ratkaisuista halu päästä keskitettyyn, joustavaan ja kehittyvään kokonaisratkaisuun**
 - data-analyysin, tekoälyn ja koneoppimisen aito hyödyntäminen voisi mahdollistua rajapinnoilla
 - **Kansainvälisen yhteistyön edellytykset on turvattava** (ml. riittävät ja tietoturvalliset dataliikenneyhteydet)
 - **Kansallinen datanhallinnan tärkeys** (vs. kolmansien osapuolten palveluissa hallittu data) on (poliittisesti) **ratkaistava**
- Budjettirahoituksen pieneminen ja hankerahoituksen kasvu ovat merkittävästi **heikentäneet DL:n organisatorisen strategisen kehittämisen kyvykkyyttä**

Esitys DL-investoinnin suuntaviivoiksi

- **DL-investoinnin tavoitteena** tulee olla **suomalaisen DL-intensiivisen tieteen kannalta toimiva kokonaisratkaisu**, jossa on **huomioitava seuraavat näkökulmat**
 - DL-kapasiteetti, -työvälineet ja -palvelut
 - DL-investoinnin **teknologinen ratkaisu ei ole** tutkimuslaitosten näkökulmasta **oleellinen kysymys** kuin tietyin reunaehdoin (KV-ulottuvuus)
 - DL-kapasiteetin tulee olla **riittävän joustava** niin **laskentaintensiivisen** (ml. GPU) kuin **kasvavan ja monimuotoistuvan dataintensiivisen tieteen** tarpeisiin
 - DL-työkalujen ja palveluiden kehityksen **avulla** tulee **tukea** avoimen **datan hyödynnettävyyttä tutkimuksen rajapinnoilla** esim. data-analyysin, tekoälyn ja koneoppimisen keinoin
 - Toimintamallit ja osaaminen
 - Toimintamallina erityisesti **DL-intensiivistä tutkimusta tekevien organisaatioiden** (korkeakoulut, tutkimuslaitokset sekä elinkeinoelämä) **yhteistyön kansalliselle koordinaatiolle** on perusteet niin tutkimuksen kuin siitä syntyvän tiedon hyödynnettävyyden näkökulmista
 - Tarve niin **kansalliselle** perus- kuin hankerahoituksella syntyvän datan yhtenäiselle, käytännönläheiselle (huomio mm. liikesalaisuuksien suojaustarpeet sekä tutkimuslaitosten omat aineistopolitiikat) **datanhallintapolitiikalle**
 - Olemassa olevien **hyvien käytänteiden** tietoisuutta tulee lisätä, levittää ja edelleen vahvistaa
 - DL:n **strategisen merkityksen tietoisuutta** tulee lisätä **kansallisella ja organisatorisella** tasolla pitkäjänteisen ja vaikuttavan kehittämisen mahdollistamiseksi
 - Kaikenlaiset **investoinnit DL-osaamiseen ovat tärkeitä** ja perusteltuja
 - **Avoimen tieteen ja datan** luomien **mahdollisuuksien integrointia korkeakoulujen opetukseen** tulee **vauhdittaa**



Yhteenveto tutkimuslaitosten tarpeita tukevasta DL-investoinnista

1

DL-intensiivinen huippu- tutkimus Suomessa

- Tutkimuslaitosten nykyinen DL-intensiivisen tutkimusrahoituksen määrä varsin pieni
- **Tutkimuslaitosten** laskenta- ja erityisesti dataintensiivisen **laskentakapasiteetin tarve** kokonaisuutena **kasvussa**
- DL-laiteinvestoinnissa painottettava myös **toimintamallien ja osaamisen** kehittämistä
- **DL-osaamisella tärkeä rooli lähes kaikessa tutkimuksessa**

2

Tutkimuslaitosten näkemys DL-investoinnista

- Budjettirahoituksen laskun ja hankerahoituksen kasvun tasapaino ratkaistava osana strategista DL-kehitystä
- Tutkimusorganisaatioiden **avoimen datan** ympärillä tehtävää **yhteistyötä kehitettävä kokonaisuutena**
- **Osaamisinvestoinnit** tärkeitä
- Tavoitteena **keskitetty, joustava ja kehittynyt DL-kokonaisratkaisu**

3

Varsinainen
DL-investointi

DL-kapasiteetti-, -työvälineet ja -palvelut

- **DL-investoinnin palveltava laskenta-** (ml. GPU) **sekä** kasvavan ja monimuotoistuvan **dataintensiivisen tieteen tarpeita**
- DL-palvelualustateknologian **tuettava datan hyödynnettävyyttä tutkimuksen raja-pinnoilla** (ml. PPP)
- **KV-yhteistyön** edellytykset on **turvettava** (ml. data-liikenteen tietoturva)

4

Toimintamallit ja osaaminen

- DL-intensiivistä tutkimus-yhteistyötä tuettava mm. tarkoituksenmukaisella **datanhallintapolitiikalla**
- **Hyviä käytänteitä** edelleen vahvistettava ja levitettävä
- DL-intensiivinen tutkimus ja siitä syntyvä data on **integroitava** korkeakoulujen **opetukseen** tehokkaasti
- DL:n **strategisen merkityksen tietoisuutta lisättävä**

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- **Toimenpide-ehdotukset**
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- Huomioitavaa jatkossa
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Pilottien tausta ja periaate

- Nykytilaselvityksen aikana ilmeni tarve tiivistää tutkimuslaitosten ja korkeakoulujen yhteistyötä datanhallintaan ja laskentaan liittyen pilotoinnin keinoin
 - Pilottien avulla voidaan päästä kiinni niihin haasteisiin, jotka tulee ratkaista datanhallintaan ja laskentaan liittyvien palveluiden käytössä
 - Pilotit toteutettaisiin mahdollisesti jo vuoden 2018 aikana siten, että palvelumallit saadaan otettua käyttöön uuden DL-palvelualustan käyttöönoton yhteydessä
- Tutkimuslaitoksilta pyydettiin ehdotusta piloteiksi, jotka täyttävät seuraavat periaatteet
 - Datan näkökulmasta pilotit muodostavat temaattisia kokonaisuuksia, joiden ympärillä tutkimuslaitokset, korkeakoulut ja mahdollisesti myös yritykset ja elinkeinoelämä voisivat rakentaa uudenlaista yhteistyötä
 - Piloteissa tulisi olla valmius tuoda nykyiset organisaatioperustaiset datasetit yhteen datavarantoon laajempien hyödyntämismahdollisuuksien aikaansaamiseksi
- Pilottien tavoitteena on toimintalähtöinen kehittäminen, jonka avulla
 - mahdollistetaan akuuttien tarpeiden täyttäminen
 - sitoutetaan kaikki toimijat yhteistyön positiivisten vaikutusten edistämiseen
 - kohdennetaan toimet nopeasti todellisiin tarpeisiin, niiden ratkaisuvaihtoehtoihin sekä esiin nouseviin ongelmiin

Tutkimuslaitos esittämät pilotit

- Luke
 - Tutkimuslaitosten tuottamien aineistojen (pitkäaikaissurannat ja rekisteriaineistot) hallinta ja säilytys
 - Valtorin ja CSC:n tuottamien palveluiden yhteensovittaminen
 - Genetiikan laskennan pilotointi esim. seuraavissa projekteissa: Rumen metagonome study, comparative aligner study, arcArk transcriptome study
- SYKE
 - Lähtöaineistojen ja datavirtojen toimintamallin määrittely ja testaus
 - Federoidut tutkimustietovarannot ja laskentapalvelut: olemassa olevan datan laajempi hyödyntäminen
 - Tutkimuslaitosten käyttötarpeet CSC:n ympäristölle
- THL
 - Isaacus – terveysalan aineistojen yhteiskäytettävyyden kehittäminen (aineistokatalogi.fi)
- GTK
 - Hyperspektriskannauksen ja muun materiaaliskannauksen kehittäminen ja aineistoon liittyvän tutkimustyön mahdollistaminen: tavoitteena toteuttaa skannaus-, analysointi- ja tallennus- sekä jakeluratkaisun kokeilu korkeakoulujen kanssa
- IL
 - Suomalainen Satelliitti Data Cube
 - Suurten SAR-kuvamäärien matemaattinen analyysi (IL, MML, Aalto)
 - LES-mallin ja alueellisen ilmanlaatumallin yhdistäminen kaupunkimittakaavassa
- EVIRA
 - Kokogenomisekvensointidatan analyysi
 - Kansallisen datapankin (EFSA:n Data Warehousea vastaavan kansallisen systeemin) perustaminen ja kehittäminen
 - Laskurin kehittäminen saanninarviointien tekemiseksi kuluttajan nettilomakkeeseen täyttämien ruoankäyttötietojen pohjalta
 - Ihmisille tautia aiheuttavien bakteerien genomidatan hallinta ja jakaminen Eviran ja THL:n kesken sekä avaaminen tutkimuskäyttöön
- TTL
 - Ohjelmistoinfraskruktuurin rakentaminen henkilökohtaisten IoT-mittalaitteiden datan keräämisen keskittämiseen ja datan yhdistämiseen tietoturvalisessä ympäristössä (osana useampaa tutkimusprojektia)
- MML
 - Paikkatietoaineistojen yhtenäistäminen, GeoCubes (MML, Luke., SYKE, GTK...)

Esitys kansallisiksi DL-piloteiksi

- DL-intensiivisen tieteen yleisten edellytysten parantaminen
 1. Kansallisen datapankin (EFSA:n Data Warehousea vastaavan kansallisen systeemin) perustaminen ja kehittäminen
 - Tutkimuslaitosten tuottamien aineistojen (pitkäaikaissurannat ja rekisteriaineistot) hallinta ja säilytys
 - Federoidut tutkimustietovarannot ja laskentapalvelut: olemassa olevan datan laajempi hyödyntäminen
 2. Lähtöaineistojen ja datavirtojen toimintamallin määrittely ja testaus
 - Ohjelmistoinfraskruktuurin rakentaminen henkilökohtaisten IoT-mittalaitteiden datan keräämisen keskittämiseen ja datan yhdistämiseen tietoturvalisessa ympäristössä
- Sisältöperustaisen yhteistyön kehittäminen
 3. Paikkatietoaineistojen yhtenäistäminen
 4. Genetiikkaan liittyvä yhteistyö
 5. Oppimisanalytiikan kehittäminen (opintotietovaranto)

Sisältö

- Tietopohja ja yhteenveto
- Nykytila-analyysi
 - Nykytila-analyysin kuvaus
 - Toimijakohtainen nykytilakuvaus
 - Substanssikohtainen nykytilakuvaus
- Laskentakapasiteetin yksilöity tarve
 - Lisätietopyyntöä koskeva kyselylomake
 - Toimijakohtaiset lisätietokuvaukset
- Toimenpide-ehdotukset
 - DL:n keskeisimmät haasteet ja kehityskohteet
 - Tutkimuslaitosten tarpeita tukeva DL-investointi
 - DL-pilottiehdotukset
- **Huomioitavaa jatkossa**
- Liite 1. Tutkimuslaitosten nimien lyhenteet

Huomioitavaa jatkossa

- Datanhallinnan ja laskennan kokonaiskoordinaatiolle ja -kehittämiselle on tarvetta tutkimuslaitoksissa sekä laajemmin korkeakoulu- ja tutkimuskentässä
 - Yhteistyön edistämisen mahdollisuudet tulee jatkossa arvioida päätöksenteon eri tasoilla
- Yhden kokonaisratkaisun sijaan oleellista on kansallisesti keskittyä tuottamaan sellaisia palveluita, jotka edistävät suomalaisen tieteen kehittymistä kokonaisuutena
 - Perus- ja tulevaisuuden palveluiden koordinaatio näyttäisi olevan tarkoituksenmukaista, jolloin alakohtaisten ratkaisujen kehittämiseen pystytään kohdentamaan resursseja organisaatioittain
 - Yhteistyön kehittäminen on oleellisen tärkeää niin yksittäisten organisaatioiden kuin kokonaisuuden innovaatiokyvyyden varmistamiseksi – DL-kyvyyden tasossa on paikoin merkittäviä eroja
- Tässä selvityksessä esitetyt toimenpiteet tulee arvioida, priorisoida sekä aikatauluttaa onnistuneen toimenpiteistämisen tukemiseksi
 - Pilottien osalta suunnittelu- ja erityisesti rajaustyö tulee käynnistää pikimmiten tavoitteena mahdollisimman laaja-alaisen yhteistyön mahdollistaminen
- Suomalaisen tutkimuksen vaikuttavuuden kasvattamiseksi olisi hyvä selvittää yritys- ja elinkeinoelämän tarpeet DL-toimintaan liittyen PPP-toimintamallin (public-private-partnership) edistämiseksi

Liite 1. Tutkimuslaitosten nimien lyhenteet

- FIIA: Ulkopoliittinen instituutti
- GTK: Geologian tutkimuskeskus
- EVIRA: Elintarviketurvallisuusvirasto
- IL: Ilmatieteen laitos
- LUKE: Luonnonvarakeskus
- MML: Maanmittauslaitos
- STUK: Säteilyturvakeskus
- SYKE: Suomen ympäristökeskus
- THL: Terveysten ja hyvinvoinnin laitos
- TTL: Työterveyslaitos
- VATT: Valtion taloudellinen tutkimuskeskus
- VTT: Teknologian tutkimuskeskus VTT Oy

Turo Kilpeläinen
044 – 708 5085
turo.kilpelainen(at)gmail.com

www.minedu.fi/DL2021