



TIETOALLAS AVOIMEN LÄHDEKODIN RATKAISUNA

ATK-PÄIVÄT 23.5.2017

SAMI HELIN, KEHITTÄMISPÄÄLLIKÖ, HUS-TIETOHALINTO / JOHDON TIETOJÄRJESTELMÄT

SISÄLTÖ

- Tausta – Tietoaltaan rooli tiedonhallinnassa
- Tietoaltaan kehitys HUSissa – vaiheet ja kokemukset
- Avoimen lähdekoodin ratkaisu – käyttö ja kehitys
- Jatkosuunnitelmat ja eteneminen

TAUSTA

Tietoaltaan rooli tiedonhallinnassa

TIETOALLAS, MÄÄRITELMIÄ JA OMINAISUUKSIA

- Usein Hadoop- teknologiaan pohjautuva tietovaranto, johon kaikki tieto koostetaan tarkimmalla mahdollisella tasolla alkuperäisessä muodossaan tukemaan eri käyttötarpeita
 - Tukee kaiken muotoisen ja tyyppisen tiedon tallennusta
 - Tiedon muoto ja käsittely kiinnitetään vasta käyttötarpeen mukaan – ei tietoa tietovarantoon talletettaessa
 - Tiedon hallintaan laaja ja kustannustehokas avoimen lähdekoodin tarjoama
 - Skaalautuu erittäin suurten datamäärien tallennukseen ja prosessointiin
- Tietoallas ei ole määrämuotoisen raportoinnin ja tietovarastoinnin ratkaisu. Voi toimia alustana myös tuon tyyppiselle käytölle, mutta vaatii tuolloin käytön mukaisen oman kehitysprojektin.
 - Helpompiakin polkuja on, jos hakee vain ratkaisua varsin rajattuun raportoinnin tarpeeseen.

TOIMINTAYMPÄRISTÖN MUUTOKSIA JA PAINeita PERINTEISELLE TIEDONHALLINNALLE

Yleisiä ja HUS ajureita muutokseen	Perinteisen tiedon hallinnan haaste
Eri tyyppiset tiedon lähteet (ei-strukturoitu tieto, streamit, kuvat..)	Tiedon käsittely hidasta tai mahdotonta
Uudet laajat tietolähteet; IOT potilasmonitoreista ja omaseurantalaitteista, genomidata...	Tallentaminen kallista ja epäkäytännöllistä
Uudet tiedon käyttökohteet perinteisen raportoinnin lisäksi; esim. analytiikka ja koneoppiminen	Etukäteen ”yksien lukujen raportointiin” rajattu tieto saattaa hävittää muiden käyttökohteiden tarvitseman tiedon
Toimintaympäristön muutoksen nopeus kasvanut; tarvitaan eksaktin tiedon lisäksi myös nopeammin suuntaa antavaa ymmärrystä	Etukäteen mallinnetun tiedon tuottamisen prosessi on hidas

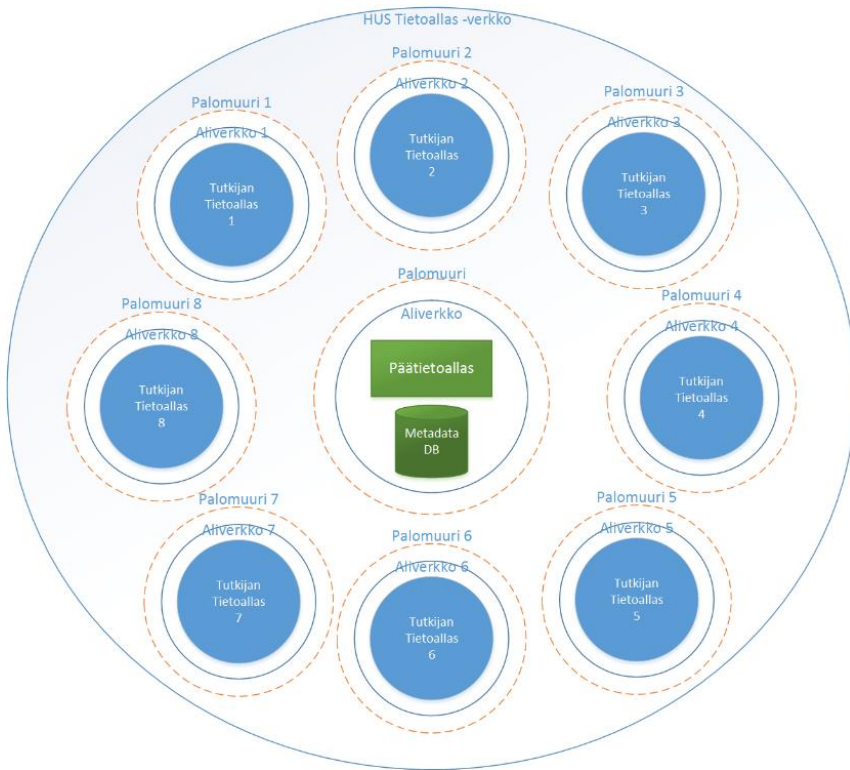
TIETOALTAAN KEHITYS HUSISSA

Vaiheet ja kokemuksia

PROJEKTIVAIHEITA HUS TIETOALLAS KEHITYKSESSÄ JA KÄYTÖSSÄ

2015				2016				2017				2018				
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
Pilotti, suunnittelu, toteutus ja koekäyttö																
				Isaacus: tietoaallas, pääsyn hallinta ja integraatiot												
								CE- merkintä ja Tietoaltaan kehityksen ja hyödyntämisen laajentaminen								
								Uuden tietovarasto- ja raportointiarkkitehtuurin kehittäminen								

HUS TIETOALLASJÄRJESTELMÄ- TOIMINTAPERIAATTEITA



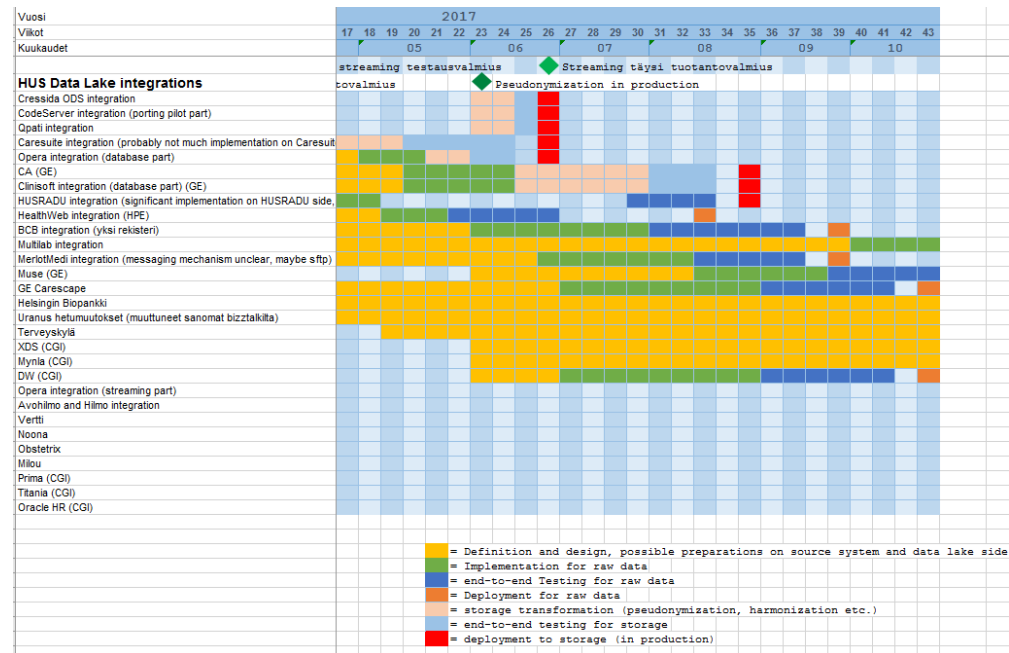
- HUS Tietoallas koostuu useista tietoaltaista jotka voivat jakaa tietoa
- Päätietoallas hoitaa tiedon integroinnin ja ajaa yleiset tiedonjalostusprosessit
- Tutkijan tietoaltaat / sivutietoaltaat toimivat tutkimustyössä ryhmän/toimijan omana altaana
- Sivutietoaltaat voivat käyttää päätietoaltaan dataa sitä monistamatta, mutta voivat sisältää myös omaan käyttöönsä spesifiä dataa
- Toteutusalueena Azure, mahdollistaa teknisesti modernin ratkaisun miniklustereilla
 - **Skaalautuvuus; tallennus ja laskenta erikseen**
 - **Tietoturvallinen Tietoaltaiden eristäminen**
 - **Ohjelmallinen automaatio ja pilvikapasiteetin tehokas käyttö.**
 - **Kytkeytyminen muuhun yritysinfraan**

ISAACUS-HANKKEEN AIKANA HUS TIETOALTAASEEN INTEGROIDAAN JO KATTAVASTI DATAA ERI LÄHDEJÄRJESTELMISTÄ

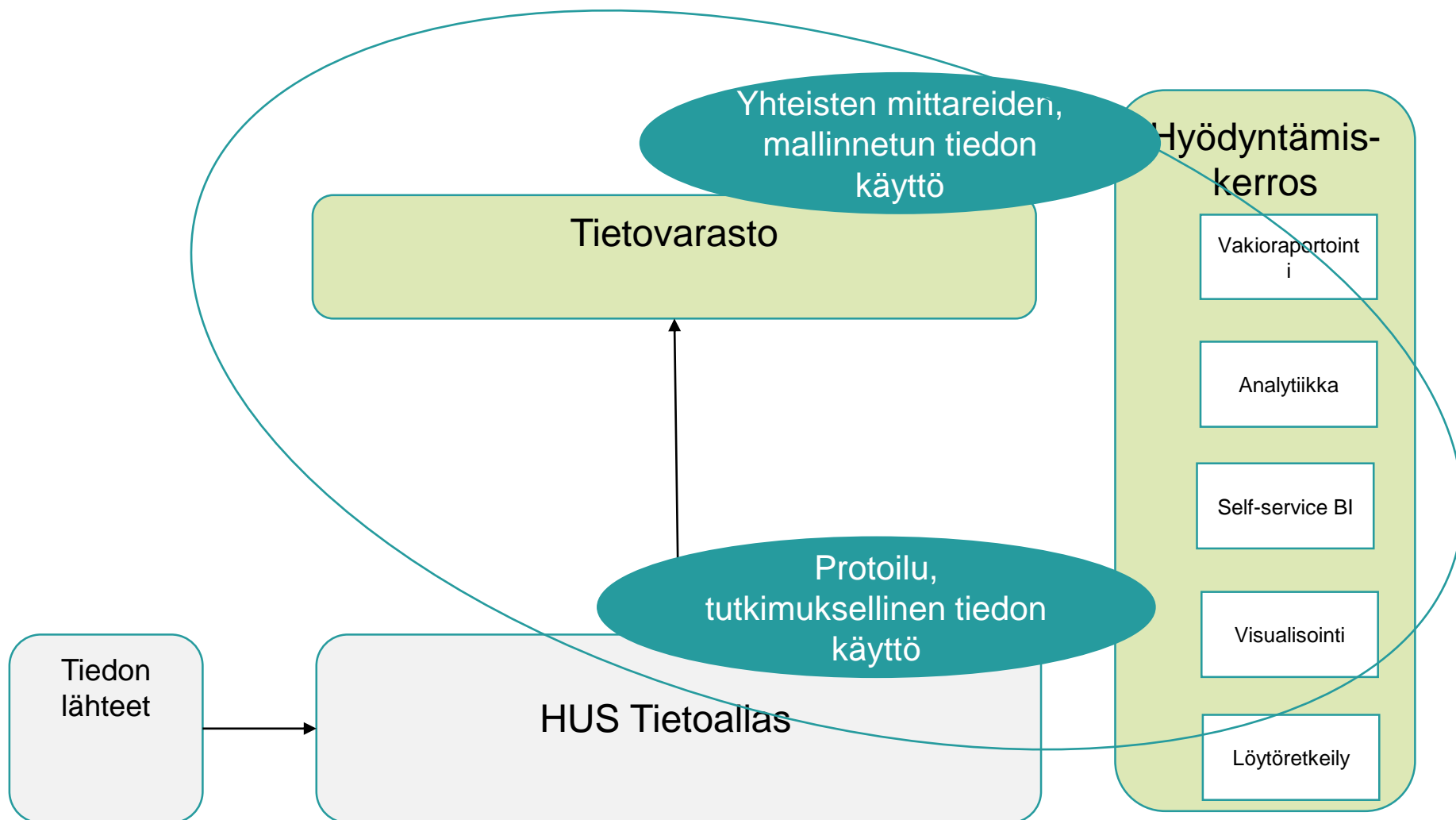
- Erikoissairaanhoidon potilaskertomuksia (Cressida/Uranus)
- Patologian ja genetiikan lausuntojen dataa (Qpati)
- Tehohoidon dataa (Caresuite/PICIS ja Clinisoft)
- Tautikohtaisia laaturekisterejä (BCB Medical)
- Ensihoidon kertomustietoja (MerlotMedi)
- Terveyshuollon Kanta-palvelun dataa (Healthweb)
- Laboratorion mittaustietoja (Multilab)
- Laskutustietoja (Mynla)
- Henkilöstöhallinnon dataa (Prima ja Titania)
- Anestesia (CA)
- Dokumenttien sähköinen arkisto (XDS)
- Raskauden ja synnytysten tiedot (Obstetrix)
- Leikkaussalien toiminnanohjaustietoja (Opera)
- ...

23.5.2017

Alustava integraatioaikataulu



TIETOVARASTO- JA RAPORTOINTIARKKITEHTUURIN KEHITYS - PROTOILUN JA VIRALLISEN OHJAUKSEN JA SEURANNAN TUKEMINEN



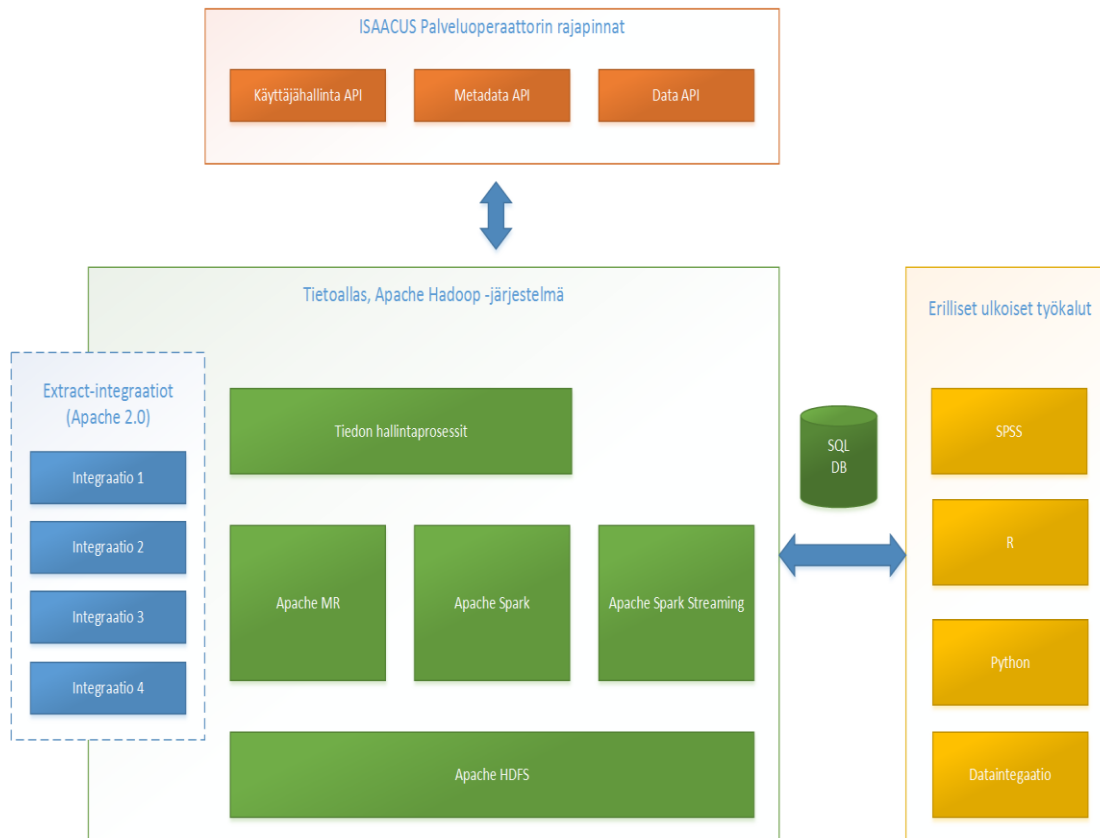
KOKEMUKSIA

- **Kalenteriaikaa kuluu paljon** – integraatioissa useita toimijoita ja omia aikatauluja
- Tietojen siirtäminen teknisesti suoraviivaista, mutta **tiedon laatuun ja ymmärrettävyyteen liittyvät perusasiat säilyvät**, jotta tietoa voidaan myöhemmin tehokkaasti käyttää (metadata, ymmärrys tiedosta, tietomuutosten ymmärrys). Huomioi riittävä tiedon laadun hallinta projektin alusta lähtien.
- Datan välittömästi saatavaksi tuomiseen liittyy paljon odotuksia. Toteutus ottaa kuitenkin aikaa. **Odotusten hallinta ja kommunikointi korostuu.**
- **Teknologia kehittyy koko ajan ja muutoksiin ja haasteisiin on syytä varautua** – jouduttu tekemään myös HUS projektin aikana. Valintoja on kuitenkin tehtävä.
- **Iso työ edessä kun data on saatu tietoaltaaseen** (tiedon jäsentäminen tiedon hyödyntämiseen)
- Perinteiset tietovarastoinnin liittymä- ja työskentelytavat puskevat helposti läpi – **ei ole vielä uusia toimintamalleja.**
- Järjestelmien **sopimusten ja toteutusvalmiuksien monimuotoisuus** datan ja metadatan toimittamiseen. Tarvittaisi yhteisiä vaatimuksia ja kansallisia sopimuksia.
- Ei ole ratkaisu kaikkeen, mutta **luo mahdollisuuden uudenlaisten toimintamallien ja ratkaisujen kehittämiseen** – edessä on kuitenkin oppimisprosessi

AVOIMEN LÄHDEKOODIN RATKAISU

Käyttö ja kehitys

HUS ARKKITEHTUURIN YLEISKUVA



- Ratkaisun perustana avoimet rajapinnat ja yhteensopivuus avoimen lähdekoodin Hadoop- ja Spark- ekosysteemeihin.
- HUS hyödyntää Tietoaltaassa Microsoft Azure:n palveluita hallinnoinnin ja operoinnin helpottamisessa
- Integraatiot ja käyttö avoimen lähdekoodin työvälineillä ja niiden käyttöä tukien

AVOIMEN LÄHDEKOODIN TIETOALLAS JA TUOTANNOLLISEN KÄYTÖN VAATIMAT PALVELUT

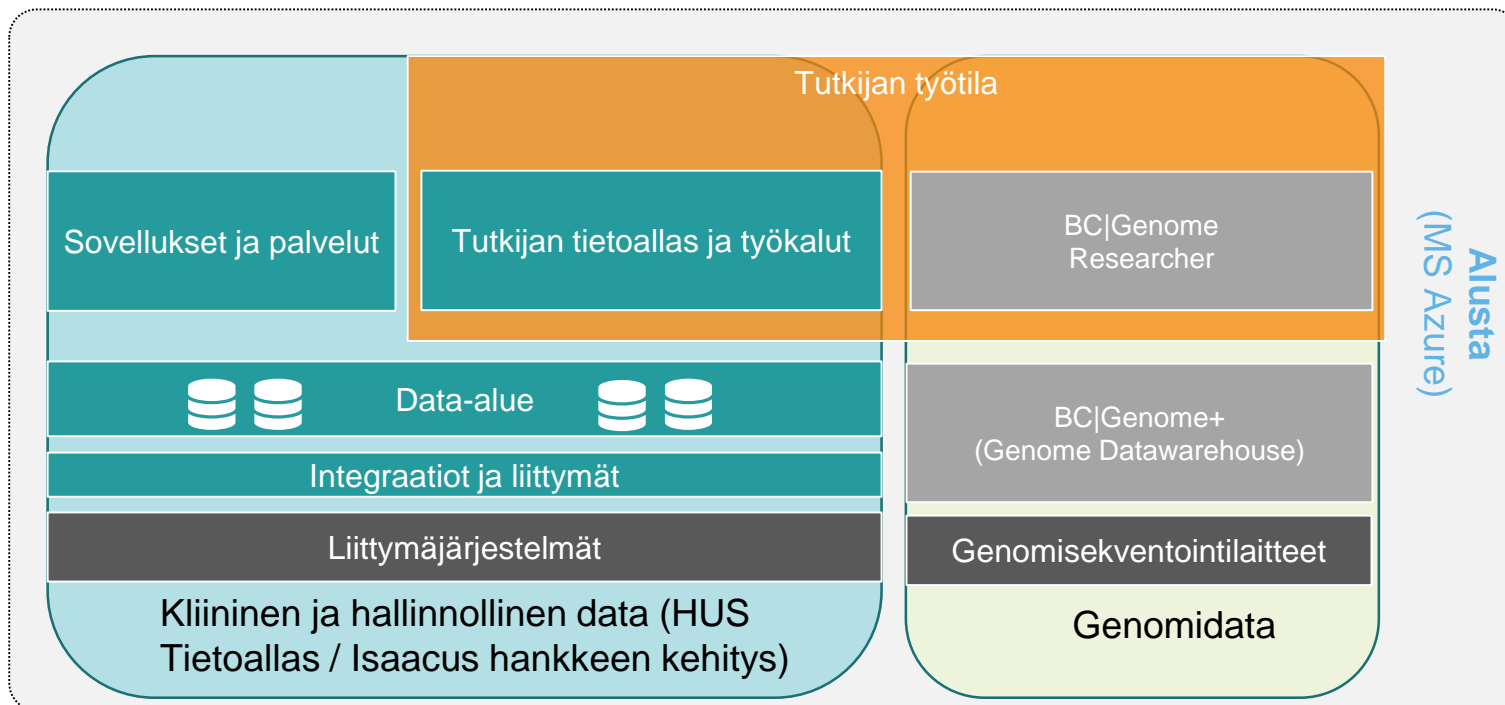
- Isaacus projektin tuotokset avoimena lähdekoodina käytettävissä Apache 2.0 lisensoituna ilman korvausta
 - Tietoallas tekninen toteutus
 - Automaatio ympäristön pystytykseen
 - Tietointegraatiot ja tiedonjalostusprosessit
 - Tarvittavat ohjelmistokonfiguraatiot
- Tuotannolliseen käyttöön huomioitava kuitenkin myös mm.
 - Hoidollinen hyödyntäminen ja tarvittaessa CE-merkitseminen vaatii tuotteistuksen ja valmistajan
 - Tietoaltaan integrointi organisaation omaan ICT-infrastruktuuriin
 - Tietoaltaan operoinnin ja jatkuvan palvelun toteutus sekä ylläpidon (ja mahdollisen oman jatkokehityksen) organisointi



JATKOSUUNNITELMIA

Mitä seuraavaksi

LOOGINEN ARKKITEHTUURI - TAVOITETILA



HUS Tietoallas

- ✓ SITRAN rahoittaman ISAACUS-hankkeen mukainen ja yhteensopiva, liittymä kansallisen palveluoperaattorin rajapintaan
- ✓ Mahdollistaa sekä ensisijaisen että toissijaisen tiedon hyödyntämisen; voidaan käyttää SOTE-alueen alueellisena tietoaletta ja hyödyntää alueen Tiedolla johtamisen alustana
- ✓ Jatkuva ylläpito ja kehittäminen → vikojen hallinta, päivitysten hallinta, jatkuva kehittäminen perustuen CE-merkinnän mukaisiin ISO laatuprosesseihin ja tuotteen elinkaarenhallintaprosesseihin
- ✓ Valmiiden integraatioiden joukko kasvaa, kun asiakkaita tulee lisää
- ✓ Datatutkijoiden kanssa kehitetty Tutkijantyötila
- ✓ Täyttää EU GDPR 2018 tietoturva-asetuksen mukaiset dataprosessorille asetetut vaatimukset
- ✓ Valmistettu MDD I prosessin mukaisesti → CE-merkityt tuotteet kaikilla tasoilla

Genomidata

- ✓ Genomidatan käsittely saatavana erikseen
- ✓ Mahdollistaa genomitiedon hyödyntämisen kliinisessä työssä sekä tutkimuksessa
- ✓ Helpottaa ja nopeuttaa genomitietojen ja kliinisen tiedon yhdistämistä sekä helpottaa tutkimukseen tarvittavien kohorttien keräämistä

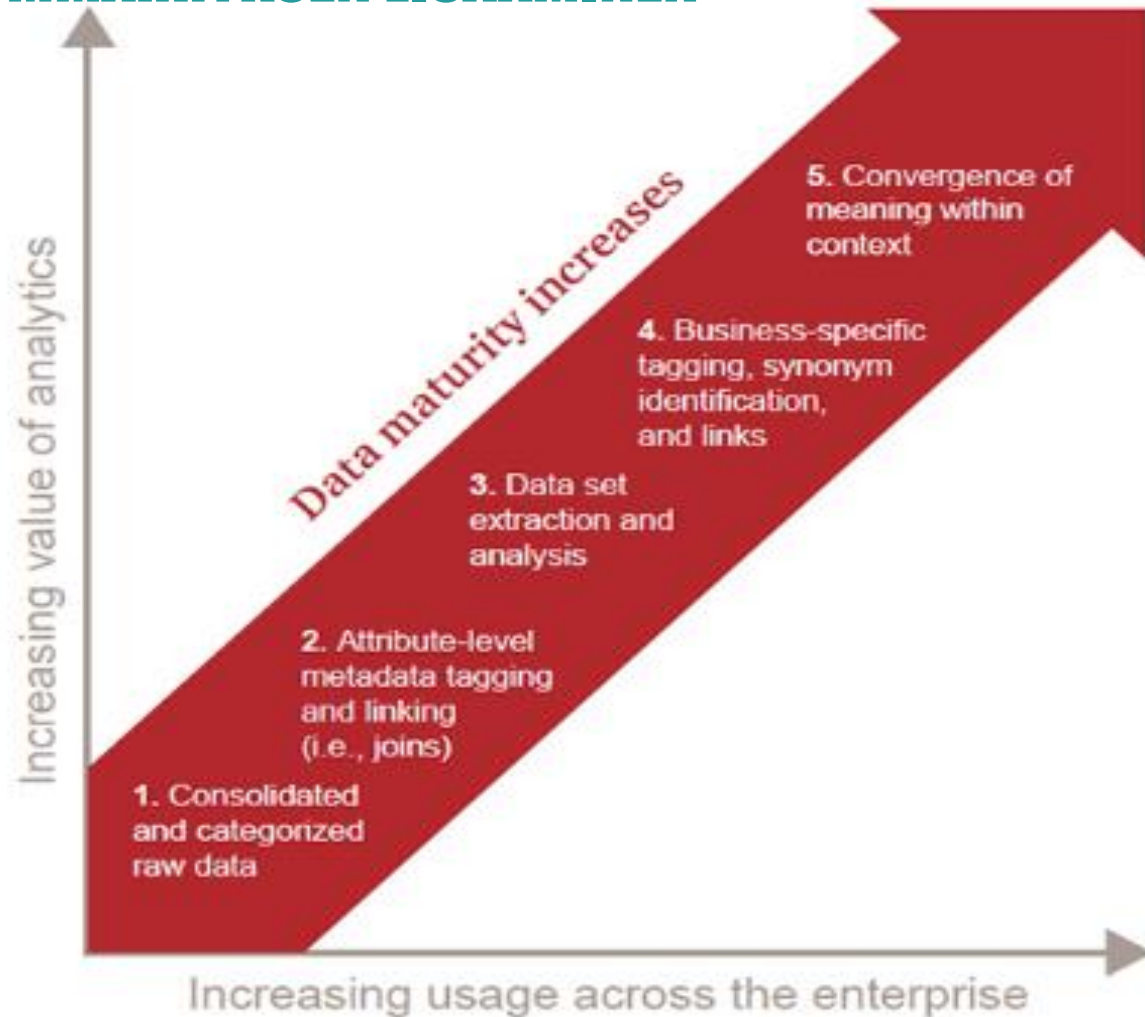
AVOIMEN LÄHDEKOODIN KOLLABORAATIO JA KEHITYS JATKOSSA

- Tavoitteena on, että avoimen lähdekoodin tietoallaskehitys jatkuu Isaacus- projektin jälkeenkin
 - Tarvitaan halua kehittää yhteistä mallia ja panostusta yhteiseen kehityshankkeeseen; nyt HUSilla ja Istekillä sama runkoratkaisu kehityksessä
 - Kehitys ei tyypillisesti avoimen lähdekoodin hankkeissakaan organisoidu itsestään vaan vaatii aktiivisia tahoja
 - Jatkokehitys tarvitsee tahon, joka koordinoi ja organisoii kehitystä ja teknisen portinvartijan kehityksen hallinnointiin
- > Tarvitaan kansallisen tason linjauksia ja toimintamalleja Sote:n ja palveluoperaattoritoiminnan käynnistämisessä

KÄYTÖN KEHITTÄMINEN ON OPPIMIS- JA YHTEISTYÖPROSESSI – HUS VAIHEITTAINEN YMMÄRRYKSEN LISÄÄMINEN

Luotava mekanismit kerätä, jäsentää ja rikastaa ymmärrys eri käyttötapauksiin.

Ymmärrys datasta kumuloimaan tehokasta käyttöä -> lisää ymmärrystä datasta.



Lähde: PWC 2015. Data lakes and the promise of unsiloed data

KIITOKSIA!

Sami Helin

050 403 3381

sami.helin@hus.fi