

Kohti kansallista sote-tietojen saatavuuspalvelua loppuraportti

Versio 1.0
31.10.2018

Minna Blomster
Arho Virkki

Tiivistelmä

Raportti esittää näkemyksen sosiaali- ja terveydenhuollon aineistojen saatavuusarkkitehtuuriksi sekä käy läpi joitakin valmisratkaisuja ja niistä saatuja kokemuksia.

Vuonna 2014 esitelty Sosiaali- ja terveysministerion ”SOTE-tieto hyötykäyttöön 2020” –strategia pyrkii edistämään Suomen kilpailukykyä kansainvälisen tutkimusrahoituksen markkinoilla. Myös Sitran Isaacus-hanke on osa tätä kokonaisuutta. Kansallisen tietopääoman hallitsijaksi on tarkoitus perustaa yksi keskitetty lupaviranomainen ja sen päätöksiä toteuttava palveluoperaattori. Näiden toiminnan pohjaksi tarvitaan sekä STM:n arkkitehtuurityö että alueellisiin tietoaltaisiin perustuvan tiedon toissijaisen hyödyntämisen arkkitehtuuri.

Tässä projektissa selvennetään saatavuuspalvelun arkkitehtuuriin ja toteutukseen liittyviä kysymyksiä alueellisten tietotaltain näkökulmasta. Saatavuuspalvelulla tarkoitetaan tässä kokonaisuutta, jonka avulla ulkopuolinen toimija voi ilman tutkimuslupaa tehdä tilastollisia kyselyitä pseudonymisoituun tietovarantoon selvittääkseen esimerkiksi tarpeellisen potilasmäärän olemassaolon kliinisen tutkimusasetelman tarpeisiin. Saatavuuskyselyn tekijä saa vastaukseksi tyypillisesti karkean lukumäärän potilaista, jotka toteuttavat annetut ikä, sukupuoli, lääkitys- ja muut ehdot.

Tätä kirjoittaessamme esiselvitykset tehdään pääsääntöisesti yksittäisinä työpyyntöinä asiantuntijoille, jotka ohjelmoivat käsin tietopyyntöön liittyvät kyselyt ja ajavat ne sairaaloiden tietotaltain tai raportointitietokantojen avulla. Lopputuloksena muodostuu arvio tutkimuksen kohortista, jolla tarkoitetaan lukumäärää potilaista ja heihin liittyvistä tapahtumista. Tutkimuksen esiselvityksellä arvioidaan, onko tutkimus toteutettavissa aineistossa. Lisäksi esiselvitys auttaa tutkimussuunnitelman muodostamisessa.

Tällä hetkellä ei ole määriteltynä yleisesti hyväksyttyä tarkkaa rajaa tutkimuksen ja esiselvityksen välillä, vaan poiminnan tekevä henkilö joutuu tapauskohtaisesti arvioimaan kysymyksen tutkimuksellisuuden. Saatavuuspalvelu arkkitehtuuriin tulee kuitenkin tukea tarkan ja jopa identifioitun datan saantia käyttöön, mikäli hyödyntäjällä on muussa palvelussa tai prosessissa todettu oikeus aineistoon. Saatavuuspalvelun arkkitehtuurissa on tarkoitus huomioida rekisterin sisäinen, alueellinen, kansallinen sekä kansainvälinen toissijainen käyttö.

Projekti sisältää saatavuuspalvelun skenaarion kuvaamisen, sen toteutuskelpoisuuden osittaisen testaamisen sekä tiekartan laatimisen ratkaisuarkkitehtuuriin toteuttamiseen kansallisena ratkaisuna. Skenaario tehdään valmisteluvaiheessa tai käytössä olevien kansallisten palvelujen pohjalta.

1 HANKKEEN AIKATAULU JA TUOTOKSET VAIHEITTAIN.....	4
2 KOKONAISKUSTANNUKSET JA SELVITYS RAHOJEN KÄYTÖSTÄ.....	7
3 TOTEUTUNEET TAVOITTEET JA HYÖDYT.....	7
3.1 TUOTOKSET.....	8
3.2 HYÖDYT	8
4 TOTEUTUNEET RISKIT	8
5 ONNISTUMISEN ARVIOINTI	9
5.1 TESTAUSYMPÄRISTÖN PYSTYTYS → IMPLEMENTOINTI ONNISTUI ERINOMAISESTI.....	9
5.2 INSITE TOTEUTUKSEN PYSTYTYS VSSH YMPÄRISTÖÖN JA TESTAUS	10
5.3 TOTEUTUKSEN SIIRTO TUOTANTOYMPÄRISTÖÖN.....	10
5.3.1 Testaus ja tuotantoon vienti → testauksessa syntyneitä huomioita tiivistettynä alla: (laajemmat arviot tuotoksissa 2.2.3 Laatuvalidointi tuotannossa ja 2.2.5 Käytettävyytestaus)	10
5.3.2 Testauksen tavoitteet	10
5.3.3 Tutkimusaineisto	11
5.3.4 Testauksen kooste:.....	11
5.2.5 Käytettävyys	14
6 ONNISTUMINEN YHTEISTYÖSSÄ MUIDEN ESITUOTANTOHANKKEIDEN KANSSA.....	15
7 SAATAVUUSPALVELUJEN NOPEASTI KASVAVA TARJONTA JA GLOBAALIKILPAILU	16
7.1 SAATAVUUSPALVELUIDEN LIIKETOIMINTAMALLIT	18
8 KANSALLISEN SAATAVUUSTIETOKANNAN TIEKARTTA	21
8.1 KOODISTON OSALTA	21
8.2 TIEKARTTA JA PERUSTELUT SAATAVUUSPALVELUN IMPLEMENTOINTIIN	23
8.3 TEKNISET EDELLYTYKSET PAIKALLISEN SAATAVUUSPALVELUN IMPLEMENTOINNILLE.....	24
9 KÄYTTÖOIKEUKSIEN MÄÄRITTÄMINEN	25
10 PALVELUOPERAATTORI ON VÄLTTÄMÄTÖN TOIMIJA	25
11 PALAUTETTA SITRAN TOIMINNASTA	29

1 Hankkeen aikataulu ja tuotokset vaiheittain

Vaihe ja vaiheen tulos	Tehtävät	Aikataulu hankkeen alusta tai ajankohta	Toteuttaja
1 Arkkitehtuuriskenaario			
1.1 Kokonaisuuden kuvaaminen <i>Arkkitehtuuriskenaario osana Raporttia 2</i>	Tehdään toiminta, tieto, järjestelmä ja teknologia-arkkitehtuurin kuvaus (ppt)	Q1/2018	VSSH, Medbit Oy
2. Testauskohteiden valinta ja toteutus			
2.1 InSite pilotti – 2/2017: InSite arkkitehtuurin toteutus VSSH tietojenkäsittely-ympäristöön ja teknisen toimivuuden ja ratkaisun kohortin muodostus -kyvykkyyden testaus kerta-aineistolla ja kerta-aineistoon soveltuvilla vanhoilla faasin 3 lääketutkimusprotokollilla			
2.1.1 Tietomäärittely <i>POCin sisällöllinen suunnitelma ja rajaukset</i> <i>Raportti 1</i>	Yksilöidään POCissa julkaistavat tiedot metatietoineen Kuvataan tietojen nykyinen jalostuspolku Tunnistetaan mahdolliset muutostarpeet	Q4/2017	VSSH, Custodix
2.1.2 Datamartin perustaminen <i>Yksityiskohtainen testattavan ratkaisun kuvaus</i> <i>Raportti 1</i>	Yksilöidään kunkin tiedon yhteis- mitallistamisen vaiheet Tunnistetaan ja otetaan käyttöön master-koodistojen lähdejärjestelmä Koodataan tarvittavat skriptit	Q4/2017	VSSH, Custodix, Medbit Oy
2.1.3 Tekninen ratkaisu <i>Teknisen alustan POC toteutus versio 1</i> <i>insite.vssh.net</i>	Perustetaan EHR4CR varten palvelin paikalliseen ympäristöön Avataan tietoliikenne-yhteydet	Q4/2017	Custodix, Medbit Oy
2.1.4 Testaus, paikallinen ja kansainvälinen, ja jatkosuunnitelma <i>kv-julkaisu</i>	Analysoidaan käyttöliittymän kyky löytää testiaineiston metatietoja ja näyttää niitä paikallisesti. Sovitaan POCin yhteistyötaho, joka toimii osapuolena testauksessa (lääketeollisuus)	Q2-Q3/2018	VSSH, Custodix Methods of Information in medicine - julkaisu toimitetaan arviointiin. BioMed Research International Volume 2015, Article ID 801436, 11 pages http://dx.doi.org/10.1155/2015/801436

	Testaan ratkaisu sovitulla aineistolla, osallistutaan julkaisuun. Arvioidaan ratkaisu ja suunnitellaan tuotantototeutus. Selvennetään IPR kysymykset		Journal of Biomedical Informatics Medical Informatics and Decision Making, Impact factor 2.134
2.2 Tuotantoon vienti VSSH tietoaalasympäristöön		Q2	
2.2.1 Tietolatausten automatisointi <i>Jatkuva palvelu</i> <i>Koodi VSSH:n Git-versionhallinnassa</i>	Suunnitellaan ja toteutetaan poiminnat tietoltaasta	Q3/2017	VSSH
2.2.2 Koodistojen siltaus kansainvälisiin koodeihin <i>Koodistosiltaus</i>	Suunnitellaan ja toteutetaan koodistojen siltaus	Q4/2017 - Q3/2018	VSSH (THL pidetään informoituna) KODISTOSILTAUKSIEN VALIDOITU VERSIO TOIMITETAAN KANSALLISELLE KODISTOPALVELIMELLE
2.2.3 Laatuvalidointi tuotannossa <i>Testatun tuotteen kehitysehdotus</i>	Selvitetään löytääkö InSite käyttöliittymä samat potilaat ja tapahtumat kuin ne, jotka mainitaan aikaisemmin käsin kerätyssä tutkimusaineistossa (TYKS aineisto FibStroke –tutkimuksessa)	Q2-Q3/2018	VSSH
2.2.4 Terminologiapalvelun testaus <i>Toimiva alueellinen terminologiapalvelu</i> <i>Ehdotus koodistosiltauksista THL:lle</i>	Testataan terminologiapalveluna THL käytössä olevaa Code Server –tuotteen VSSH installaatiota ja tunnistetaan THL koodistojen täydennystarpeet	Q2-Q3/2018	VSSH, Elisa Appelsiini Oyj, THL
2.2.5 Käytettävyydestä <i>Tieto alimman oikeustason saatavuustiedon käyttöliittymän käytettävyydestä ja sen kehittämistarpeista</i>	Testataan aggregoidun tiedon käyttöliittymän käytettävyys ja käyttökelpoisuus klinikon näkökulmasta	Q2-Q3/2018	VSSH

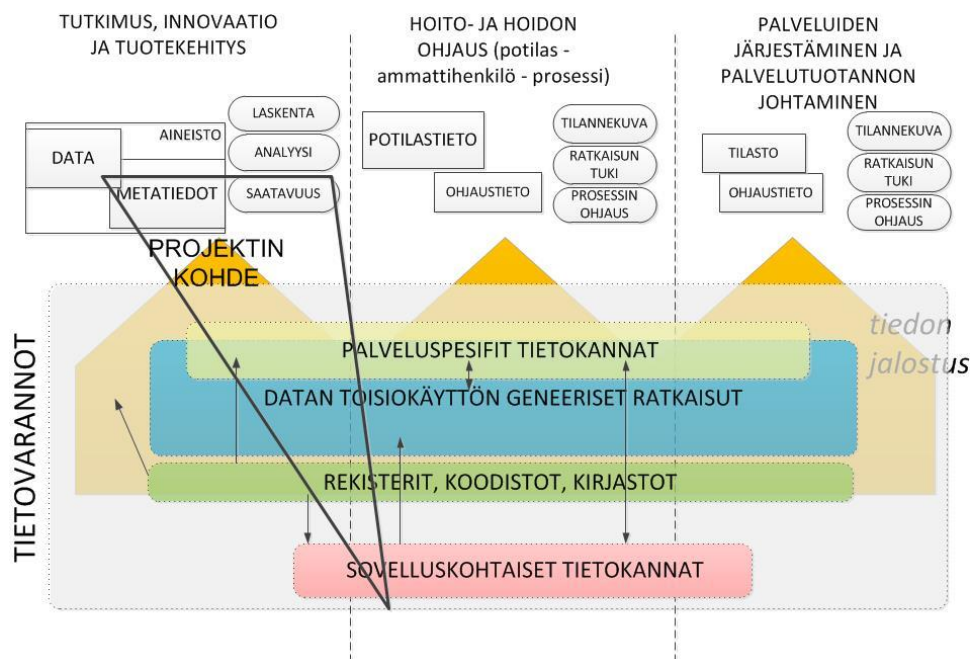
2.2.6 Arviointi ja paikallinen levityssuunnitelma	Arvioidaan tuotteen paikallinen käytettävyys, tehdään käyttöönottosuunnitelma ja käyttöönotto	Q2/2018	VSSH, 2M-IT Oy, Auria Biopankki, Custodix
2.2.7 Rinnakkaisratkaisu Väli­raportti 2	Kuvataan ja demonstroidaan BC Platforms ratkaisun arkkitehtuuri ja sen erot toiseen ratkaisuun	Q4/2017- Q1/2018	VSSH, BC Platforms, Auria Biopankki
3 Kansallisen toteutuksen tiekartta ja riskit			
3.1 Synteettinen data GitHub: https://github.com/Sitra-Isaacus/	Generoidaan synteettinen data, jonka avulla saatavuusratkaisuja voidaan todentaa ja testata ilman tietoturvarajoitteita	Q3/2017	VSSH
3.2 Välttämättömät sairaalakohtaiset tekniset edellytykset <i>Väli­raportti 2</i>	Kuvataan todetut tekniset edellytykset	Q1/2018	VSSH ja Medbit Oy
3.3 Kansallinen koodistopalvelimen ylläpitoprosessi ja vastuut	Kuvataan terminologiapalvelun hallintamalli	Q2Q3/ 2018	VSSH, Elisa Appelsiini Oyj
3.4 Palveluoperaattorin rooli potilaskohtaisen tiedon yhdistämisessä ja saatavuustiedon tietohakemiston laskennassa	Analysoidaan toteutuksessa esiin tulleet haasteet kansallisesti toteutettuna	Q3/2018	VSSH
3.5 Hajautetun tiedontallennuksen mahdollisuudet ja ongelmat	Analysoidaan esim. geenitiedon yhdistäminen kansalliseen arkistoon vs. pitäminen erillään	Q3/2018	VSSH, Sitra

2 Kokonaiskustannukset ja selvitys rahojen käytöstä

		Kausi 1	Kausi 2	Kausi 3	Kausi 4	Yhteensä	Jäljellä
		15.9.-31.12.2017	1.1.-31.3.2018	1.4.-30.6.2018	1.7.2018-30.9.2018		
234 060,00		42 941,74	100 648,11	111 120,13	61 994,07	316 704,05	-82 644,05
118 920,00		28 549,26	4 444,17			32 993,43	85 926,57
352 980,00		71 491,00	105 092,28	111 120,13	61 994,07	349 697,48	3 282,52
175 000,00	49,58 %	35 443,72	52 104,75	55 093,36	30 736,66	173 378,49	1 621,51
177 980,00	50,42 %	36 047,28	52 987,53	56 026,77	31 257,41	176 318,99	1 661,01
352 980,00	100,00 %	71 491,00	105 092,28	111 120,13	61 994,07	349 697,48	3 282,52

Projekti päättyi syyskuun lopussa 30.9. Projektin kokonaiskustannukset ovat 349 697€ Sitra rahoitti hanketta 173 380 €.

3 Toteutuneet tavoitteet ja hyödyt



Kuva1: Tiedon toisiökäytön käsitteelliset linjat ja projektin raja-
aus.

Projektin tavoite oli kuvata kansallisen saatavuuspalveluratkaisun tiekartta sen kokemuksen pohjalta, joka saadaan tehdessä koearkkitehtuuri kahden saatavuuspalvelua tuottavan yrityksen kanssa sekä ratkaisun testaamisesta VSSH:n tietoaalusratkaisun ja Auria Biopankin yhteydessä.

3.1 Tuotokset

Projekti tulos, alueellisen Saatavuuspalveluratkaisun testaus ja analyysi vastaavan toteutuksen edellytyksistä, rajoituksista, haasteista ja hyödyistä kansallisesta näkökulmasta esitellään tässä dokumentissa ja sen liitteissä.

Sivutuotos oli testaustarkoituksiin synteettistä dataa, jota voidaan käyttää myös opetus ja tuotekehitykseen sekä Hackathontyyppisten tapahtumien materiaalina. Sosiaali- ja terveydenhuollon uudistus (SoTe) on Suomen historian suurin yksittäinen hallinnollinen muutos, mutta dataa on ollut käytettävissä suhteellisen niukasti kahdesta syystä: (1) aineistoa on ollut hankalaa yhdistää teknisesti ja (2) tietosuojalait rajoittavat aineiston käyttöä. Nämä kaksi ongelmaan ovat ratkaistavissa informaatioteknologian, muuttuvan lainsäädännön ja datan tunnistamattomaksi tekemisen avulla.

Synteettisen datan ohjelmakirjasto on valmis ja saatavilla MIT-lisenssillä osoitteesta:

<https://github.com/Sitra-Isaacus/VSSHP-synthetic-data>

Synteettinen Data on satunnaislukugeneraattorin avulla keinotekoisesti tuotettua dataa, eikä se sisällä todellisia tapahtumia. Datan tilastolliset ominaisuudet mukailevat kuitenkin alkuperäistä havaintoaineistoa. Etuna on, että dataa voidaan käyttää vaihtoehtona oikealle datalle ilman, että henkilön tietosuoja vaarantuu mm. opiskelussa, tuotekehityksessä, koneoppimismallien kehityksessä jne. Matemaattisten mallien ja menetelmien validoitumisen myötä synteettisen datan käyttö yleistyy. Synteettinen data muodostetaan empiirisistä kumulatiivisista jakaumista, jotka on johdettu käyttäjän toimittamasta todellisesta datasta. Mikäli tämäkään datan käyttöä ei katsota sallittavan, on parametristen matemaattisten mallien avulla tuotettu data oikea, mutta paljon työläämpi vaihtoehto ja silloin menetetään myös tilastolliset piirteet, jotka synteettisessä datassa jossain määrin ovat näkyvissä.

3.2 Hyödyt

Projekti on tärkeä käytännön askel kohti kansallista saatavuusratkaisua. Projekti testaa arkkitehtuuria toimivassa tietoallasympäristössä. Kansallisen saatavuusratkaisu on puolestaan olennainen osa, jotta palveluoperaattorilla on työkaluja eri aineistojen löytämiselle sekä niiden metatietojen esittelemiselle ja näyttämiselle asiakkaille. Lisäksi lupaviranomainen tarvitsee testitoteutuksen omia palveluita pystyttäessään.

4 Toteutuneet riskit

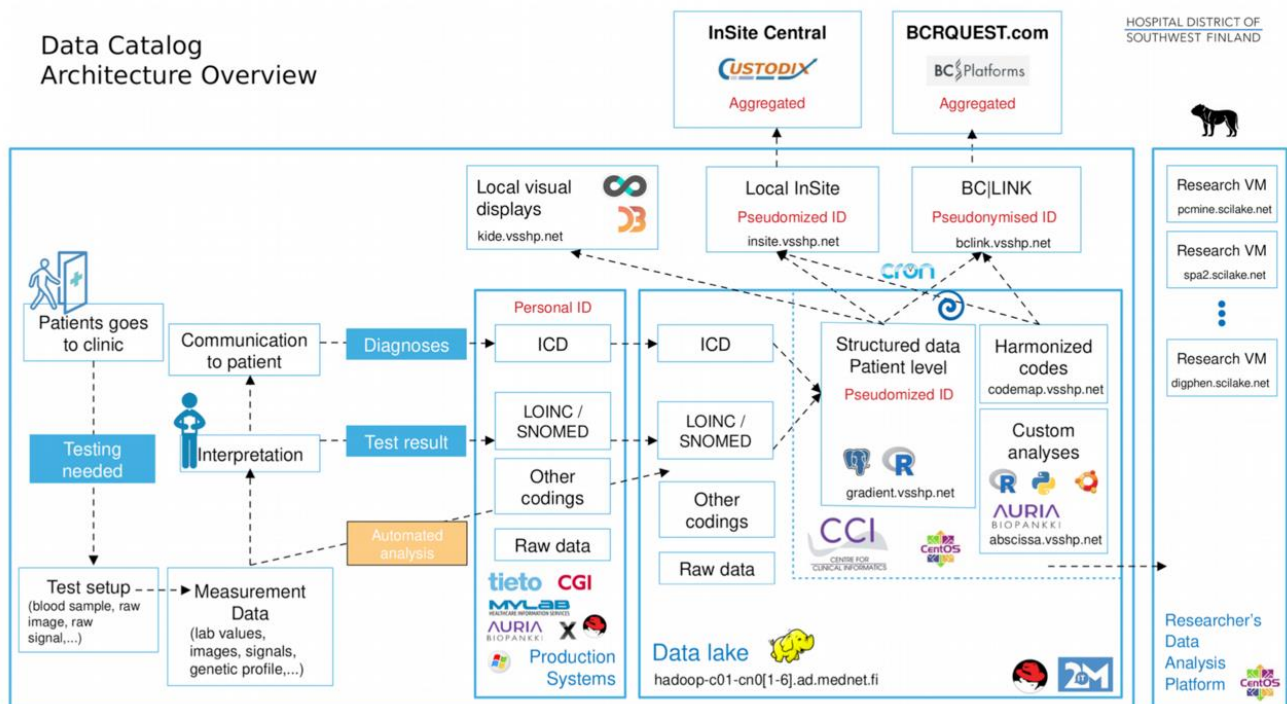
Projektisuunnitelman mukaan oli tarkoitus testata 2 saatavuuspalveluratkaisua, Custodixin Insite-ratkaisua ja BC Platformin R|QUEST saatavuuspalvelu alustaa. BC Platformsin R|QUEST -testaus päästiin aloittamaan 6 kuukautta myöhässä ja 08/2018. Tästä syystä BC Platformin R|QUEST -palvelun testaaminen jäi tämän projektin osana pintapuoleisemmaksi kuin Custodixin Insite-ratkaisun.

5 Onnistumisen arviointi

Testauksen osia olivat alkuperäisen suunnitelman mukaan:

5.1 Testausympäristön pystytys

→ implementointi onnistui erinomaisesti



Kuva 2. Tieto-, järjestelmä- ja teknologia-arkkitehtuurin kuvaus saatavuuspalvelun käyttäjän näkökulmasta (VSSH).

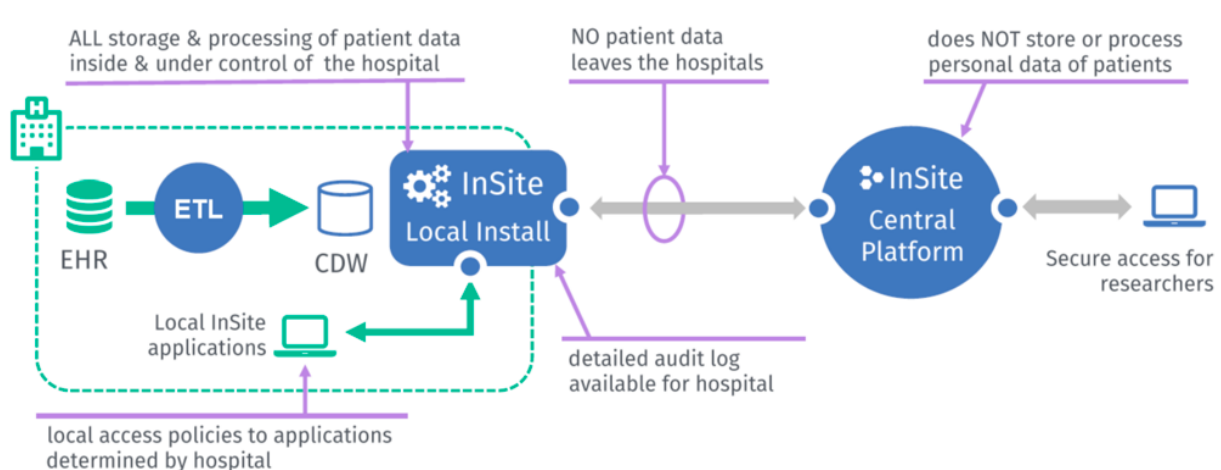
Tieto kertyy operatiivisiin järjestelmiin, josta se ladataan ajastetusti M2-IT:n ylläpitämään Hadoop-pohjaiseen raakatietoaalaseen (Cloudera Hadoop ja Red Hat Enterprise Linux).

VSSH:n Aurian tietopalvelu (Auria Clinical Informatics; ACI) yhteismitallisti datan rakenteiksi, joista ladattiin kaikkien saatavuuspalveluiden ja käyttöliittymien raakatieto (kuvassa paikallinen Custodix InSite, ja BC|Link). Koodistojen keskitetystä harmonisoinnista hyötyvät ekosysteemissä laadullisesti ja työmäärällisesti kaikki: Tätä varten on erillinen palvelin (codemap.vssh.net), jonne jokaisella koodistojen käsittelijällä on luku- ja kirjoitusoikeus. Myös paikalliset käyttöliittymät (Apache Superset) ja analyysiympäristö (Aurian tietopalvelu, talouspalvelut ja Auria Biopankki; Ubuntu 16.04, R ja Python) käyttävät samaa yhteismitallistettua dataa.

VSSH:n Aurian Tietopalvelussa tutkimusluvalla suoritettavaan tutkimukseen on mahdollista perustaa tutkijan allas (Researcher's Data Analysis Platform; DAP). Data luovutetaan myös tähän ympäristöön samasta yhteismitallisesta joukosta, jota käytetään saatavuuspalveluiden pohjana. Tutkimusympäristöön voi kirjautua vahvasti tunnistautumalla RSA-avaimella ja salasanan yhdistelmällä ja se avataan tietosuojasitoutumusta ja hyväksyttyä tutkimuslupaa vastaan.

5.2 InSite toteutuksen pystytys VSSHP ympäristöön ja testaus

Insite järjestelmän implementointi tietoaaltaaseen meni 5.1 osiossa kuvatun arkkitehtuurin mukaisesti.



Kuva 3. Insite arkkitehtuurissa oletusarvoisesti edes metadata ei näy sairaalan ulkopuolelle. Vain tieto siitä että VSSHP on verkostossa mukana näkyy muille käyttäjille.

5.3 Toteutuksen siirto tuotantoympäristöön

Tietolatausten automatisointi ja yhdistäminen kansainvälisiin koodistoihin

→ myös tämä vaihe onnistui erinomaisesti Custodix Insite-alustalla.

→ BC Platformin R|QUEST alustan implementointi myöhästyi 6 kuukautta ja käyttöoikeuksia palveluun saimme rajoitetusti vasta projektin loppuvaiheessa, tuosta syystä BC|RQUEST testaus jäi pintapuoleiseksi. BC Platformsin saatavuuspalvelualusta oli vielä kehityksen alla ja ei testausaikalla tarjonnut toiminnallisuutta, kuten aikamäärään huomioimista osana hakukriteerien muodostusta.

Autovalidate-tutkimus perustuu Fibstroke-tutkimuksessa kerättyyn kardiologiseen potilasaineistoon vuosilta 2003-2012, sekä lisäanalyysiin, jossa TYKSin tietokannoista tehdyistä poiminnoina selvitetään tietyin kriteerein potilastapahtumat vuosilta 2013-2017.

5.3.1 Testaus ja tuotantoon vienti

→ testauksessa syntyneitä huomioita tiivistettynä alla:

(laajemmat arviot tuotoksissa 2.2.3 Laatuvalidointi tuotannossa ja 2.2.5 Käytettävyytestaus)

5.3.2 Testauksen tavoitteet

Tutkimuksen tavoitteena on selvittää seuraavat kysymykset:

1. Miten hyvin InSite-ohjelmistolla löydetään TYKS:in tietokannasta potilaat, jotka oli manuaalisesti haettu Fibstroke-tutkimukseen vuosilta 2003-2012
2. Miten hyvin InSite-ohjelmistolla löydetty FibStroke-tutkimuksen potilaiden tiedot vastaavat potilaiden terveyttä, sairautta ja interventioita koskevia tietoja FibStroke aineistossa
3. Ovatko aineistojen välillä mahdollisesti todettavat eroavaisuudet lähdeaineiston puutteita vai InSite-ohjelman toiminnallisuuteen liittyviä ilmiöitä
4. Mitä haasteita InSite ohjelmiston käyttämiseen klinikan kannalta liittyy käytettävyyden tai palvelun käyttäytymisen kannalta

5.3.3 Tutkimusaineisto

FibStroke tutkimuksessa on tunnistettu retrospektiivisesti sairaskertomuksista potilaat, joilla on sekä eteisvärinädiagnoosi että aivoverenkiertohäiriödiagnoosi (stroke/TIA) milloin tahansa 10-vuotistutkimusajanjakson (vuodet 2003-2012) aikana. Tutkimukseen otettiin mukaan vain potilaat, jotka ovat asuneet sairaalan alueella aivohalvauksen ilmetessä ja vähintään vuoden ajan sen jälkeen, jotta kattavat tiedot komplikaatioista voidaan kerätä. Myös lääkitystiedot ja riskitekijöitä CHA2DS2-VASc ja HAS-BLED arvioitiin.

5.3.4 Testauksen kooste:

Insite Platform on kahdesta testaamastamme saatavuuspalvelusta kehittyneempi. Käyttöliittymän käyttäjälähtöisyydessä jää vielä Custodix Insitenkin rajapinnassa toivomisen varaa. Käyttöä haittaa mm. tulkinnan vaikeudet filterikriteerien asetuksissa, jotka syntyvät jokseenkin monimutkaisesta logiikasta käyttöliittymässä silloin kun kriteereitä on useita.

Molemmat saatavuuspalvelut löytävät enemmän potilaita kuin laadunvalitointiin valitussa Fibstroke väitöskirjatutkimuksessa on löydetty manuaalisin poiminnoin annetuin kriteerein.

KOHORTIN MUODOSTUS KRITEERISTÖ AUTOVALIDATE TUTKIMUKSEEN PERUSTUEN

1. Aikuisia (>18 vuotiaita) 1.1.2003 diagnoosihetkeltä alkaen
2. Potilaalla Diagnosoitu eteisvärinä à *Atrial Fibrillation (AF) or atrial flutter I48*
3. Yksi tai useampi seuraavista aivoverenkierron häiriöistä à *Disturbances in cerebral blood flow during 01.01.2003-31.12.2012:*
 - a. Aivoinfarkti à *Ischemic stroke*, I60.0-I60.9, I61.0 - I61.9, I62.0 - I62.9, I63.0 - I63.9, I65.0-I65.9, I66.0-I66.9, I69.0 - I69.9 **or**
 - b. Ohimenevä aivoverenkiertohäiriö à TIA, G45.0 - G45.9, G46.0 - G46.9 **or**
 - c. kallonsis. Verenvuoto à *Intracranial bleeding*, S06.0- S06.9
- Kuinka monella potilaalla oli eteisvärinä **AF diagnoosi ENNEN ensimmäistä diagnoosia a, b tai c.** (ENNEN määritelty, ainakin 1 päivä ennen).

saatavuuspalvelu	Kohortin koko	Huomio
Insite Platform, Custodix	6230 Potilasta	Insight Platformilla kohortin muodostus pystytään tekemään. Haku verifioitiin myös Custodixin asiantuntijoiden toimesta sillä käyttöliittymä ei ole näin monimutkaisen (Fibstroke aineiston Autovalidate study)

BC-R QUEST, BC Platforms	10934 potilasta	kohortin luomiselle ihan yksiselitteinen. Kohortin rakentaminen ei onnistu autovalidate tutkimuksen edellyttämällä tavalla. -Pvm ehtoja ei kyetty asettamaan. -Ehtoja mikä tapahtuma tapahtuu ennen toista tapahtumaa ei kyetä asettamaan.
VERTAILUKOHDE: AUTOVALIDATE STUDY (SAS:illa poimittu FIBSTROKE aineisto)	2920 Potilasta	tutkimusaineistoon/väitöskirjaan ei ollut kirjattu kriteereitä tai tietoja potilaista, joita ei oltu kelpuutettu mukaan manuaalisen haun perusteella.

Fibstroke-tutkimuksen ottaminen vertailututkimukseksi ei ehkä ollut paras vaihtoehto validoinnin kannalta, sillä aineistoon/väitöskirjaan ei ollut kirjattu kriteereitä tai tietoja potilaista, joita ei oltu kelpuutettu mukaan manuaalisen haun perusteella. Myöskään tietopoiminnan toteutuksen skriptejä ei ole tallella organisaatiomuutos syistä. Nykyään tietoluovutuksen prosessiin kuuluu tallentaa sekä SQL-, että R-tilastoajon skriptit, että tutkimusluvut samaan paikkaan jolloin skriptit ovat helposti uudelleen ajettavissa lähdejärjestelmistä. Tämän takia suora vertailu on mahdotonta. Myöskään ei ollut mahdollista määritellä samaa aikakriteeriä kuin Fibstrokeissa sisällyttää sellaiset potilaat, jotka olivat VSSH:n potilaina vähintään vuoden aivotapahtuman jälkeen.

Insitellä saatu kohdepopulaation määrä oli kaksinkertainen Fibstroke-tutkimuksen hakuihin. Ei ole olemassa yhtä selittävää tekijää suuresta erosta.

InSite central platform screenshot

PFS Authoring criteria (queries)

Clinical concept quick search

Clinical concepts to design queries with (ICD, LOINC, SNOMED, etc.)

Designing feasibility queries (i.e. eligibility criteria)

Kuva 4. Insite kuvankaappaus tutkimusta suorittavan organisaation näkökulmasta.

Saatavuuspalvelun validointiin valitut Fibstroke Autovalidate-studyn kohortin muodostus kriteeristön määritelmä alla:

-----TESTICASE KOHORTTI-----

Aikuisia (>18 vuotiaita) 1.1.2003 diagnoosihetkeltä alkaen

1. Potilaalla Diagnosoitu eteisvärinä → Atrial Fibrillation (AF) or atrial flutter I48
2. Yksi tai useampi seuraavista aivoverenkierron häiriöistä → Disturbances in cerebral blood flow during 01.01.2003-31.12.2012:
 - a. Aivoinfarkti → Ischemic stroke, I60.0-I60.9, I61.0 - I61.9, I62.0 - I62.9, I63.0 - I63.9, I65.0-I65.9, I66.0-I66.9, I69.0 - I69.9 or
 - b. Ohimenevä aivoverenkiertohäiriö → TIA, G45.0 - G45.9, G46.0 - G46.9 or
 - c. kallonsis. Verenvuoto → Intracranial bleeding, S06.0- S06.9

- Kuinka monella potilaalla oli eteisvärinä AF diagnoosi ENNEN ensimmäistä diagnoosia a, b tai c. ("ENNEN" määritelty, ainakin 1 päivä ennen diagnoosia a, b tai c).

Tulokset:

	Insite search	Fibstroke manual search
Number of patients who suffered from stroke, TIA or Intracranial bleeding during 2003-2013 and who had ever been diagnosed of AF (see Fig 1)	6230	2920
Number of events (stroke, TIA or/and intracranial bleeding) in total during 2003-2012 (as some had more than one event).	9114	3762
AF diagnosed before the event. Please see explanation in the end of this document why the last values (4493 and 2177) are not comparable with each other.	4493	2177

→ Ei yhtä kohortin suuruuseroja selittävää tekijää.

→ Tällaisia syitä voisi olla mm. se, että väitöskirjassa on alkuperäisen kohortin siivoamiseen/data curatointiin käytetty tapoja joita ei ole selitetty niin että tutkimus voitaisiin toistaa mielekkäästi. → Poiminta tuolloisesta SAS-järjestelmästä on voitu toteuttaa hieman eri tavoin → ei selitetty osana väitöskirjaa.

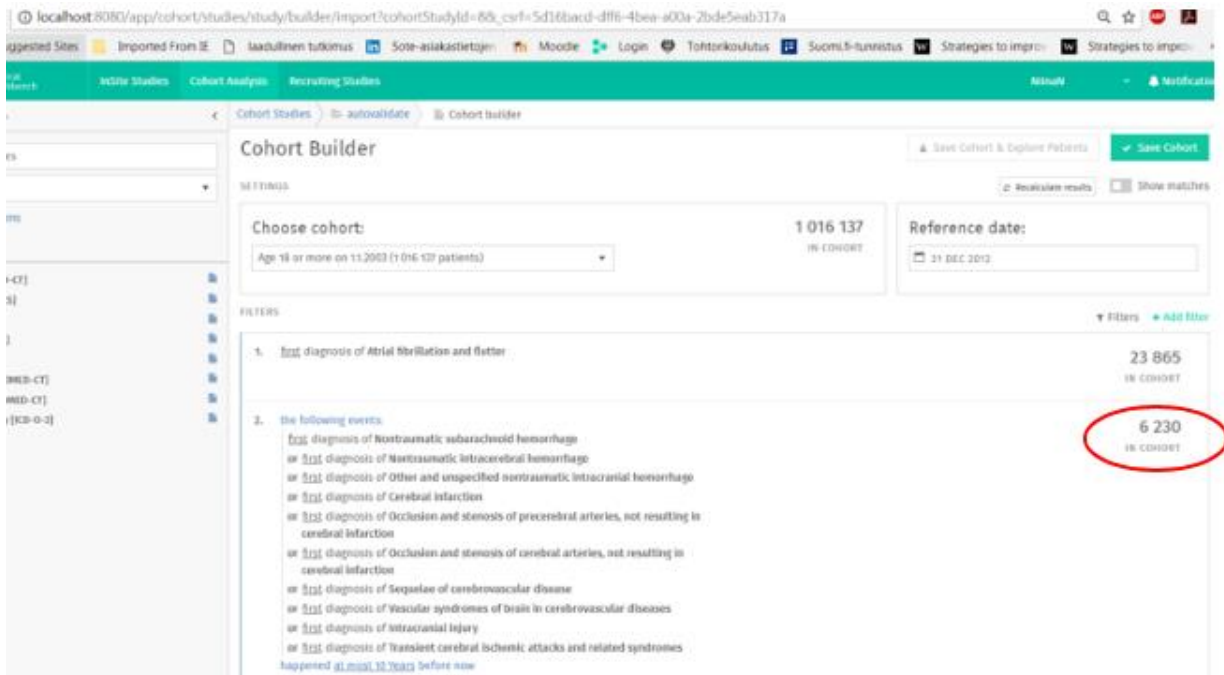
→ Toki voi olla, että Insite on hauissaan tarkempi.

→ Mainittakoon vielä, että tiedustelimme useampaan kertaan Insitelta asiaa ja tunnistimme, että insite ottaa hakuihinsa mukaan myös potilaita, joilla exclude kriteeri päällä, eli potilaita jotka eivät pitäisi tulla

Fibstroke verrokkitutkimusasetelmassa oli kaksi keskeistä haastetta suhteessa testausasetelmaan:

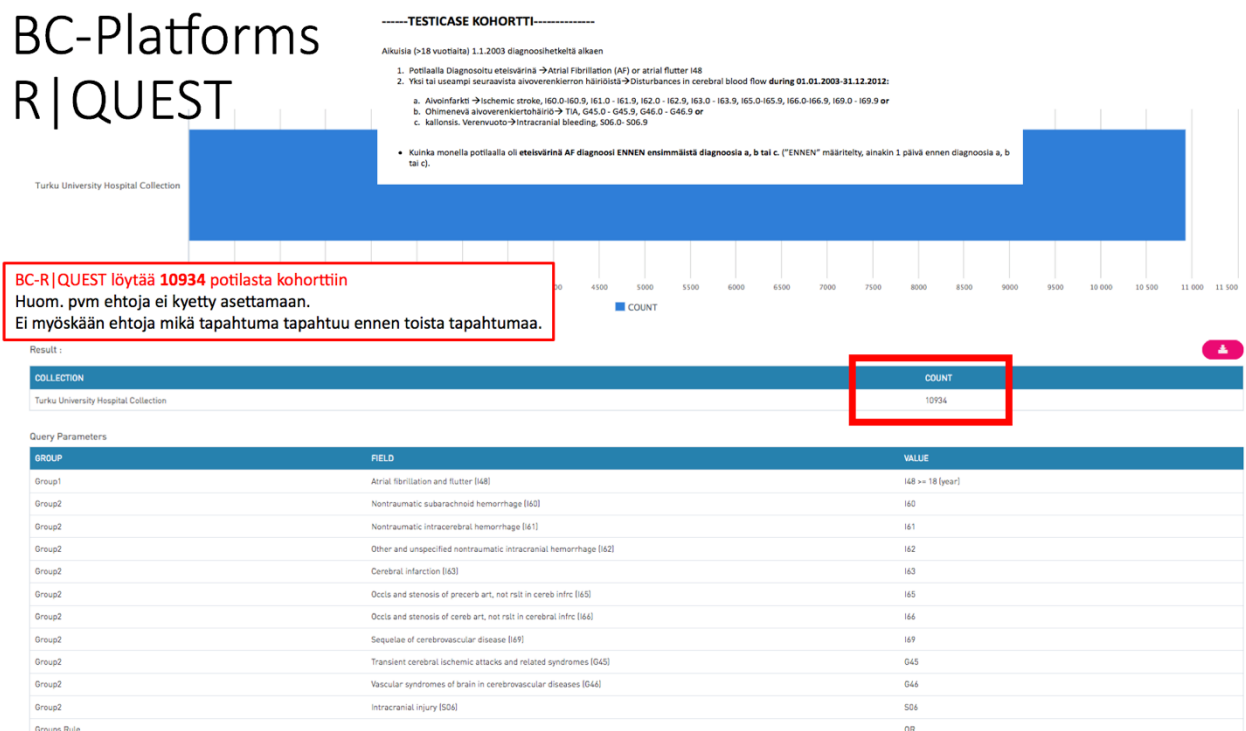
1. Filtterikriteereitä oli useita ja niiden toteuttamisessa Insite Platform käyttöliittymän avulla tarvittiin useita validointeja kehittäjien kanssa, sekä verifiointeja R- ja SQL koodein. Osoittautui haasteelliseksi tehdä hakuja juuri Fibstroke kriteerejä simuloiden.

2. Haasteita aiheutti myös väitöskirjan puutteellinen ilmaisu siitä kuinka alkuperäinen materiaali oli poimittu, kuratoitu ja mistä lähteistä poiminta oli tarkalleen ottaen alun perin tehty ja millaisin työvaihein. Kuratoinnilla tarkoitetaan on tiedon keräämistä useista eri lähteistä uudeksi sisältöyksiköksi ja sen muokkaamista yhtenäisemmäksi kokonaisuudeksi tutkimuksen tarpeita paremmin palvelevaan muotoon. Käytännössä tämä tarkoittaa olemassa olevan tiedon muokkaamista tarpeisiin sopivaksi.



KUVA: kuvankaappaus Custodixin Insite saatavuuspalvelualustalta kohortin rakennuksesta.

BC-Platforms R|QUEST



kuvankaappaus BC-Platforms sin saatavuuspalvelualustalta kohortin muodostuksen kriteereistä ja tuloksista.

5.2.5 Käytettävyys

Saatavuutta on ollut mahdollista tutkia VSSH:ssä sekä sisäisen InSite-palvelun, jonne tiedot päivittyivät kerran päivässä ja ja BC|RQUEST palveluun data vietiin bulkkilatauksena. Insite käyttöliittymään pääsee sairaalan sisäverkosta ja siihen kirjaututaan erikseen. Johtajaylilääkäriin luvalla on valtuudet lisätä





palveluun käyttäjiä ja voimme halutessamme valita, että edes metatietoja ei jaeta näkyviin sairaalan ulkopuolelle. Tällä hetkellä Insite ratkaisussa on toimittu niin että hakuja saa suorittaa sairaalamme dataan vain sairaalan antamin valtuutuksin eikä edes kohortti kokoja näytetä ulospäin.

BC Platformsin R|QUEST implementointi myöhästyi puoli vuotta ja tuosta syystä sen laajempaa käytettävyyttä ei ehditty testaamaan tässä hankkeessa yhdessä klinikkojen kanssa. Tuosta syystä käyttökelpoisuutta on testattu laajemmin vain Custodixin Insite alustalla.

Kliinikon näkökulmasta positiiviset huomiot Custodixin Insiten käyttökelpoisuus

- ✓ Insite palvelee hyvin automatisoitua kohortin muodostusta.
- ✓ perusidea selvittää **tietyt kriteerit omaavat potilasmäärät** per sairaala (aikamääre, diagnoosi, toimenpide, lääkitys, sukupuoli, ikä diagnoosihetkellä jne) Tähän tarkoitukseen se soveltuukin hyvin ja sen käyttö/implementointi osaksi sairaalan järjestelmää ei maksa mitään ja mahdollistaa näyteikkunan klinisen tutkimuksen maailmanmarkkinoille. Maksajana toimii tutkimuksen tekijä.
- ✓ Käyttöliittymä näyttää selkeästi potilasmäärän vähenemisen sitä mukaa kun kriteereitä lisätään. **Insite listaa potilasmääriä.** Fibstroke-tutkimuksessa oli listattu myös mm. aivotapahtumien lukumääriä per potilas, minkälainen haku on Insitellä mahdotonta tai ainakin hyvin hankalaa.

kliinikon näkökulmasta Custodixin Insiten jälleenkehitystä kaipaavia asioita

-  Insitellä ei niinkään pystytä havaitsemaan diagnoosien ja lääkehoitojen onnistumisen trendejä tai analysoimaan dataa potilastapahtumien aikajanaa ajatellen. Kliinisen tutkimuksen näkökulmasta Insitea voidaan käyttää alustavaan arvioon, siitä kuinka monta potilasta kohorttiin saataisiin kyseisestä sairaalasta.
-  Insiten käyttökokemuksista: oli vaikea hahmottaa Insite termejä "at most"(maksimissaan) ja "at least"(ainakin).
-  Inklusio- ja eksklusio-kriteerien erotelu ja tarkastelu kaipaa selkeytystä käyttöliittymässä.
-  Monessa kohdassa inhimillisen käyttäjätulkinnan virheen riski on suuri, sillä käyttöliittymä ei ole vielä tarpeeksi selkeä.

6 Onnistuminen yhteistyössä muiden esituotantohankkeiden kanssa

Olemme soveltanut jo aiemmin ISAACUS projektissa saatuja oppeja ja tehnyt laaja-alaista yhteistyötä aiemmin esituotantohankkeissa syntyneiden poikkitieteellisten verkostojen kanssa. Koimme saavamme sieltä synergiaetua hankkeen toteutuksessa useissa eri vaiheissa.

7 Saatavuuspalvelujen nopeasti kasvava tarjonta ja globaalikilpailu

Kansainvälisessä mittakaavassa Suomessa on erinomaiset edellytykset kliinisen tutkimusten suorittamiseen maassamme. Meillä on kansakuntana tarjota kehittyvät ja luottamusta nauttiva terveydenhuoltojärjestelmä, mielenkiintoinen geeniperimä, laadukas rekisteridata jo vuosikymmenten ajalta, yksilöllinen henkilöturvatus ja teknistä kyvykkyyttä. Lisäksi tutkimuslainsäädäntöömme ollaan modernisoimassa, mikä osaltaan voisi helpottaa ja keventää Suomessa toteutetun tutkimuksen kustannusrakennetta. Lakiuudistuksista suurimpana valmistelussa on sosiaali- ja terveystietojen toisiokäyttöön liittyvää uutta lainsäädäntöä, jonka myötä eri rekistereistä ja lähteistä noudetun tiedon (RWD, real world data) hyödyntämismahdollisuuksia tulisi entisestään laajentamaan.

Custodix pyrkii laajentamaan Insite-ohjelmiston asiakaskuntaa Euroopassa, jossa sillä on EU:n IMI-hankkeen jälkeen asiakkaanaan 10 sairaalaa. Yhdysvaltalainen TrinetX on käytössä 24 sairaalassa, joista valtaosa on Yhdysvalloissa. Sairaalle liittyminen Custodixin Insite-käyttäjäksi on ollut tähän asti ilmaista Champion ohjelman avulla, jossa on partneroiduttu ja sitouduttu kehittämään Insite-ohjelmistoa yhdessä julkisen ja yksityisen puolen toimijoiden kanssa.

Saatavuuspalveluja kehitetään nyt globaalisti vauhdikkaasti ja fokus on pitkälti huomioida paitsi lääketeollisuuden tarpeita myös kunkin kehitykseen osallistuvan partnerin hyöty. Projektimme aikana löysimme hankkeessa testattujen Custodixin Insite palvelun ja BC Platformsin alustan lisäksi myös Clinerion ja TrinetX -nimiset saatavuuspalvelun mahdollistavat palvelut. **Kaikkien saatavuuspalveluiden suurin lupaus on nopeuttaa kliinisen tutkimuksen eri vaiheita niin, että tutkimussuunnittelusta tulee paitsi laadukkaampaa niin myös nopeampaa ja edullisempaa automatisoinnin ansiosta.** Myös potilasidentifiointi, -rekrytoinnin toivotaan nopeutuvan ja tutkimuslupien haun harmonisoituvan lääketeollisuuden näkökulmasta. Tämä ei tapahdu yksin teknisen kehityksen myötä vaan muutos edellyttää myös uudenlaisia lakeja. Jotta saatavuuspalvelu houkuttelisi mukaan myös datan hallinnoijia eli sairaaloita tarjoaa **osa toimijoista saatavuuspalveluratkaisusta ratkaisunsa käyttöön ilmaiseksi kyseiselle sairaalalle ja datan hallitsijoille työkaluja esim. sairaalan resurssien allokointiin ja seurantaan, sekä yhteistyön fasilitointia sairaaloiden välillä järjestelmänsä avulla. Myös mahdollisuus reaaliaikaisen datan laajemmasta käytöstä osana johtamista on huomioitu kaikista kehittyneimmissä saatavuuspalveluratkaisuisissa.**

Suomella on mahdollisuus nousta kansainvälisesti merkittäväksi rekisteri ja kliinisen tiedon hyödyntäjäksi, sekä maailmalla laadustaan tunnistetuksi tutkimuspartneriksi. Globaalit lääkefirmat ovat kiinnostuneita laadukkaasta tietovarannon hyödyntämisen mahdollisuuksista päätöksenteon tukena. Useista rekistereistä yhdisteltyä tietoa voidaan hyödyntää muun muassa tutkimuspotilaiden identifioinnissa, satunnaistettujen lääketutkimusten tulosten validoinnissa sekä uusien hoitojen kustannusvaikuttavuuden arvioinnissa. Lääketeollisuuden tutkijat voisivat jo nykyisen tietovarantomme avulla pyrkiä osoittamaan tiettyjen lääkkeiden tehokkuutta tai riskejä suhteessa vanhempiin hoitomuotoihin ja lääkityksiin. Harvinaissairauksiin liittyvät tutkimukset, hyödyttävät merkittävästi potilasta ja juuri niissä laajat ja laadullisesti tasokkaat tietovarannot ovat erityisen tärkeitä kustannustehokkuus näkökulmasta. Kansainvälisessä mittakaavassa Suomessa on erinomaiset edellytykset kliinisen tutkimusten suorittamiseen maassamme. Meillä on kansakuntana tarjota kehittyvät ja luottamusta nauttivat terveydenhuoltojärjestelmä, mielenkiintoinen geeniperimä, laadukas rekisteridata jo vuosikymmenten ajalta, yksilöllinen henkilöturvatus ja teknistä kyvykkyyttä. Lisäksi tutkimuslainsäädäntöömme ollaan modernisoimassa, mikä osaltaan voisi helpottaa ja keventää Suomessa toteutetun tutkimuksen kustannusrakennetta. **Lakiuudistuksista suurimpana valmistelussa on sosiaali- ja terveys- tietojen toisiokäyttöön liittyvää uutta lainsäädäntöä, jonka myötä eri rekistereistä ja lähteistä noudetun tiedon (RWD, real world data) hyödyntämismahdollisuuksia tulisi entisestään laajentamaan.**

Sairaanhoitopiiri, jossa on paikallinen saatavuuspalvelu tai jopa useampi eri saatavuuspalvelu näyteikkunoita, näkyy tutkimuksen maailmanmarkkinoilla kuhunkin laajaan verkostoon ja mahdollistaa tutkimusvolyymin kasvun. Saatavuuspalvelu tukee kliinisen tutkimuksista feasibility-kysymysten selvittelyn prosessia ja vastaa kysymykseen, että miten paljon tutkimus asetelmaan sopivia, tietynlaisia kriteereitä omaavia potilaita, kyseisen sairaanhoitopiirin alueelta löytyy.

Suurin etu sairaanhoitopiirille kuitenkin saatavuuspalvelussa syntyy siitä, kun sitä hyödyntävät suuret tutkimusorganisaatiot ja lääkeyritykset näkevät sairaanhoitopiirin potilaskohortin ja voivat ottaa sen suunnitelmissaan huomioon. Mikäli tutkijaorganisaatiot ja lääkeyhtiöt alkavat käyttämään enemmän kaupallisia saatavuuspalvelutuotteita, niiden käyttöönotto todennäköisesti lisää kliinisiä tutkimuksia myös Suomessa. Vastaavasti niiden käyttöpiirin ulkopuolelle jääminen voisi vastaavasti heikentää suomalaisten sairaaloiden markkina-asemaa.

Lisäetuna on graafisen helppokäyttöisen käyttöliittymän käyttöönotossa kohortin muodostamiseksi on, että sen tekeminen helpottuu ja jatkossa myös muut kuin asiaan vihkiytyneet tekniset osaajat voivat sen toteuttaa. Nykyiseltään poiminnan tekeminen vaatii paitsi datan asiantuntemusta, niin myös SQL ja R-skriptustaitoja.



Suurin osa VSSH:lle nykyiseltään saapuvista kansainvälisistä kliinisistä tutkimuksista saapuu kahdesta vaihtoehdoisesta syystä:

- 1) Tutkimus tehdään kansallisesti kaikissa yliopistollisissa sairaaloissa ja sen takia myös VSSH:llä.
- 2) Kliinikan johtava lääkäri tai muu maailmalla luottamusta herättänyt henkilö on tutkimusmyönteinen ja saa jalkautettua kansainvälisiä tutkimuksia myös VSSH:lle

On hyvä pitää mielessä, että lääkefirmat puntaroivat VSSH yhteistyössä sopivan yhteistyökliinikon lisäksi myös tutkimuksen hintaa, nopeutta rekrytoida potilaita tutkimukseen, sekä kohortin sopivuutta ja datan laatua suhteessa tutkimusasetelmaansa.

7.1 Saatavuuspalveluiden liiketoimintamallit

Liiketoiminnan malli

Saatavuuspalvelut joita testasimme	Ansaintamalli – liiketoiminnan malli
Insite Platform, Custodix 	Maksajina toimivat tutkimusorganisaatio - Champignon ohjelman aikana sairaala ei maksanut verkostoon liittymisestä. (CRO/Pharma/Akateemikot) jotka saavat kohortin Insiten kautta.
BC R QUEST, BC Platform 	Ansaintamalli on vielä kehitteillä. BC Platformsin edustajan mukaan on todennäköistä, että ansaintamalli on pitkälle sama kuin Eurooppalaisten markkinojen muilla saatavuuspalvelua tarjoavilla toimittajilla.



TrinetX

TrinetX-Ratkaisun testaus ei ollut osana projektia. Osallistuimme 18.10.2018 videotapaamisessamme demo.

TrinetX tarjoaa aggregoidun datan pohjalta kohortin muodostuksen lisäksi kohorteille suoraan mm. myös statistiikan: esim tunnusluvut, riskisuhteet p-arvoineen ja Kaplan-Meierit yms.

Kokonaisuutena alusta oli vaikuttava verrattuna projektissa testaamiimme alustoihin. Sairaallalle järjestelmän implementointi, tekninen ylläpito ja käyttö ei maksa mitään ja sitoutumispakkoa pitkälle ajalle ei ole ja erilaisia hallintamalleja on helppo toteuttaa osana tätä ratkaisua.

TrinetX on parhaillaan laajentamassa toimintaansa Eurooppaan ja moni toimija joka on ollut mukana investoimassa EHR4CR IMI projektin kautta Custodixin Insite Platform ratkaisuun ovat ottaneet ratkaisut rinnakkain. TrinetX liiketoiminnan mallissa maksajana toimii CRO, Pharma ja tutkimuksen teettäjä.

Voi ostaa oikeuden nähdä kaikki verkoston sairaalat kaikkine terapia-aluein tai sitten rajoitetusti ennalta valitut terapia-alueet pienempää kertakorvausta vastaan.

TrinetX on investoinut pian 100milj dollaria alustan kehitykseen ja se näkyy käyttäjäliittymän intuitiivisena suunnitteluna ja lisäpalveluina, jotka auttavat käyttäjää statistisin infograafein muodostamaan potilaskohorin tarjoamista tutkimusmahdollisuuksista selkeämmän kokonaiskuvan.

Mukana on jo 100 terveyden huollon tarjoajaa 17 eri maasta verkostossa, pääosin USA:n markkinoilta ja 32 pharma yhtiötä. 145 milj. potilasta/ 14.5 miljardia kliinistä faktaa

(Vertailuna EHR4CR IMI projektin budjetti josta Custodixin ratkaisu syntyi oli 16milj €, ja mukana Pharma-firmoja joilla mm. AstraZenecalla ja Sanofilla nyt molemmat järjestelmät käytössä.)

TrinetX:n demon perusteella vaikuttaa siltä että heidän pian 100 miljoonan dollarin investointi on tuottanut tuloksia ja on syntynyt paitsi potilaskohortteja muodostava saatavuuspalvelu niin myös tutkimuksen laajempia tarpeita tukeva alusta, joka hyödyntää myös strukturoimatonta tietoa tekstinlouhinnan ratkaisuja ja tekoälyä hyödyntäen.

TrinetX vaikuttaa demon perusteella käyttäjälähtöisesti suunniteltulta, mutta vaatisi paikallista testausta väittämien validoimiseksi. Alusta tekee paitsi kohortinmuodostusta aggregoidun datan pohjalta palvelun tutkijoita myös estimaatteja potilastapahtumiin liittyen. TrinetX kertoo datan pohjalta esim. mikäli joku lääke kohottaa riskiä sairastua tiettyyn tautiin.

Ehdotamme TrinetX testausta ja sen vertaamista muihin saatavuuspalveluratkaisuihin. TrinetX maksaa implementoinnin ja palvelimen ylläpidonkin, eikä velvoita sairaanhoitopiiriä mihinkään. Asiakas voi itse vastata hallintamallista ja niin halutessaan pysytellä testin aikana anonyyminä suhteessa verkoston muihin jäseniin. Sensitiivinen potilasdata ei missään vaiheessa poistu asiakkaan verkon ulkopuolelle vaan ainoastaan asiakkaan niin salliessamme meta-analyysin tiedot lähtevät TrinetX linkki-palvelimelta ulospäin. TrinetX-liiketoiminnanmalli on selkeä ja 32 lääketeollisuuden yritystä on jo mukana.

Hospital Effort vs TriNetX Proposed Investment Comparison

	Hospital Activity	Hospital Cost	TriNetX Activity	TriNetX Investment
Platform Cost				
○ Appliance hardware		\$0		\$20,000
○ Network onboarding		\$0	Y	\$10,000
○ Network monitoring / administration		\$0	Y	\$800 ¹
Initial Data Loading				
○ Extract Data from Source Systems	Y			
○ IT Engineer (one time requirement for data extract creation)	Y			
○ Data Analyst (Informatics) Resource (one time requirement for data extract creation)	Y			
○ Patient De-Identification Tool (one time cost)		\$0		\$8,000
○ Map Data & Load Data Extract Data Ingestion Engineers (e.g. mapping, cleansing, de-duping & testing)		\$0	Y	\$10,000
Periodical Data Refresh Loading				
○ IT Engineer (periodical data extract creation)	Y			
○ Data Analyst (Informatics) Resource (periodical data extract creation)	Y			
Operational Costs				
○ Network Connectivity Monitoring, Systems Management, Issue Triage and User Administration		\$0	Y	\$17,496 ¹
○ Clinical Operations Support Team			Y	\$24,000 ¹
○ Master Terminology / Code Cross Reference Management			Y	\$45,000 ¹
Project Cost / Investment Summary				
Year One Cost / Investment		\$0		\$135,296
Annual Re-Occurring Cost / Investment		\$0		\$87,296

Erona esim: Custodixin Insite-ratkaisuun on suurempi lääketeollisuuden näkyvyys. Käyttöliittumän helpompi tulkinta parantaa myös käyttökokemusta. Lisäksi hallintamallin luvataan olevan aiempien vaihtoehtoihin rinnastettuna paremmin huomioitu. Myös sairaalaa itseään hyödyttävät lisäinformatiikat, joilla voi tehdä nopeasti havaintoja datasta, toimivat niin klinikon kuin sairaanhoitopiirin johdon työn tukena. TrinetX tekoälyratkaisulla voi arvioida odotettavissa olevia käyntejä ja niiden ajankohtaa. TrinetX tietomallit ovat i2b2 ja OMOP Common Data Model (CDM).

Alla muutamia kuvia TrinetX-alustasta:

TriNetX

South West Oncology Symposium

Steve Provider - HCO Research Admin Help

295 PATIENTS 6 HCOs
Oct 16, 2018, 2:33 pm Steve Provider Research

Count Patients

Network Research 18 of 18 HCOs online

Population Any age / Any sex 19,195,906 patients on network

MUST Have

Search Term...

CANNOT Have

Search Term...

C67 Malignant neoplasm of bladder Stage 4 45,461

AND 100211 Chemotherapy, line 1 493,627

Event 1A + Add Time Constraint + Add Related Event

Add Terms

Drag and drop terms here

+ Create an event

TriNetX

South West Oncology Symposium Sales Demonstration Study

Steve Provider - HCO Research Admin Help

Count Patients

Network Collaboration - HCO A / HCO B / HCO D 3 of 3 HCOs online

Population Any age / Any sex 110,000 patients on network

MUST Have

Diabetes mellitus

CANNOT Have

Search Term...

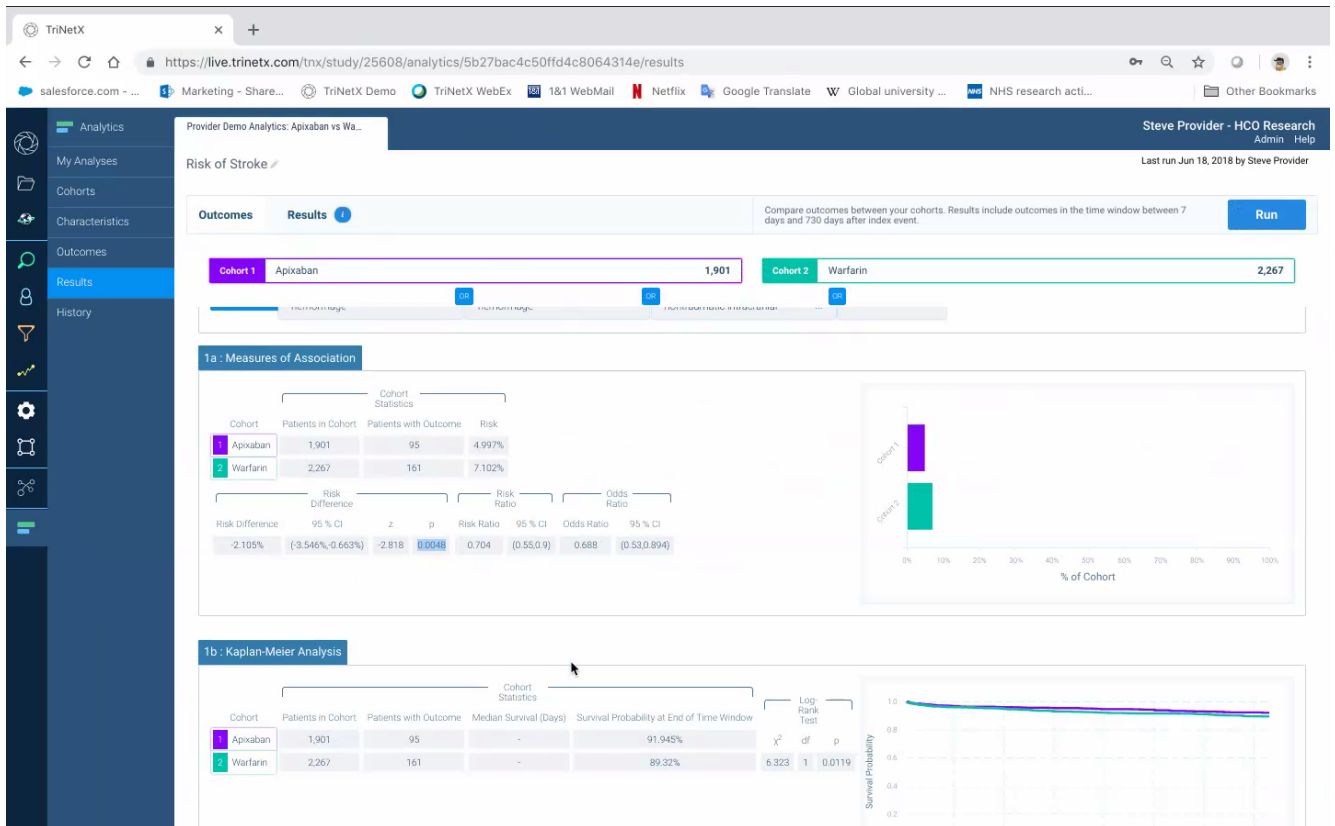
All D Demographics Dx Diagnoses P Procedures M Medications L Labs G Genomics

Organized by ICD-10 Show ICD-9 terms

	Patients
<input checked="" type="checkbox"/> E08-E13 Diabetes mellitus	15,770
<input checked="" type="checkbox"/> E08 Diabetes mellitus due to underlying condition	1,290
<input checked="" type="checkbox"/> E08.0 Diabetes mellitus due to underlying condition with hyperosmolality	110
<input type="checkbox"/> E08.00 Diabetes mellitus due to underlying condition with hyperosmolality without nonketotic hyperglycemic-hyperosmolar coma (NKHHC)	10
<input type="checkbox"/> E08.01 Diabetes mellitus due to underlying condition with hyperosmolality with coma	90
<input type="checkbox"/> E08.1 Diabetes mellitus due to underlying condition with ketoacidosis	190
<input type="checkbox"/> E08.2 Diabetes mellitus due to underlying condition with kidney complications	120
<input type="checkbox"/> E08.3 Diabetes mellitus due to underlying condition with ophthalmic complications	260

Show Terms with Zero Patients

Add to Query Cancel



8 Kansallisen saatavuustietokannan tiekartta

8.1 Koodiston osalta

Alla yhteenveto koodistoista, joista tarkempi selvitys projektin tuotoksessa: *2.2.2 Koodistojen siltaus kansainvälisiin koodeihin.pdf*.

Lääketieteellisten luokitusten yhdenmukaisella käytöllä mahdollistetaan teoriassa potilastietojen käsittely algoritmien avulla sekä muun muassa saatavuuspalvelu, tutkimus- ja tilastointitoiminnan, sekä kansainvälinen vertailu. Rakenteisempia tietoja syntyy uusien yhä kattavampien käsiteluo- kituksen, käyttöliittymien luokitukset huomioivien päivitysten, tekoälyn ja toki myös tekstinlouhinnan menetelmien kehittyessä. Ydintietojen (Master Datan) laadun varmistamiseksi tarvitaan paitsi käyttöliittymä, joka tukee yhdenmukaista kirjaamista, myös hyvät vakiintuneet käytännöt, joiden jalkautuksen merkitys vaikuttaa usein unohtuvan.

Rakenteisempia tietoja syntyy:

- uusien yhä kattavampien **käsiteluo- kituksen myötä**
- käyttöliittymien luokitukset huomioivien **päivitysten myötä** (mikäli sillataan osa kertomusmuotoisesta potilastiedosta rakenteisiin)
- tekoälyn & tekstinlouhinnan menetelmien kehittyessä**.

Ydintietojen laadun varmistamiseksi tarvitaan paitsi käyttöliittymä, joka tukee yhdenmukaista kirjaamista, myös hyvät vakiintuneet käytännöt, joiden jalkautuksen merkitys vaikuttaa usein unohtuvan.

Luokitellun tiedon avulla voidaan tulevaisuudessa yhä paremmin rakentaa osa kertomusmuotoisesta potilastiedosta. Suomessa nykyisin käytössä olevista luokituksista mahdollisesti tunnetuin on Maailman terveysjärjestön (WHO) määrittelemä ICD-tautiluokitus (International Classification of Diseases and

Related Health Problems), josta tällä hetkellä on käytössä kymmenes versio (ICD-10). ICD-10 tautiluokitus on ollut maassamme käytössä 22 vuoden ajan saavuttaen vakiintuneen aseman. Luokituksen alkuperäinen käyttötarkoitus sairastavuuden ja kuolemansyiden tilastoinnissa on laajentunut käsittämään myös kliiniset, potilaan hoitoon liittyvät käyttötarkoitukset. Kansalliset hoitosuosituksien sekä alueelliset ja paikalliset hoitoketjut on suurelta osin määritelty ja indeksoitu ICD-10 diagnoosien mukaisesti. Luokituksen mukaisia diagnooseja hyödynnetään myös kliinisen päätöksentuen järjestelmissä. Suomalainen ICD-10 luokitusversio noudattaa suurimmalta osin WHO:n luokitusta. WHO on aloittanut ICD-11 tautiluokituksen laatimisen. THL seuraa luokitustyötä ja on varautunut osallistumaan siihen. Uudempi alaa hiljalleen valtaava lääketieteellinen käsitejärjestelmä on graafiseen hilamalliin pohjautuva SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), jota ylläpitää kansainvälinen järjestö IHTSDO (International Health Terminology Standards Development Organisation). SNOMED CT:tä ei tosin ole vielä yhdessäkään kansallisesti Suomessa käytössä olevassa terveydenhuollon lähdejärjestelmissä implementoitu (EPIC aikoo implementoida sen Q3/2018 aikana).



KOODISTOSILTAUKSISSA SUOSITAMME

Custodix Insite-saatavuuspalvelu (datakatalogi) ratkaisussa suosittelemme käyttämään lähdejärjestelmien käyttämiä alkuperäiskoodistoja niin pitkälle kuin mahdollista, koska mitä vähemmän tarvitsee sillata, sitä vähemmän virheitä syntyy hakumootorin tuottamissa kyselyissä. Toisin sanoen mitä lähemmällä alkuperäistä rakenteista dataa ollaan, sitä pienempi virhemarginaali tuloksissa ja validointiresurssien allokoinnin tarve.

Terveydenhuollon organisaatiot nimittäin osaltaan myös tietoisesti poikkeavat koodistojen topographyasta suunnitelmallisesti eri perusteluin ja toteuttavat räätälöityä sisältöä myös lääketieteellinen käsitejärjestelmiin eri sairaanhoitopiireissä, jolloin koodistot eivät ole täysin yhteneväisiä edes kansallisesti, saatikka kansainvälisesti. Tästä syntyy laadullisesti epäluotettavaa päällekkäisiä tietoja, jotka eivät ole verrattavissa. Esimerkkinä eri murteita omaavasta luokituksesta voidaan käyttää Suomessakin käytettyä kansainvälistä patologian SNOMED koodistoa.

Tulevaisuudessa koodistojen implementointi lienee helpottuu, kun koodistojen käyttöä jatkuvasti yhteismitallistetaan paremmin kansainvälisesti yhdenmukaiseksi. Luokitusten ja rakenteellisen tiedon harmonisointia eteenpäin vievänä voimana järjestelmäkehityksen lisäksi ja katalyyttinä toimii mm. tekoälyn hyödyntämisen tarve hoidossa. Samaan aikaan tekstinlouhinta kehittyy ja nykyiseltään tekstikentästä olevasta tekstistringistä voidaan rakenteistaa dataa tietyllä luottamuksella. Vapaana tekstinä kirjatun potilastiedon koneellinen hyödyntäminen on kuitenkin nykyiseltään jokseenkin haastavaa, sillä yksinkertainenkin asia voidaan ilmaista lukuisin eri tavoin. Edeltä määriteltyjen ilmiöiden löytäminen potilaskertomuksista on kuitenkin kasvavissa määrin mahdollista: voimme palauttaa sanat perusmuotoonsa ja rikastaa niitä semanttisella tiedolla.

Esimerkkinä voimme käyttää Custodix Insite -ratkaisussa hyödynnettyjä koodistosiltauksia, jotka ovat Aurian tietopalvelussa parhaillaan validoitavina. Insitessa siltaukset on tehty ainoastaan TYKS-Lab laboratorio koodistojen mittauksista (82,3% on sillattu Custodix Insiteen) LOINC-koodistoon (Logical Observation Identifiers Names and Codes (LOINC)) ja toimenpiteitä on sillattu Suomen toimenpideluokituksesta, joka pohjautuu Pohjoismaiseen toimenpideluokitukseen NCPS:ään. NCSP:tä on siis sillattu ICD10PCS-koodistoon on 40% kattavuudella suhteessa erinlaisiin toimenpiteisiin. Muita siltauksia ei ole katsottu tarpeelliseksi toteuttaa tässä Custodix vetoisessa pilotissa huolimatta koodistojen kansainvälisistä käyttöeroista.

BC R|QUEST saatiin ulkopuoliseen testaus käyttöön vasta hankkeen lopussa, jonka takia sen testaus on jäänyt hyvin vähäiseksi. BCPodottavaa tällä hetkellä Aurian tietopalvelun koodistosiltausverifiointeja Custodixin Insite siltauksien pohjalta. Insite-datakatalogi ei tällä hetkellä siltaa toimenpiteitä SNOMED CT:lle sen takia, että se ei tuo lisähyötyjä, koska vain harva lähdejärjestelmä tukee SNOMED CT:tä koodistoa.

Keskeisimmät kansainvälisistä referenssiterminologioista joita Custodixin Insite ratkaisussa hyödynnetään ovat: ICD-10, SNOMED-CT, ICD-O(onkologia), NCSP(toimenpide), ATC(lääkkeet), ja LOINC (laboratorio).

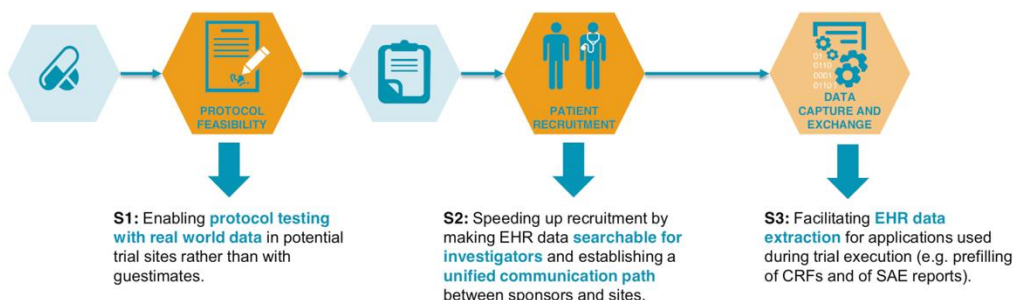
8.2 Tiekartta ja perustelut saatavuuspalvelun implementointiin

Varsinais-Suomen sairaanhoitopiirissä, VSSH:ssä, Custodixin Insite ratkaisussa tarjolla oleva tieto kattaa nyt diagnoosit, lääkityksen, laboratorioarvot sekä operaatiot. Nämä yleensä riittävät saatavuuskyselyihin, koska jos tarvitaan jotain enemmän, käyttäjien täytyy joka tapauksessa olla yhteydessä VSSH:hen datan saatavuutta ja lupamenettelyä koskien. Tällä hetkellä VSSH:ssä ei edes metatietoja näytetä ulkopuolisille ilman sairaalan myöntämiä oikeuksia.

Saatavuuskyselyiden käyttöliittymään yhteyteen on tunnistettu myös tarve pystyä määrittämään/tallentamaan kyselyssä haluttu data kohortti (potilaat ja muuttajat) saatavuuspalvelun avulla, ja mahdollistaa prosessi erilaisten tutkimslupien haku menettelyjen ja datan analysointialustalle pääsyn mahdollistamiseksi järjestelmän kautta. Terveystieteellisen tutkimuksen voi aloittaa sen jälkeen kun siihen tarvittavat luvat ja lausunnot on saatu ja ilmoitukset tehty. Tarvittavien lupien, lausuntojen ja ilmoitusten määrä riippuu siitä, minkä tyyppinen tutkimus on lainsäädännön kannalta.

"Automatisoidut saatavuuskyselyt nopeuttavat ja helpottavat huomattavasti työtämme uusien projektien suunnittelussa ja ideoinnissa ja lisäävät asiakkaidemme mielenkiintoa Suomalaisten biopankki- ja rekisteridatojen hyödyntämiseen tutkimuksessa" - Jaana Ahlmaa MD, PhD, Medical Affairs Director, Medaffcon Oy "

- Research & develop a **trustworthy service platform able to unlock clinical information stored in EHRs** for improving clinical research
 - Clear focus on three (3) relevant use cases

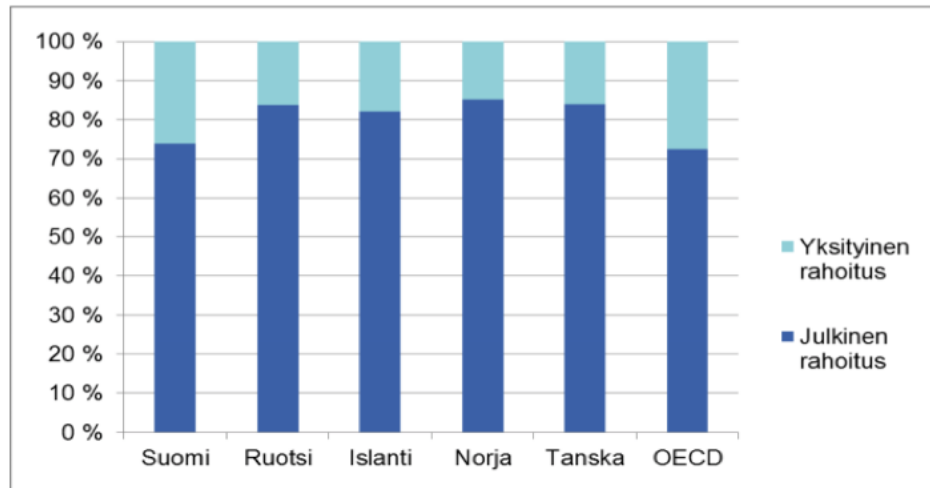


Kuva: EHR4CR lupaus siitä millä tavoin kliininen tutkimus nopeutuu EHR-ratkaisua hyödyntäessä. S1-Tutkimusasetelman testaaminen tosi elämän datalla/saatavuus. S2- potilaiden tunnistuksen ja rekrytoinnin nopeutus uuden kommunikointikanavan mahdollistamana. S3. Trial vaiheen dokumentit esitetytynä järjestelmässä nostaa nopeutta ja laatua.

Tutkimusrahoituksen saajat perustelevat kotimaassa tehtävän tutkimuksen tuottavan tehokkaiksi sekä turvallisiksi osoittautuvat diagnoosimenetelmät ja hoitomuodot hyödyntävän Suomalaisia potilaita nopeammin, kun niitä koskevaa tutkimusta tehdään jatkossakin suomalaisen terveydenhuoltojärjestelmän sisällä. Uusia kliinisen tutkimuksen rahoittajia etsitään myös yksityiseltä puolelta sillä valtion budjettirahoitus terveyden- huollon yksiköille yliopistotasoiseen tutkimukseen on supistunut 15 miljoonaa

euroon oltuaan 59 miljoonaa euroa vielä 2000-luvun alkuvuosina. **Saatavuuspalvelun hyödyntäminen on siis perusteltua myös sairaanhoitopiirin tutkimuksen rahoituksen näkökulmasta.**

Kuvio 7. Terveystenhuoltomenojen rahoitus Pohjoismaissa ja OECD-maissa 2016, %



Lähde: OECD, stats.oecd.org 2.5.2018 tilanteen mukaan. Suomen luvut tilastoraportin mukaiset.

Terveystenhuoltomenojen julkinen rahoitus pohjoismaissa suhteutettuna kulloiseenkin BKT:n Laina THL:n tilastosta.

<http://urn.fi/URN:NBN:fi-fe2018070326792>

Tilastoraportti : 20/2018

Parhaassa tapauksessa, lääkeyritykset tunnistaisivat Suomen ja linkki-implementoinnin omaavan sairaanhoitopiirin yhdeksi yhteistyökumppanikseen, jonka kanssa tutkimusta toteutettaisiin. Toki parhaan hyödyn realisoimiseksi tiivis yhteistyö ja käyttöliittymän laajempi validointi (koodisiltauksien verifiointi), käytettävyyden kehitys yhdessä Insite-ratkaisua myyvän Custodixin kanssa loisi paremmat menestysedellytykset. Tiiviimpi yhteistyö ja käyttäjien kouluttaminen paikallisesti ja asiantuntevasti varmistaisi parhaan näkyvyyden palvelun kautta.

8.3 Tekniset edellytykset paikallisen saatavuuspalvelun implementoinnille

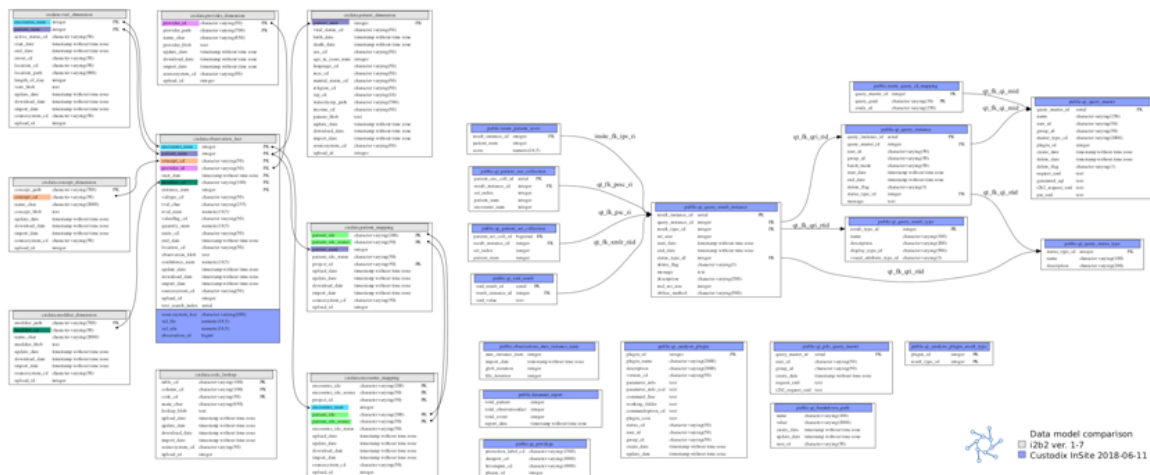
Kaikki saatavuuspalvelut edellyttävät sen toteuttajalta tietovarastoratkaisua. Tietovarasto on järjestelmä, johon kerätään organisaation tietoa eri lähteistä yhdenmukaiseen rakenteeseen. Tietovarastossa säilytettävien tietojen tallentamiseen on olemassa erilaisia tietomalleja.

VSSH:n tietoaaltaan päälle on rakennettu erilaisia tietomalleja ja näkymiä eri lähteistä. Näistä tiedoista valikoidaan ydintiedot, joista toimitetaan pseudonymisoitu kopio sairaalan palomuurien sisälle olevalle linkkipalvelimelle. Linkkipalvelin kääntää datan i2b2-tietomalliin sopivaan muotoon, laskee siitä kohortteja ja tuottaa tilastolliset näkymät aineistolle. (viite: 20160527 - InSite Core dataset - specification - v.1.2.docx)

Saatavuuspalveluratkaisut suosivat sairaalan palomuurin sisään implementoitavaa linkkipalvelua. Useimmat linkit on toteutettu i2b2-tietomallilla, jonne tietovaraston data käännetään. I2b2-tähtimallin mukainen relaatiotietokanta koostuu yhdestä tapahtumatietoa sisältävästä faktataulusta ja useasta tapahtumaa kuvaavasta dimensiotaulusta. Tähtimalli on kompromissi toisteisuuden välttämisen ja kyselyiden tehokkuuden välillä. Rakenteeltaan tähtimallinen tietomalli on erittäin selkeä ja ymmärrettävä. Mainittakoon, että monet tietovarastoissa käytettävät ohjelmistot on optimoitu erityisesti tähtimallille.

I2B2-tietomalli vs. Custodix implementaatio

I2B2 suunniteltu nimenomaan "**kohortin tunnistukseen**" ympäristöön jossa on useita erilaisia lähdekirjastoja. Toimii myös pseudonymisoidussa ja aggregoidussa tietoympäristössä. Custodixin tietomalli pohjaa I2B2-tietomalliin ja on sen kanssa yhteensopiva.



9 Käyttöoikeuksien määrittäminen

Tuotantovaiheessa sairaalan sisäisessä käytössä saatavuuspalvelun käyttöoikeudet tulisi jakaa seuraaviin ryhmiin:

Pääkäyttäjät (Power users)

- o Pääsy tunnistelliseen dataan (maskattu tai maskaamaton). Pääsyoikeuksien myöntämisluva.

Peruskäyttäjät (Regular users)

- o Ei pääsyä tunnistelliseen dataan.
- o Pääsyaggregoituundataan(n-määriin tms.)
- o Pääsy myös tunnistelliseen dataan, mikäli hoitosuhde potilasjoukkoon.

Tehokäyttäjät

- o Voi luoda kohortteja ja pääsy pseodonymisoituun dataan

Lisensioitu / osoitettu tutkija

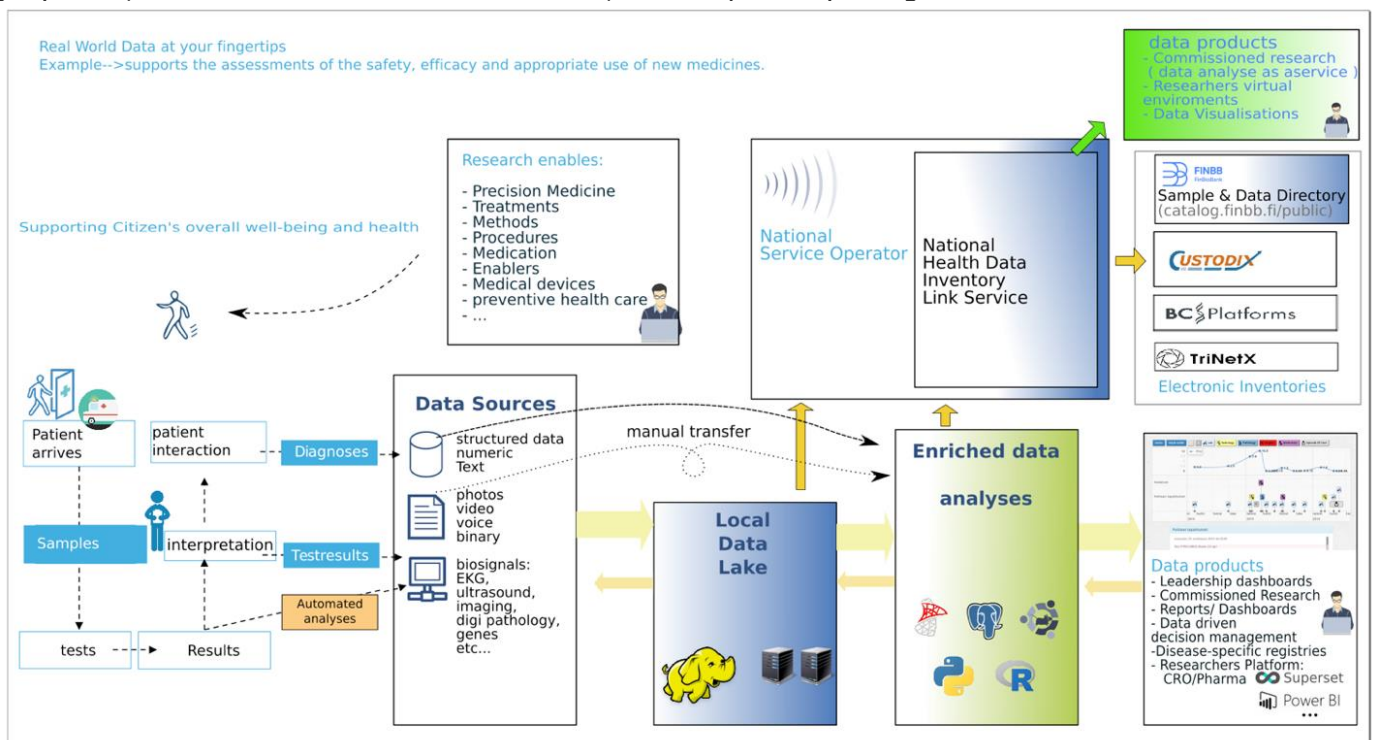
- o Pääsyvaintutkimusluvanmukaiseenkohorttiin/datasettiin

10 Palveluoperaattori on välttämätön toimija

Hankkeessa tehty johtopäätös on, että kansalliset tutkimuskäyttöön soveltuvat terveystiedot olisi hyvä tarjota yhden ikkunan ja prosessin kautta koko kliinisen tutkimuksen maailmanmarkkinoille. Se tarkoittaa

yhä kansallista näyteikkunaa (one-stop-shop), yhteisiä pelisääntöjä ja projektinhallintaa. Lisäksi tarvitaan valmius vastaanottaa rikastettua dataa ja huolehtia, että tietoluovutukset tehdään keskitetysti ja tietoturvasestisesti. Tämä vaatii sopivia lakiuudistuksia, laaja-alaista yhteistyötä sosiaali- ja terveyssektorilla sekä teknisen osaamisen hankkimista toimijoille itselleen, jotta todellinen kokonaisvastuu voidaan kantaa.

Saatavuuspalvelun kansallinen toteutus kannattaa tehdä yhdessä Suomen Biopankkiosuuskunnan palveluoperaattorihankkeen kautta. Palveluoperaattori on käytännössä välttämätön toimija, jotta yhteinen saatavuuspalvelu voidaan toteuttaa. Saatavuustiedon yhdistämiseen on useita teknisiä tapoja, joita voidaan käyttää myös yhtä aikaa. Datan kerääminen ja yhteismitallistaminen useista lähteistä on vaikeaa ja työlästä, mutta tämä voidaan tehdä vaiheittain, mikä helpottaa työn organisointia.



Kuva: Palveluoperaattorin rooli potilaskohtaisen tiedon yhdistämisessä ja saatavuustiedon tietohakemiston laskennassa

Kansallisen Palveluoperaattorin haasteet kiteytyvät tällä hetkellä seuraavasti:

- (1) datan saantivaikeuksiin eri lähteistä – resurssit ja osaaminen
- (2) tietojen yhteismitattomuuteen
- (3) tietoturva-vaatimusten ja suuren tallennustilan aiheuttamiin – toisinaan hyvin perusteltuihin – hidasteisiin
- (4) kaupallisen mallin mahdollisiin ongelmiin ja
- (5) loppukäyttäjän kokemaan palvelun toimimattomuuteen
- (6) erityyppisiin lakitulkintoihin käytännön tasolla.

Ehdotamme, että

- THL vastaa terminologiapalvelusta keskitetysti.
- Analyysi kansallisesta saatavuuspalvelun haasteesta tehdään, kun kansallista yhteistyötä on kokeiltu Biopankkiosuuskunnan piloteissa

- Geenitedon ja muun massadatan yhdistämisestä tiedetään jo, että kansallinen, Suomessa sijaitseva master-kopio datasta helpottaa sekä analyysejä että luvituksen järjestämistä. Vaikka datalla olisi monta omistajaa, datasta tarvitsee olla vain yksi fyysinen kopio (joka on varmistettu riittävällä luotettavuustasolla).

	Saatavuustilastojen toteutustapa	Toteutus	Hyödyt	Haitat
1	Hajautettu tilastointi	Jokainen rekisterinpitäjä laskee määrämuotoisen tilaston aineistoistaan	Siirrettävä data on täysin anonyymiä	Tilastot voivat vääristyä pahasti, koska rekisterinpitäjiä käsitellään erillisinä ja yksittäisen henkilön tiedot eivät yhdisty
2	Yhdistetty tilastointi	Jokainen rekisterinpitäjä tuottaa tilastointiin riittävän rivitason datan, minkä palveluoperaattori yhdistää	Tilastointi on tarkkaa, koska tieto yhdistyy rekisterinpitäjien yli	Palveluoperaattorin aineisto on rajattu
3	Rivitason datan yhdistäminen	Palveluoperaattori säilyttää kattavaa ja laajaa kopiota rekisterinpitäjien rakenteisesta datasta	Ennalta määrättyjen tilastojen lisäksi voidaan tehdä datan muutakin analyysiä, kuten yksittäisen asiakkaan toivomia erityistilastoja	Palveluoperaattorin keskitetyn järjestelmän pitää olla erittäin huolellisesti suojattu
4	Laskenta pyydettyäessä	Kuten yhdistetty rivitason data, mutta laskenta tehdään lennosta rajapintakutsujen avulla aina uusimmalla datalla	Tarkka ajanmukainen tilasto	Monimutkainen tekninen toteutus ja laskennan hitaus. Soveltuu todennäköisesti vain eräajoihin.

Taulukko. Projektin kokemusten nojalla tapa (3) "Yhdistetty rivitason data" vaikuttaa toteutuskelpoisimmalta vaihtoehdolta. Kunhan palveluoperaattorin järjestelmien tietoturva huolehditaan. Tapa (3) tarjoaa tehokkaan tavan laskea tavanomaisimmat saatavuustilastot tarkasti, sekä mahdollisuuden kehittää tilastoja edelleen.

Hankkeessa on havaittu ongelmia palveluoperaattorin toteutuksessa, alla asia esitetty mahdollisine ratkaisuineen.

	Ongelma	Ongelman syy	Mahdollinen ratkaisu
1	Datan saanti eri lähteistä	Laillisuuskysymykset	Velvoittava lainsäädäntö
2	- " -	Suuri määrä dataa tuottavia lähdejärjestelmiä	Joukko maakunnallisia altaita, joihin data kerätään ja yhteismitallistetaan. Palveluoperaattorin data yhdistetään tästä tietojoukosta.
3	- " -	Haluttomuus muuttaa vallitsevia toimitatapoja (liittyen datan käsittelyyn ja oletettuun "omistajuuteen")	Kirkas kansallinen strategia ja valtion ohjaus.
4	- " -	Kustannukset	Riittävä alkurahoitus ja palkitseminen yhdistetyn datan perusteella. Avoimen lähdekoodin käyttö yksinoikeudellisten tekniikoiden sijaan (jolloin toteutus voidaan aina kilpailuttaa)
5	- " -	Puutteet teknisessä osaamisessa	Osaamisen kehittäminen itse rekisterinpitäjille
6	- " -	Huomattavat viiveet toteutuksessa.	IT-alihankintaketjujen välttäminen: Rekisterinpitäjällä pitää olla riittävät tekniset valtuudet toimia esimerkiksi pitämällä tuotantojärjestelmien ylikäyttäjät omassa organisaatiossa.
7	- " -	Epäselvyydet ansaintalogiikassa tai kustannusten jakamisessa	Yksikäsitteiset sopimukset palveluoperaattorin ja rekisterinpitäjien välillä. Malli kustannusten ja tulojen jakamiselle.
8	Datan yhteismitattomuus	Erilaiset lähdejärjestelmät, erilainen lähtödata ja koodistojen poikkeavuus.	Kansalliset koodistot ja esim. THL:n kansallisen koodistopalvelimen parempi hyödyntäminen. Ymmärrys, että tarvitaan huomattava rahallinen ja henkinen investointi datan yhteismitallistamiseen.
9	- " -	Tilastot eivät ole vertailukelpoisia	Mittareiden yhtenäistäminen kansallisesti. Esimerkkinä Kuva-mittaristot.
10	Tietoturvaan liittyvät epävarmuudet	Teknisen tietoturvaosaamisen puute, jolloin ehdotettuja teknisiä ratkaisuja syystäkin epäillään	Kansallisen tietoturvaryhmän muodostaminen asiantuntijoista tekemään käytännön testiympäristöjä joita myös testataan. Testiympäristöt ja tulokset julkaistaan avoimena lähdekoodina, koska "security by obscurity" ei tunnetusti toimi

			(esimerkiksi sisäistä hyökkääjää vastaan, joka tuntee järjestelmän toiminnan).
11	- " -	Inhimillinen pelko tuntemattomasta	Kannustin datan hyödyntämiseen. Jos ainoastaan virheistä rangaistaan mutta onnistumisista ei palkita, syntyy kannustin olla tekemättä mitään.
12	Palveluoperaattorin kaupallisen mallin epäselvyys	Useat yksityisetkin toimijat pyrkivät tarjoamaan palveluoperaattorin ydintoiminnan. Ansaintalogiikka ei ole vielä täysin ratkaistu. Insite ja TrinetX järjestelmien maksajana toimii tutkimuksen tekevä organisaatio.	Selvä malli, jossa palveluoperaattorin ydintoiminnot, kuten keskeiset tietojärjestelmät ja tietomalli, ovat avointa lähdekoodia. Rajapinnat eri palveluntarjoajille avataan tämän ytimen päälle ja kukin tarjoaa omanlaisensa näyteikkunan. Toisen näyteikkunassa geenitiedon käsittely on mielekkäämpää, kun taas toinen tarjoaa statistisia automatisointeja kohortin tunnistuksen lisäksi.
13	Tietohakujen käytettävyys	Käyttöliittymän puutteet ja hitaus tietohauissa	Riittävän tehokas palvelinalusta sekä käyttöliittymän kehittäminen palautteen perusteella.
14	- " -	Käyttäjän vaikeus ymmärtää hakemaansa aineistoa	Infografiikan käyttö ja kattavat visualisoinnit (joita nyt testatut tekniikat eivät vielä tarjonneet)
15	- " -	Hakuehtojen rajallisuus käyttöliittymässä	Mahdollisuus pyytää erityistilastoja erillistä maksua vastaan, jolloin tietokoordinaattori suorittaa haun käsin (ja lisää mahdollisesti kehitetyn toiminnallisuuden käyttöliittymään)
16	Jatkokysymysten ja palvelupyyntöjen käsittely	Tietokoordinaattoreiden vaikeudet vastata datasta nousseisiin kysymyksiin, koska kaikkien erikoisalojen tuntemusta on mahdotonta keskittää.	Datan syntymekanismit tuntevien asiantuntijoiden verkosto ja riittävä rahallinen tai tunnustuksellinen korvaus konsultoinnista. Tunnustuksellinen korvaus voidaan toteuttaa avoimen lähdekoodin kehitysalustoilta tuttujen kontribuutiolistojen ja tähtiluokitusten avulla.

11 Palautetta Sitran toiminnasta

Yhteistyö Sitran kanssa on ollut mielekästä, toiminut sovitusti ja suunnitelman mukaan. Olemme jatkossakin kiinnostuneita tekemään yhteistyötä Sitran kanssa.