

Kohti kansallista sote-tietojen saatavuuspalvelua

Sosiaali- ja terveydenhuollon aineistojen
saatavuusarkkitehtuurin skenaario ja valmisratkaisujen testaus

Väliuportti

Versiohistoria:

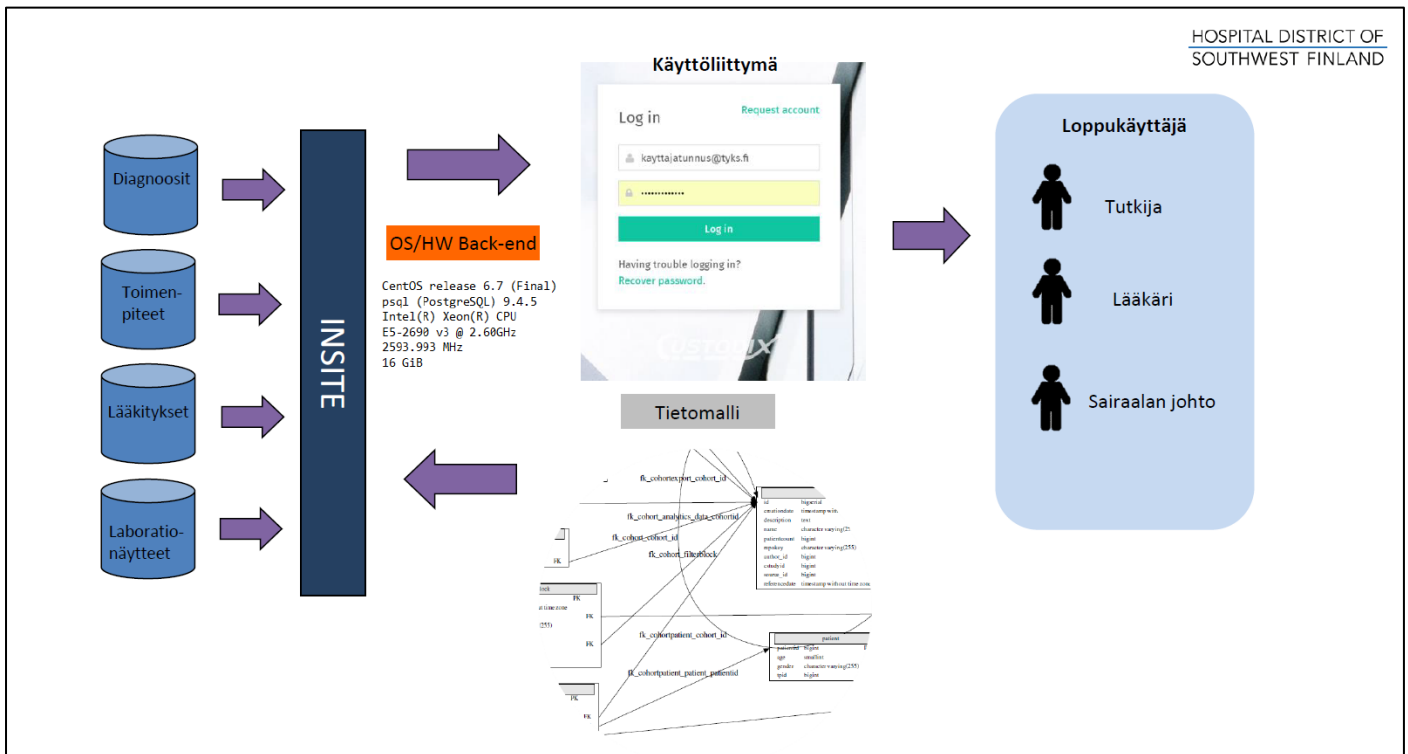
Versio	Päivämäärä	Valmiusaste (vedos/ehdotus/päivitys)	Kirjoittaja(t)	Huomautukset
0.9	1.12.2017	Vedos	Arho Virkki ja Juha Varjonen	Ensimmäinen versio

Sisällys

Saatavuuspalvelu paikallisesti VSSHP:n käyttäen InSite -ratkaisua	3
Custodixin EHR4CR Insite – koodisto: ICD10PCS ja LOINC koodistoa koskevat siltaukset NCSP ja TYKSLAB koodistoa vasten	8
Custodixin EHR4CR Insite –käyttöoikeudet	8
Havaintoja käyttöoikeuksista:	8
Analyysi saatavuuspalvelun toteuttamisen edellytyksistä, rajoituksista, haasteista ja hyödyistä kansallisesta näkökulmasta	9
Arkkitehtuuriskenaario	9
Terminologioiden siltauksen toteutuksesta analyysi	9
Vertailu käytetyn ratkaisun ja toisen saatavuuspalvelun välillä	9
Synteettisen datan generointiin ohjelmakirjasto	10

Saatavuuspalvelu paikallisesti VSSHP:n käyttäen InSite -ratkaisua

Ensimmäisen pilottivaiheen raakadata siirrettiin kertaluontoisesti vuonna 2016 Custodixin ylläpitämään tietokantaan (PostgreSQL), jolloin käytössä ei vielä ollut nykyistä tietoallasta. Medbit suoritti latauksen tekstitiedostoihin, josta data kerättiin testimielessä saatavuuskantaan.



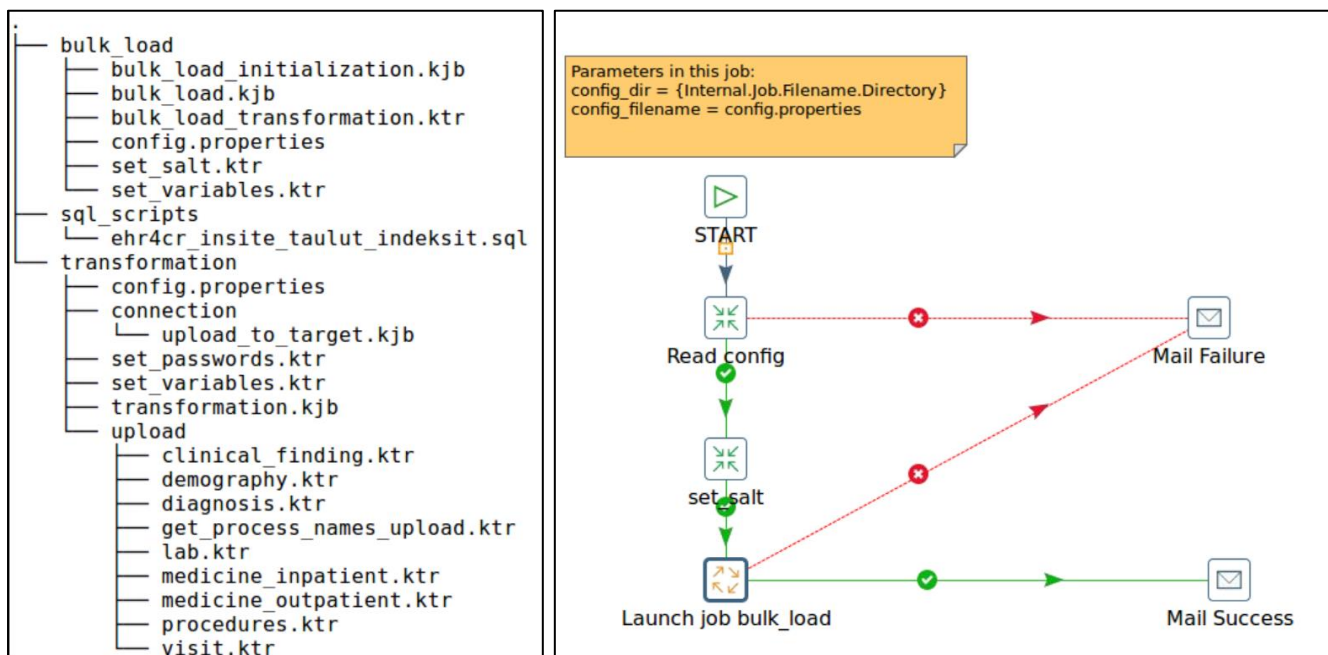
Data päivitettiin Q3 /2017 kertaluontoisesti siirtymään uudesta lähteestä (KTP altaan kautta yhteismitallistama tietokanta) Pentahon PDI (Kettle) ETL-työkaluja käyttäen. Datan päivityksen lisäksi koko ETL-prosessi meni tässä vaiheessa uusiksi. Vuosineljänneksellä Q4/2017 data on tarkoitus muuttaa päivittymään automaattisesti (inkrementaalinen lataus) joka päivä klo 06:00 (Suomen aikaa). Data on tässä vaiheessa jo peitenimillä suojattua eli pseudodataa, ja suojauksen tekijä (tässä tapauksessa KTP) voi palauttaa datan tunnistettavaksi. Custodix jatkaa edelleen joulukuun 2017 aikana datan jalostamistoimenpiteitä uudesta tietokantadatasta. Kun loppuharmonisointi on tehty, uusi data saadaan näkyville InSite-käyttöliittymään. Tämä mahdollistaa lähes reaaliaikaisen saatavuuspalvelun sairaalan sisäiseen käyttöön. Tästä seuraavat jatkotoimenpiteet kohdistuvat datan validointiin.

VSSHP palkkasi syksyllä uuden arviointiylilääkärin (Juuso Blomster), joka perehtyy työhönsä ja vastualueisiinsa 2017 loppuun mennessä, jaalkaa johtaa vuoden 2018 alusta InSite-alustan validointia Finbstroke-tutkimuksen aineistolla. Erityisesti selvitetään,

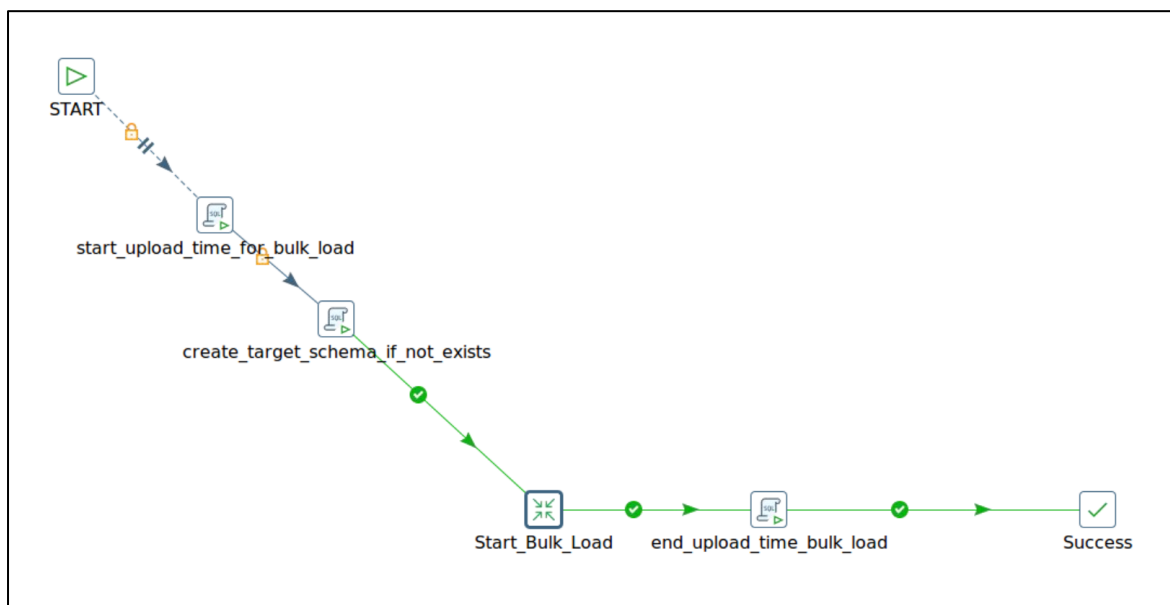
- löytyvätkö (a) samat potilaat ja
- (b) samat tapahtumat InSite-palvelun avulla, kun mitä

aikaisemmin on kerätty käsin tätä tutkimusta varten? Toimintaan on haettu tutkimuslupa.

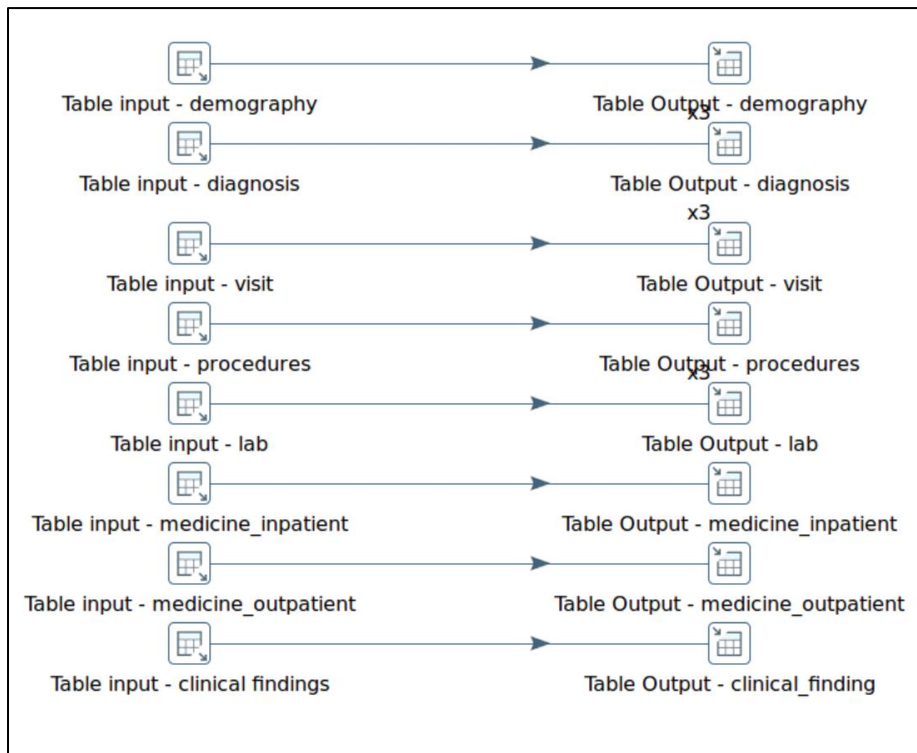
Insite -palvelun luovutusprosessin ETL-scriptit on tallennettu Git-versionhallintaan (Luovutusprosessit.git/ehr4cr/). Hakemistorakenne näyttää seuraavalta (Kuvankaappaus 1):



Näyttökuva 1. Pentaho PDI (Kettle) – hakemistorakenne ja massalatauksen päätaso.



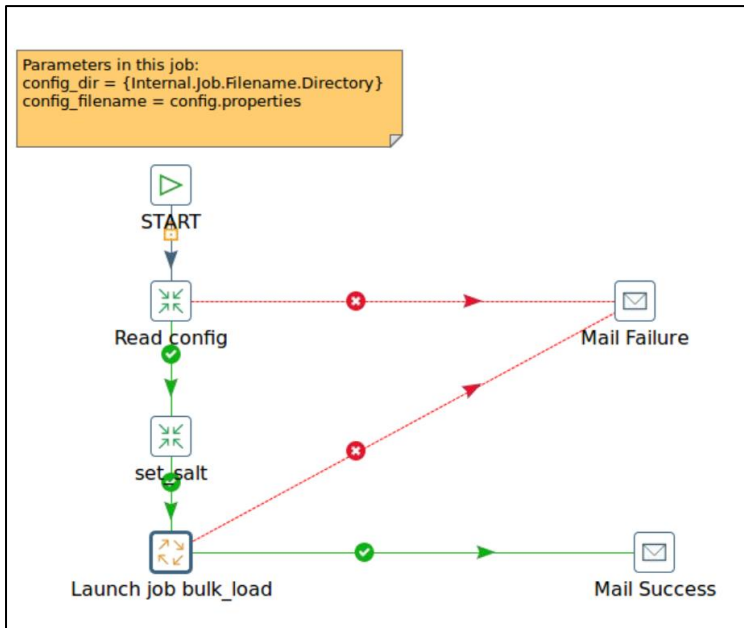
Näyttökuva 2. Pentaho PDI (Kettle) – ehr4cr Insite bulk load – alustustaso



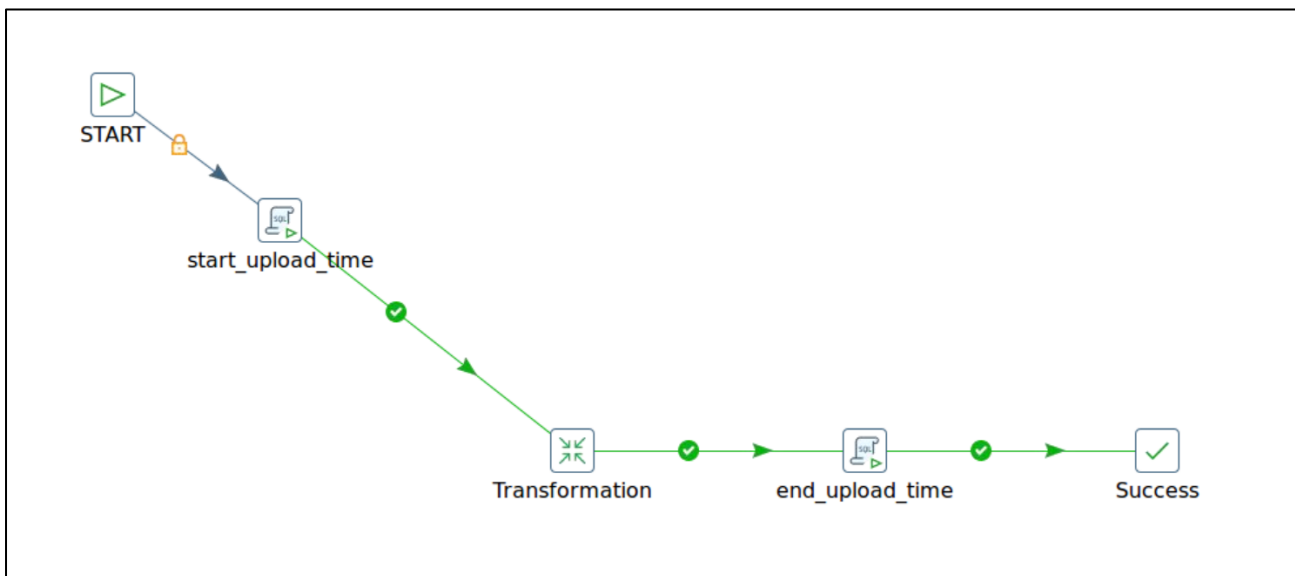
Näyttökuva 3 Pentaho PDI (Kettle) – ehr4cr Insite incremental load – datansiirtovaihe

File Edit View Tools Help						
100 rows						
	upload_id [PK] serial	is_uploaded boolean	target_table character varying	last_load timestamp without time zone	current_load timestamp without time zone	
1	1	TRUE	demography	2017-09-09 11:26:40.137043		
2	2	TRUE	diagnosis	2017-11-10 13:50:19.702709		
3	3	TRUE	procedures	2017-09-09 11:28:17.877127		
4	4	TRUE	lab	2017-09-09 11:26:47.518221		
5	5	TRUE	medicine_inpatient	2017-11-10 14:48:56.737905		
6	6	TRUE	medicine_outpatient	2017-09-09 11:27:45.578777		
7	7	TRUE	visit	2017-09-09 11:28:21.290352		
8	8	TRUE	clinical_finding	2017-09-09 11:26:35.530496		
*						

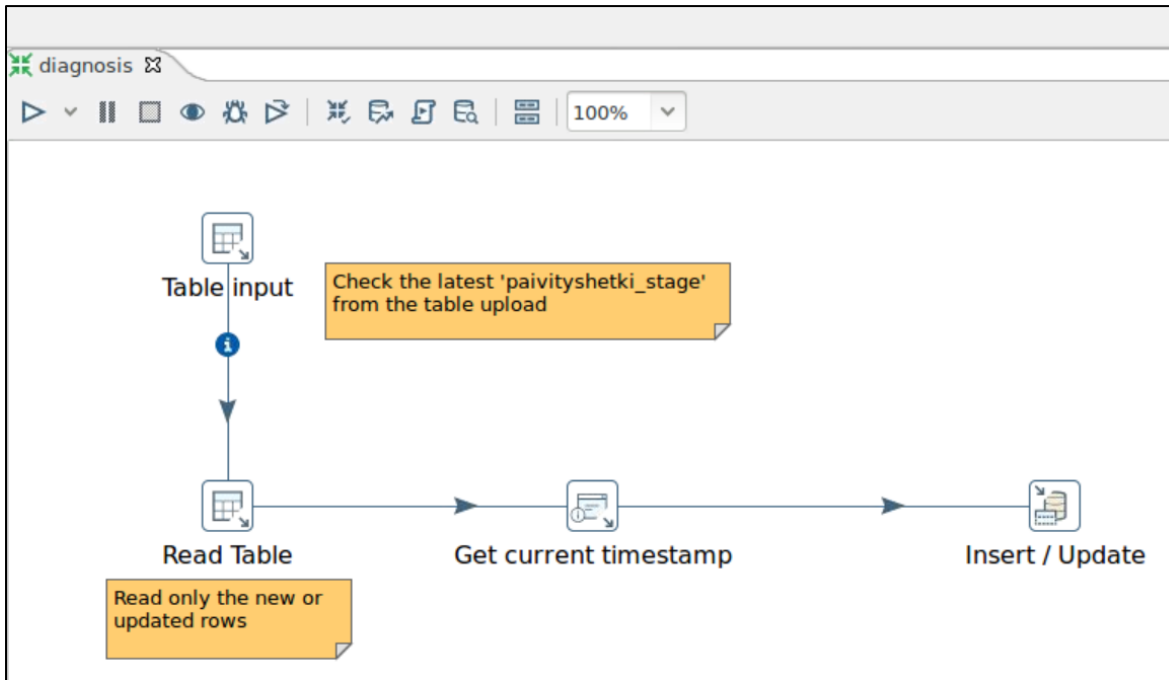
Näyttökuva 4. Siirtoaikojen aikaleimoja ylläpidetään tietokannan taulussa päivitystä varten.



Näyttökuva 5. ETL transformaatio (inkrementaalinen lataus) päätasolta katsottuna (kuvankaappaus 6):

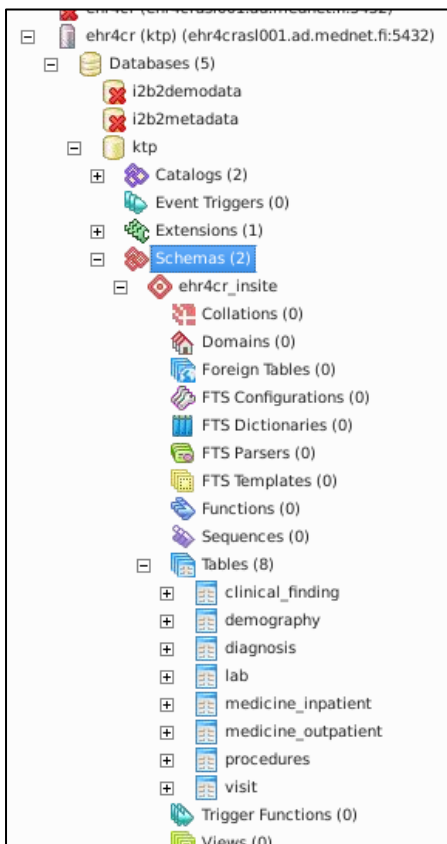


Näyttökuva 6. ETL transformaatio (inkrementaalinen lataus) alustustasolta katsottuna



Näyttökuva 7. ETL transformaatio (inkrementaalinen lataus) datansiirtovaiheesta katsottuna, esimerkissä alla kohteena diagnoosien siirto

Datsiirron vastaanottavassa päässä toimii Medbitin ylläpitämä Linux RedHat -palvelin, jossa implementoituna PostgreSQL tietokanta. KTP:lla on järjestelmävalvojan oikeudet kyseiseen palvelimeen. Vastaanottavassa päässä taulurakenne raakadatalle näyttää seuraavanlaiselta (kuvankaappaus 9):



Näyttökuva 8. PostgreSQL-taulu – ehr4cr Insite – vastaanottava pää

Custodixin EHR4CR Insite – koodisto: ICD10PCS ja LOINC koodistoa koskevat siltaukset NCSP ja TYKSLAB koodistoa vasten.

Koodistosiltaukset perustuvat pilottivaiheen tuotokseen, tarkoitus on saattaa koodistojenkin siltaukset jatkuvan päivityksen alaisuuteen. Custodixilta vaadittava täysi raportointi toteutusratkaisusta ja lähdekoodit ovat vielä tekeillä. Koodistosiltaukset ovat ilmeisimmin toteutettu taulussa insite (stage_ehr4cr.insite). Se millä tavalla itse insite-taulu on muodostettu, on vielä epäselvää ja se miten Custodixin prosessi käsittelee ja muokkaa siltaukset harmonisointivaiheessa, ei myöskään ole vielä selvää.

Mappauksien määrä insite-tilussa:

```
> i3[, .N, V1][order(-N)]
```

	V1	N
1:	SNOMED-CT	1486415
2:	ICD10PCS	184326
3:	LOINC	131709
4:	ICD10CM	106065
5:	ICD9CM	88982
6:	ATC	6113
7:	2.16.840.1.113883.6.3	2643
8:	ICD-O	1435
9:	FI_TURKU_LAB	843
10:	NCSP	439
11:	MOD	6
12:	INSITE	1
13:	Anatomic_pathology	1

Esimerkki NCSP-mappauksesta:

```
1:\\insite\\71388002\\ICD10PCS_G\\ICD10PCS_GZ\\ICD10PCS_GZB\\((NCSP_AA400))\\
```

R-scriptillä puhdistettu ICD10PCS->NCSP-mappaus:

	ryhmatunnus_insit	koodi	ryhmatunnus_ic10pcs_insit	koodi_ic10pcs_insit	selite_insit
1:	NCSP	AA400	INCD10PCS	GZB	Sähköshokki psyykkisen sairauden hoitona
2:	NCSP	AA5HE	INCD10PCS	B040	Lapsen aivojen ultraäänitutkimus
3:	NCSP	ABC16	INCD10PCS	0SB20	Mikrokirurginen välilevytyrän poisto lannerangasta
4:	NCSP	ABC36	INCD10PCS	01NB	Hermojuuridekompressio lannerangan alueella
5:	NCSP	ACC51	INCD10PCS	01N5	Keskihermon vapautus

Esimerkki LOINC-mappauksesta:

```
1:\\insite\\275924004\\LOINC_LP32742-6\\LOINC_LP16738-4\\LOINC_LP43937-9\\LOINC_31046-6\\((FI_TURKU_LAB_-RF-O (3377)))\\
```

R-scriptillä puhdistettu LOINC->TYKSLAB-mappaus:

	ryhmatunnus_insit	koodi	ryhmatunnus_loinc_insit	koodi_loinc_insit	selite_insit
1:	FI_TURKU_LAB	-RF-O (3377)	LOINC	31046-6	-Reumafaktori (kval)
2:	FI_TURKU_LAB	12hKrea-C1 (574)	LOINC	2163-4	12 tunnin kreatiniinin poistuma
3:	FI_TURKU_LAB	Afer-CD19 (11564)	LOINC	20592-2	Afer-CD19-positiiviset B-lymfosyytit
4:	FI_TURKU_LAB	Afer-CD3 (11561)	LOINC	8122-4	Afer-CD3-positiiviset T-lymfosyytit
5:	FI_TURKU_LAB	Afer-CD4 (11562)	LOINC	80223-1	Afer-CD4-positiiviset T-lymfosyytit

Custodixin EHR4CR Insite –käyttöoikeudet

Tämä osuus nojautuu Christina Müller christina@custodix.com toimittamaan esitykseen 6.9.2017 (slide 16 -20), jossa esityksen lopussa kerrotaan alustavasti käyttöoikeuksista. Tämä prosessi ei ole lopullinen.

Havaintoja käyttöoikeuksista:

Tuotantovaiheessa sairaalan sisäisessä käytössä, InSite käyttöoikeustasoja täytyy tarkentaa. Käyttöoikeudet tulisi Jakaa seuraaviin ryhmiin:

2017-12-01

Väli raportti

8

- Pääkäyttäjät (Power users)
 - Pääsy tunnistelliseen dataan (maskattu tai maskaamaton). Pääsyoikeuksien myöntämislupa.
- Peruskäyttäjät (Regular users)
 - Ei pääsyä tunnistelliseen dataan.
 - Pääsy aggregoituun dataan (n-määriin tms.)
 - Pääsy myös tunnistelliseen dataan, mikäli hoitosuhde potilasjoukkoon.
- Tehokäyttäjä
 - Voi luoda kohortteja ja pääsy pseodonymisoituun dataan
- Lisensoitu / osoitettu tutkija
 - Pääsy vain tutkimusluvan mukaiseen kohorttiin / datasettiin

Analyysi saatavuuspalvelun toteuttamisen edellytyksistä, rajoituksista, haasteista ja hyödyistä kansallisesta näkökulmasta

Analyysin aloitetaan 2018 ja saadaan valmiiksi projektin loppuun.

Arkkitehtuuriskenaario

Arkkitehtuuriskenaariota on aloitettu keräämällä Custodix InSiten ja BC|RQuestin teknistä dokumentaatiota. Skenaarion kirjoittaminen aloitetaan 2018 alussa ja tehdään valmiiksi projektin loppuun mennessä.

Terminologioiden siltauksen toteutuksesta analyysi

Alustava tutkimus on aloitettu. Tärkeintä on selvittää, voidaanko saada aikaiseksi kansallinen koodistopalvelu, vai onko koodistojen ylläpito jokaisen saatavuuspalvelun oman toteutuksen varassa.

Vertailu käytetyn ratkaisun ja toisen saatavuuspalvelun välillä

- Custodix InSite-palvelun integrointi etenee
- BCRQuest-palvelun käyttöönotto on vielä sopimusteknisiä yksityiskohtia vailla
- InSite ja RQuest –palveluiden vaatima lisäkapasiteetti on hankittu kertaluontoisesti Kliinisen tietopalvelun laskenta-alustalle sekä Medbitin konesaliin. (Jatkuvan ylläpidon sopimus solmitaan myöhemmin erikseen projektin loputtua, kun on tiedossa, mitkä ratkaisut VSSHJ jättää käyttöön)
- Edellisten lisäksi on tunnistettu muitakin ratkaisuja, kuten DataShield ja TriNetX. Nämä teknologiat otetaan mahdollisuuksien salliessa vertailuun mukaan (ilman paikallista implementointia)

Synteettisen datan generointiin ohjelmakirjasto

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. Datan avoimuudesta olisi paljon hyötyjä. Hyödyntäminen voitaisiin esimerkiksi joukkoistaa (crowd source) ongelmanratkaisutalkoiden muotoon (esim. Kuva 5, Junction Hackathon 2017).

HS Teknologia

Tänä viikonloppuna Otaniemessä 1 500 nuorta ratkaisee yritysten ongelmia ilmaiseksi – ”Tästä tulee unohtumaton kokemus”, sanoo kiinalainen Rex Huang

Euroopan suurinta ratkaisutalkootapahtumaa eli hackathonia vietetään Espoon Otaniemessä.



Rex Huang tuli Junctioniin Kiinasta asti. (KUVA: VIRVE RISSANEN / HS)



Tiedolla työskentelevä Tuukka Arola (vas.) perehtyy hackathon-kilpailija Ari Siitosen aineisto-ongelmaan. (KUVA: VIRVE RISSANEN / HS)



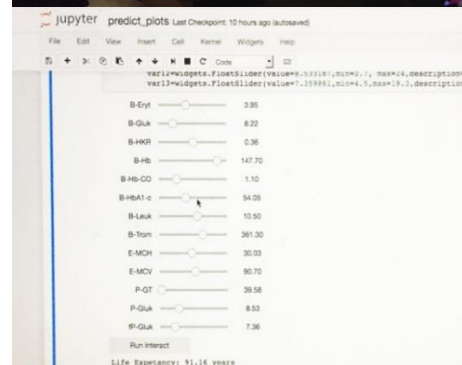
Virve Rissanen HS

HETI Otaniemen uudella metroasemalla vastaan tulee poikaporukka Junction-passit kaulassaan. Samaa väkeä pyörii Otaniemen Alepalla ostamassa eväitä.

”Slushissa puhutaan, Junctionissa tehdään”, kuvailee eräs tapahtuman osallistuja. Sekä Slushin että Junctionin taustalla on Startup-säätiö, josta HS kertoi tarkemmin keskiviikkona.

Junctionin erottaa muista hackathoneista lähinnä koko: Yrityksiä on mukana yli 60 ja osallistujia 1 500, keski-ikältään noin 24-vuotiaita. Yritysten haasteet on jaettu 13 eri teemaan.

Osallistujat muodostivat tiimensä perjantaina, ja sunnuntaina jokainen yritys palkitsee parhaan ratkaisun. Jaossa on myös 20 000 euron pääpalkinto.



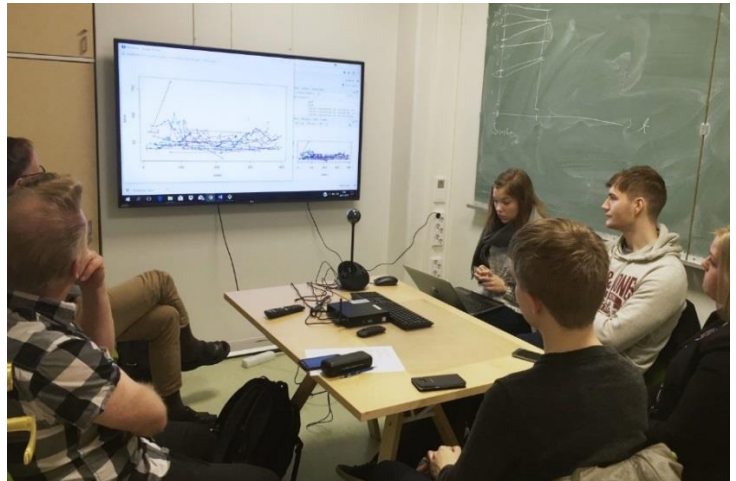
Kuva 1. Junction Hackathon ongelmanratkaisukilpailu Espoossa syksyllä 2017. Kilpailijat saivat käyttöönsä anonymisoidun diabeteeaineiston (N=21726 potilasta), jonka TYKSin Kliininen tietopalvelu toimitti järjestäjien käyttöön. Neuroverkkoa käyttävän ja veriarvoihin perustuvan elinaikamallin kehittänyt tiimi arvioitiin Sitran-isännöimän osakilpailun parhaaksi joukkueeksi (oikea sarake).

Synteettisen datan käyttöä testattiin ongelmanratkaisukilpailun lisäksi matematiikan perusopetuksessa. Kliinisen tietopalvelun tämänhetkinen etäkäyttö sisältää mahdollisuuden www-pohjaisen R Studio Server –ympäristön ajamiseen, jossa opiskelijat pääsevät harjoittelemaan R-kielisten algoritmien soveltamista ja ohjelmointia, paikallista asennusta omalla koneellaan (Kuva 2).

Kuva 4. Turun yliopiston sovelletun matematiikan projektikurssin ryhmä ja ohjaajia syksyllä 2017.

TY:n Sovelletun matematiikan opiskelijaprojektissa tutkittiin syksyllä 2017 mahdollisuuksia datan anonymisointiin ja testattiin erilaisia olemassa olevia menetelmiä.

Kuvassa näkyy www-selaimessa ajettava data-analyysiympäristö ja opiskelijoiden tekemä diabetespotilaisen hoitopolkugraafi. Yksilötason potilasdata sijaitsee TYKSin ympäristössä, eikä sitä siirrellä väliaikaisestikaan analysoitavaksi muualle.



Synteettisen datan ohjelmakirjasto on valmis ja saatavilla MIT-lisenssillä osoitteesta

<http://github.com/avirkki/synergetr>