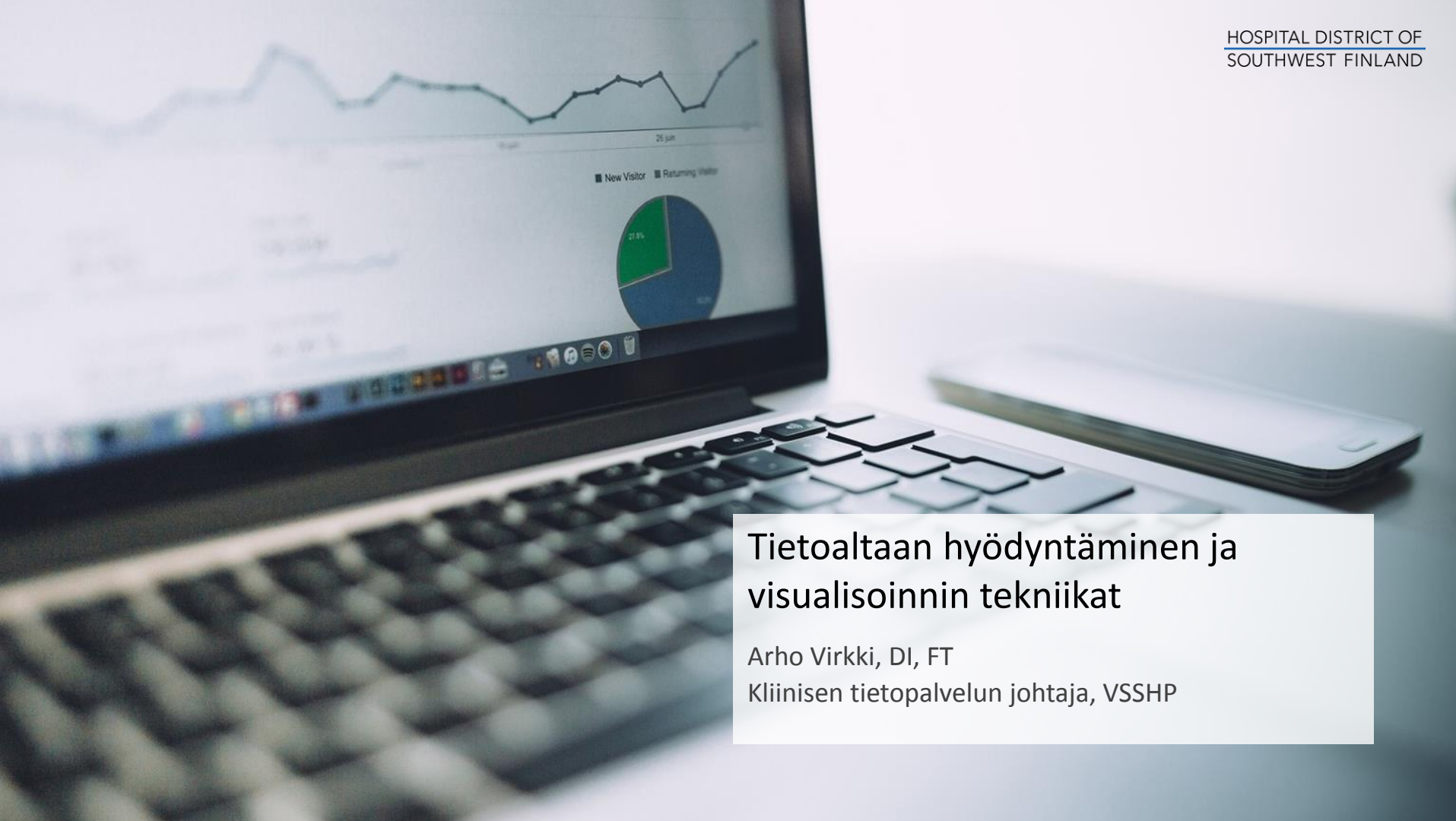




Tietoaltaan hyödyntäminen ja visualisoinnin tekniikat

Arho Virkki, DI, FT

Kliinisen tietopalvelun johtaja, VSSH

Kliininen tietopalvelu (KTP)

- VSSHP:n tutkimuspalveluiden yksikkö, perustettu vuonna 2014
- Kokoa potilastietoa operatiivisista järjestelmistä yhteen tietojärjestelmään
- Jalostaa tietoa yhdenmukaiseksi ja laadukkaaksi
- Kehittää analytiikka- ja visualisointityökaluja mm. hoidon vaikuttavuuden seurantaan
- Tuottaa potilasrekisteristä aineistoja ja analyyskejä tutkimuksiin ja laadun seurantahankkeisiin

Perinteinen potilaskertomus

- Juridinen asiakirja – dokumentoi potilaan ja hoitohenkilökunnan kontaktin
- 'Sähköistetty paperiversio' – suositetaan ihmissilmälle sopivia esitysmuotoja
- Potilaan terveystieto hajallaan
 - Sairaanhoidopiirin kymmenissä potilastietojärjestelmissä
 - Perusterveydenhuollossa ja yksityisten palveluntuottajien järjestelmissä
 - Kliinisten mittalaitteiden ja sensoreiden ohjelmistoissa
- Tietojärjestelmät tukevat pääasiassa yhden potilaan tietojen katselua kerrallaan
- Tieto vain osittain rakenteista, suurin osa saneluja
- Pääsy dataan vaikeaa – tekniset ratkaisut ja henkilötietosuoja

Kliinisen tietopalvelun tehtävä on rikastaa data ymmärrykseksi

Tietoaltaaseen kerätään raakadataa, jota
kliininen tietopalvelu jäsentää, järjestää ja
rikastaa.

Mitä datasta voidaan tutkia?

1. Annetut hoidot
2. Hoidon laatu
3. Hoidon vaikuttavuus
4. Kustannusrakenne
5. Tieteelliset kysymykset



Suomeksi

- Mitä tehtiin?
- Miten tehtiin?
- Auttoiko?
- Paljonko maksoi?
- Miksi näin?
- Miten jatkossa paremmin?

Miten tämä onnistuu?



Aineistot: Rakenteinen data

- Oberon – syntymä- ja kuolinajat, osastohoidot, avohoidot, lähetteet, toimenpiteet, diagnoosit
- Opera, Toti – leikkaukset, toimenpiteet ja diagnoosit
- Miranda/Desktop – sairauskertomusteksti, hoitotaulukot, lääkeosio, sähköiset reseptit
- Radu – kuvantamistutkimukset, toimenpidekoodit ja päivämäärät
- TYKSLAB – laboratoriotutkimukset, tulokset ja viitearvot
- Q-pati – esitiedot, patologin lausunnot ja patologian diagnoosit
- Kemokur – sytostaattihoitojärjestelmä
- Aria – sädehoidon annossuunnittelujärjestelmä (ei sis. brakihoitoja)
- WebMarela – apteekin toimittamat iv-sytostaatit
- Lukuisa joukko muita pienempiä tietokantoja, Ipana, BCB Medicalin laatu järjestelmät (tulossa),...

```

<Diagnoses>
<DiagnoseRow>
<RowNumber>1</RowNumber>
<Organ>LIQUOR</Organ>
<T-Snomed>TX1000</T-Snomed>
<Diagnose>MORPHOLOGICAL DESCRIPTION ONLY</Diagnose>
<M-Snomed>M09350</M-Snomed>
<NoteDiagnose>X</NoteDiagnose>

```

outsour text	weekday text	custorg0 text	custorg1 text	custorg2 text	custorg3 text	custorgt text	projid text	payer1 text	payer2 text	provorg1 text	labspecd text	labspec text	provorg2 text	testorg text	datetime1 timesta...
0	0	KO-HTR	KO-SIS	017	017	V	0	0		57	10	33	57	2017	
0	1	SSSova	STK	STK734	STK734	A	0	0		4463	99A	3	4463		2006-01-...
0	3	TYKS	TYKSKI	235	235	V	0	0		3827	20	2	3827	3131/01	2008-10-...
0	3	TKU	TYÖT	TYÖT	TTHKAU	A	0	0		90	99B	2	90	2136	2008-10-...
0	3	TYKS	TYKS02	022	022	P	0	0		93	10	2	93		2008-10-...
0	3	TKU	TYÖT	TYÖT	TTHKAS	A	0	0		90	99B	1	90	3132	2008-10-...
0	3	TKU	VARIS	VARIS	TAVS1	A	0	0	3793	90	99A	2	90	3133	2008-10-...
0	1	UKITK	UKITK	UKITK	U/SOFI	V	0	0		4462	99L	2	4462		2009-02-...
0	4	STK	STK	STKSAL	STK734	A	0	0		4463	99A	2	4463	2184	2012-09-...
0	4	AL-ESH	AL-LAS	LASKIR	L/MOTO	V		0		4460	20	2	4460		2010-03-...
0	4	LOITK	LOITK	LOITK	LOPTVO	A	0	0		4460	99A	2	4460	2183	2010-03-...
0	3	STK	STK	STKSAL	STK734	A	0	0		4463	99A	1	4463		2008-09-...
0	1	TYKS	KANTA	014	014	V	0	0		93	10	2	93		2006-01-...

Aineistot: Signaalidata (etä- ja osastoseuranta)

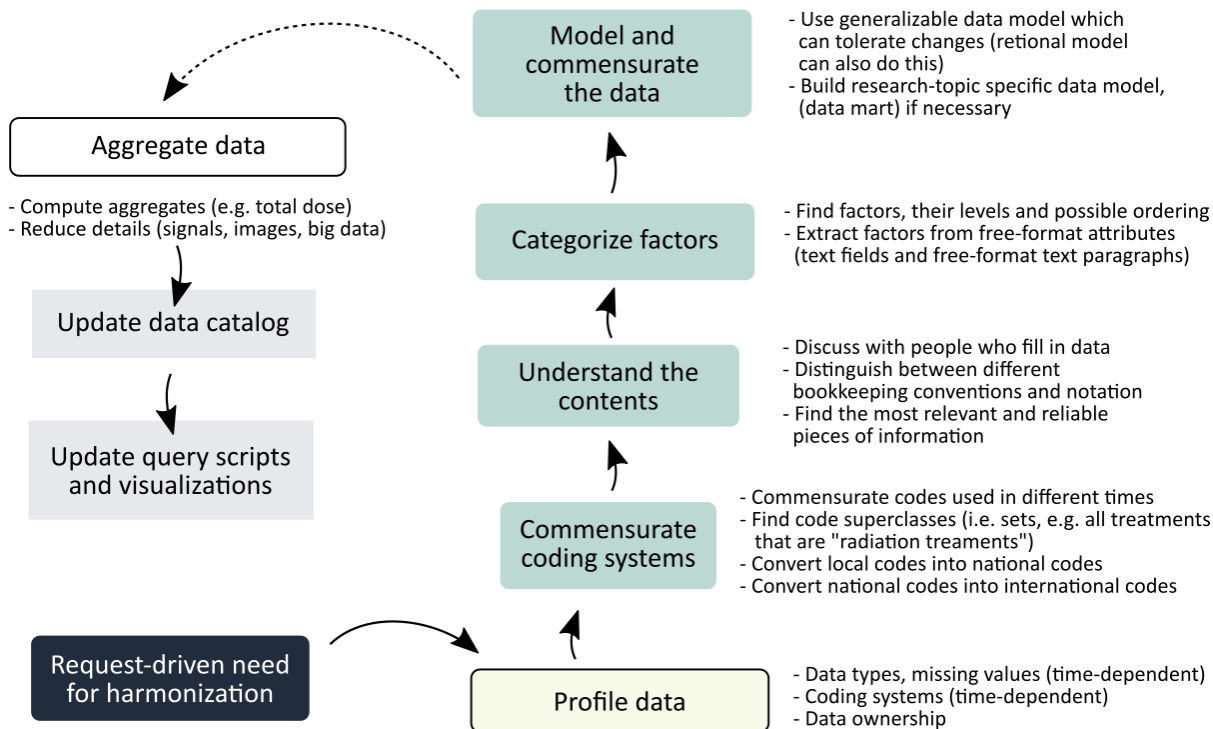


Aineistot: Kuvantamisdata

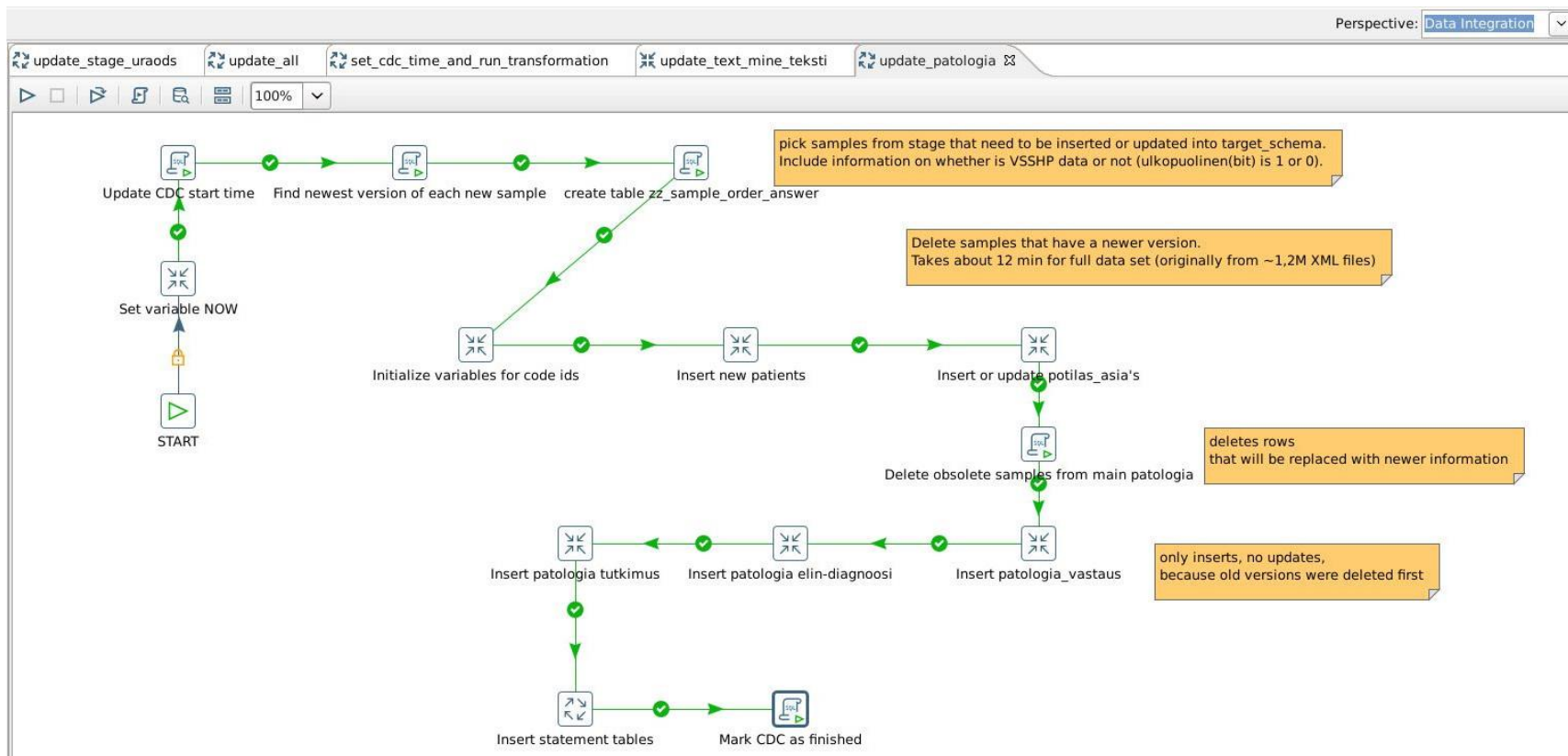


Datan yhteismitallistus

Data harmonization Process



Datan yhteismitallistus käytännössä



Datan jalostus

The screenshot shows the RStudio IDE with a script named `postgres_connection.R`. The script connects to a PostgreSQL database named `research` on `localhost` using the `RPostgreSQL` package. It reads a table named `asiakas` and a subset of data from `aria_sadehoito` (limited to 100 rows). The Environment pane shows the `con` object as a `PostgreSQLConnection` and the `drv` object as a `PostgreSQLDriver`. The Console shows the script being sourced. The bottom pane shows a terminal window with a command to connect to a remote server via SSH.

```
1 library("RPostgreSQL")
2
3
4 if( !exists("con") ) {
5   drv <- dbDriver("PostgreSQL")
6   con <- dbConnect(drv, host="localhost", dbname="research",
7                     user="<username>", password="<password>")
8 }
9
10
11 asiakas <- dbReadTable(con, "asiakas")
12 sadesample <- dbGetQuery(con, "SELECT * FROM aria_sadehoito LIMIT 100")
13
14
```

Console:

```
> source("/opt/ktp/examples/postgres_connection.R")
> |
```

Terminal:

```
arho@DESKTOP-M5E1QKK: ~$ ssh -L 8787:localhost:8787 ktp@192.168.122.100 -o ProxyCommand="ssh bastion@analytics.tyks.fi -W %h:%p -i /mnt/c/Users/arhov/ssh/KTP_Key"
```

POSTATA
POSTATAE
PROS
PROSPESÄ
PROSR
PROST
PROSTA
PROSTAE
PROSTAH
PROSTAPESÄ
PROSTAPVMAT
PROSTARA
PROSTAT
PROSTATE
PROSTATH
PROSTH
PROSTIMRT
PROSTP
PROSTPE
PROSTPES
PROSTPESA
PROSTPESÄ
PROSTPESÄVMAT
PROSTPESÄWMAT
PROSTATAT
PROSTUUSI
PROSTVMAT
PROT
PROTATA
PROTATAE
RPOSTATA
SIMRTPROSTP
SPROPESÄ
SPROST
SPROSTPVMAT
SPROSTVMAT

Tiedon laadunvarmistus ja oikeellisuus



Tietolähdekohtainen oikeellisuuden tarkistaminen

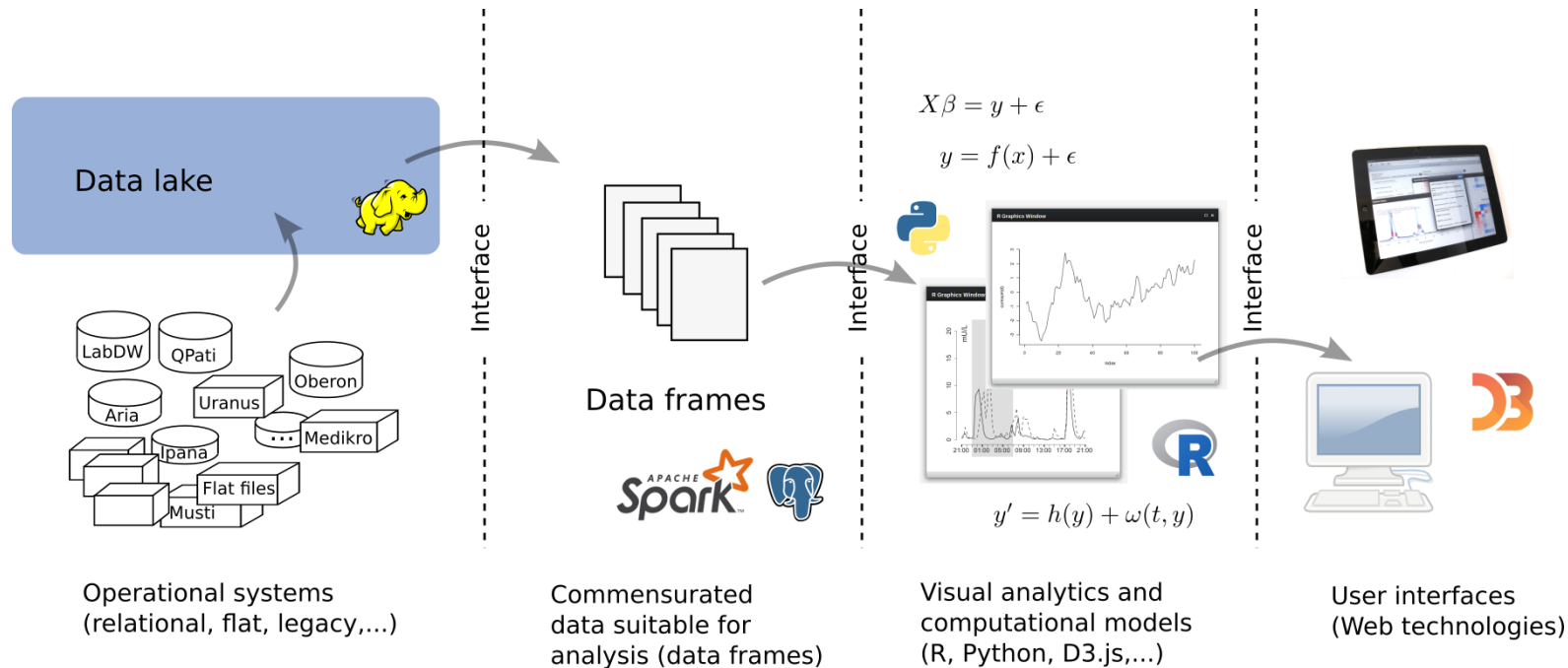
- Päivittyykö data rajapinnan kautta niin kuin pitäisi?
- Vastaako data siitä annettua kuvausta?
- Onko datassa virheitä?

Useiden lähteiden yhdistäminen varmentaa tietoa

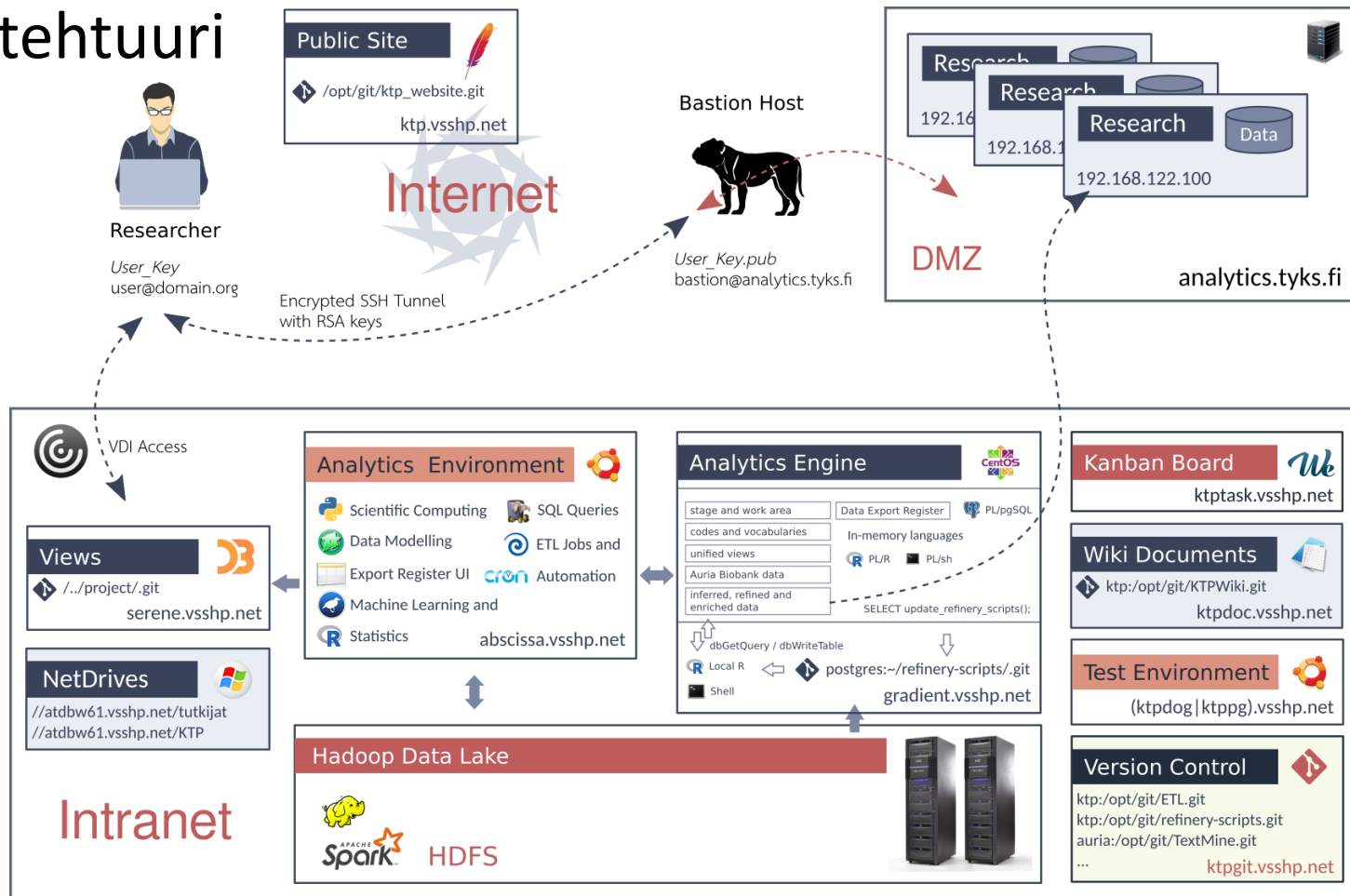
- esim. syöpä kliinisenä diagnoosikirjauksena varmentuu patologiasta, toimenpidetiedoista, kuvantamisesta ja labrasta

Tutkimuksissa käytetään usein rakenteisen datan lisäksi myös suoraa potilaskertomustekstiä tarkentamaan ja varmentamaan

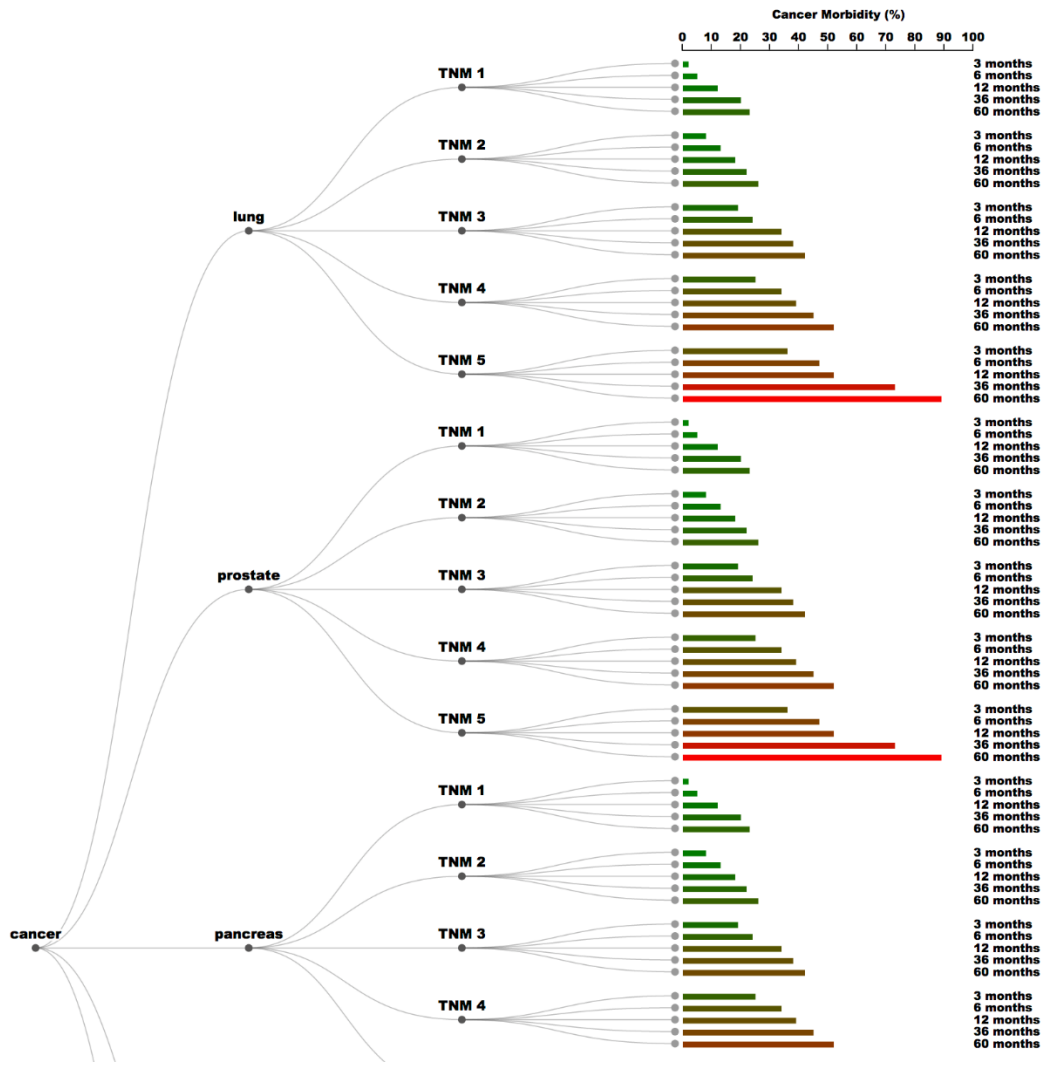
Tiedon polku altaasta näkymäksi



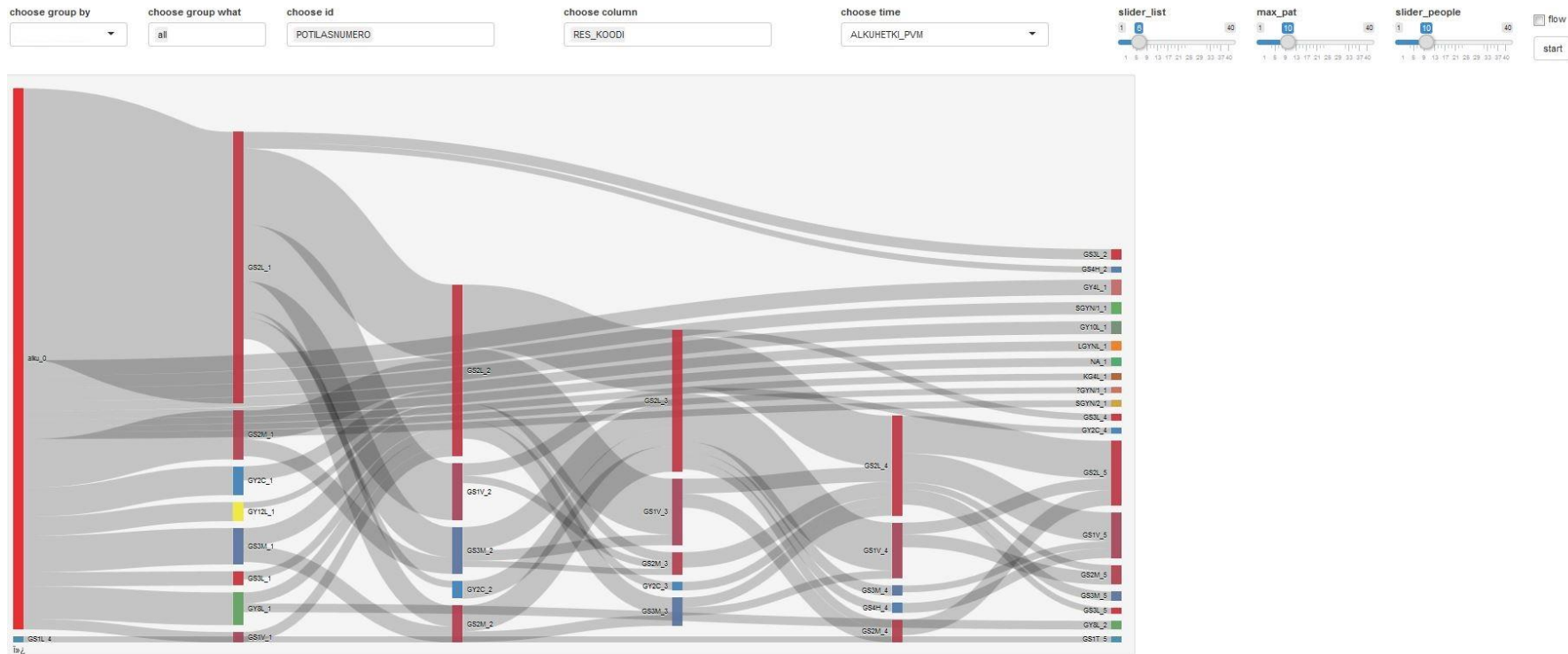
IT-arkkitehtuuri



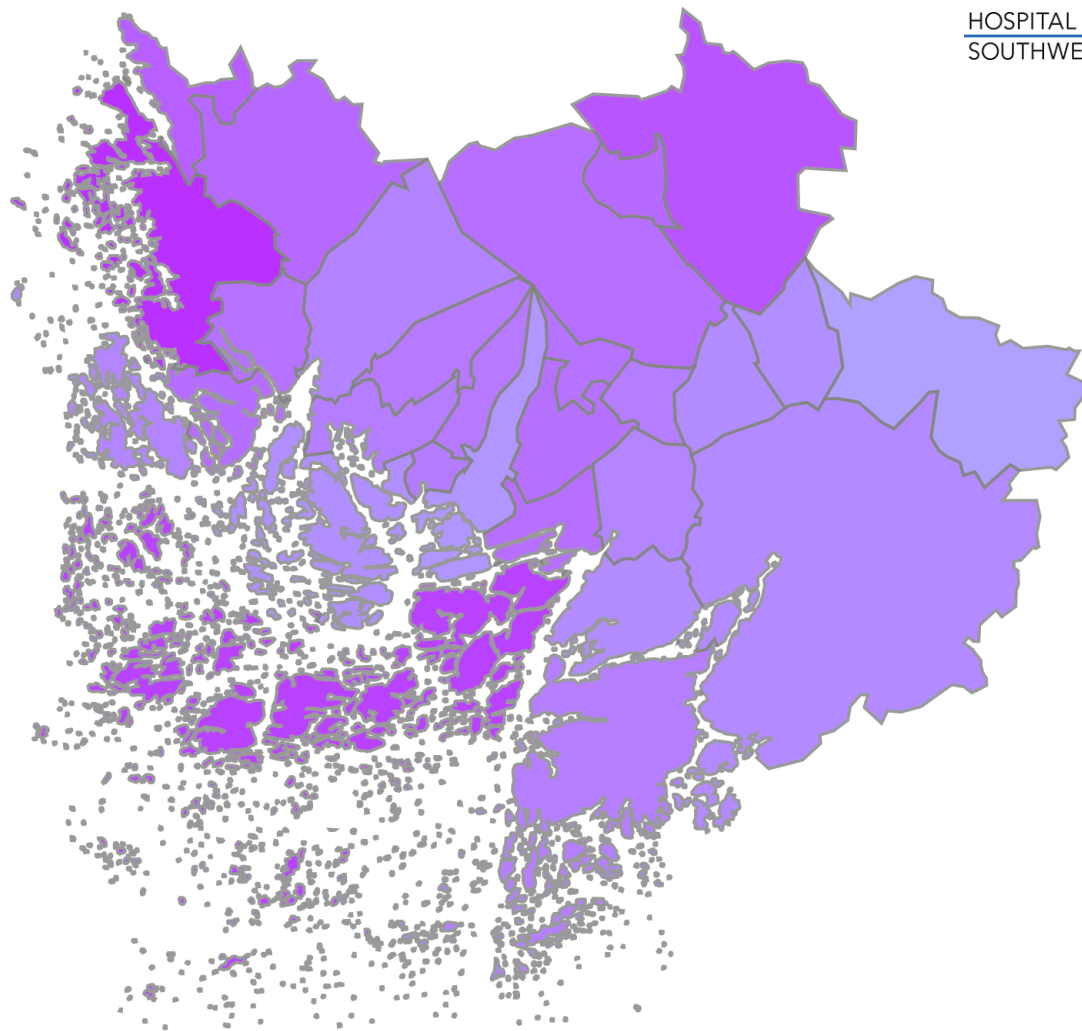
Näkymä: Potilasvirta- analyysi

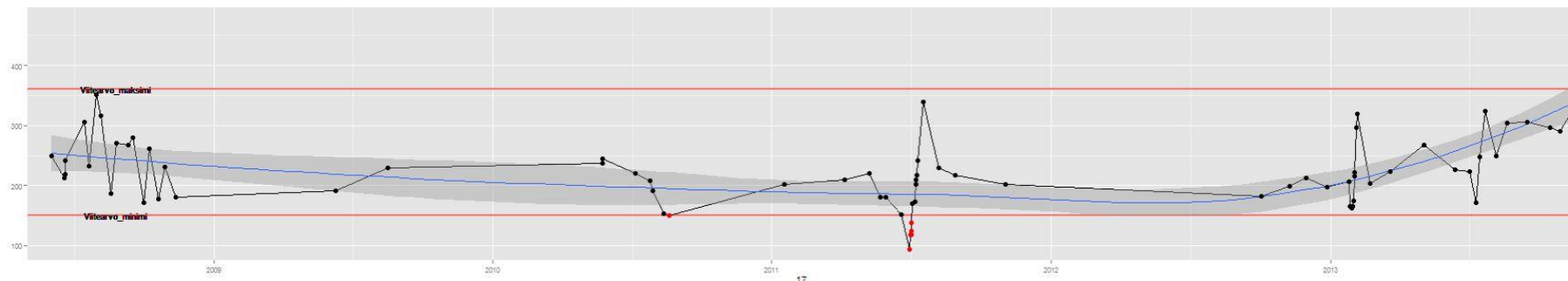
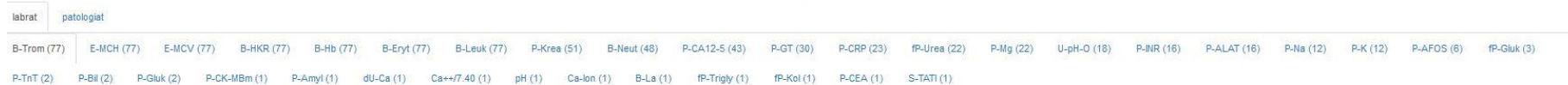


Näkymä: Potilasvirta-analyysi 2

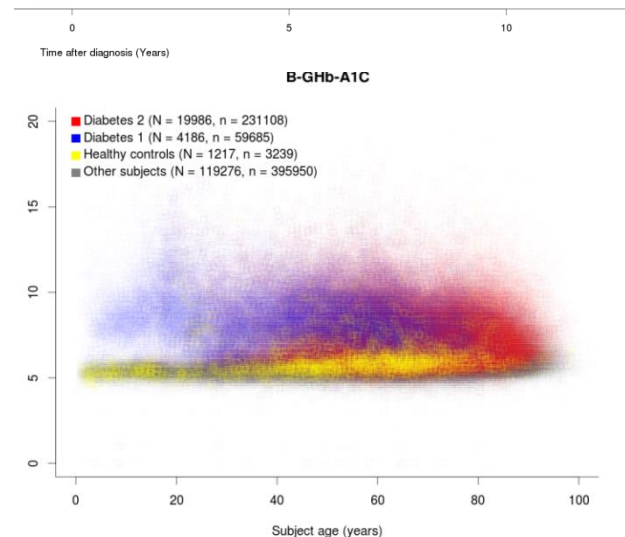
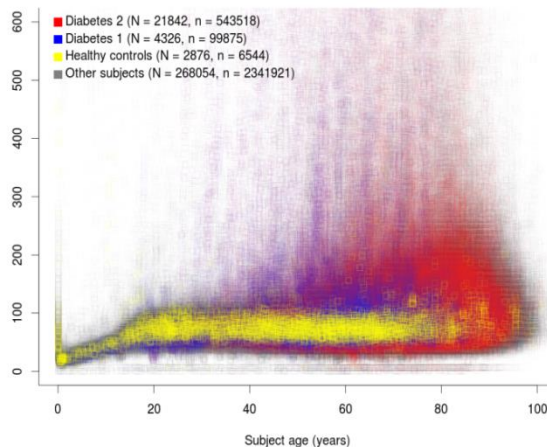
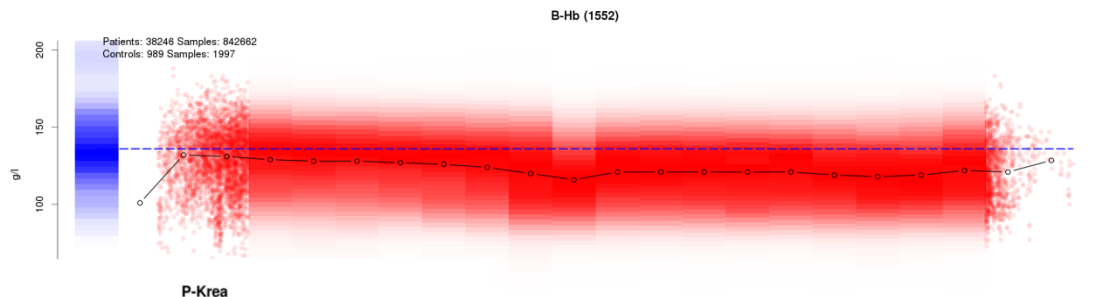
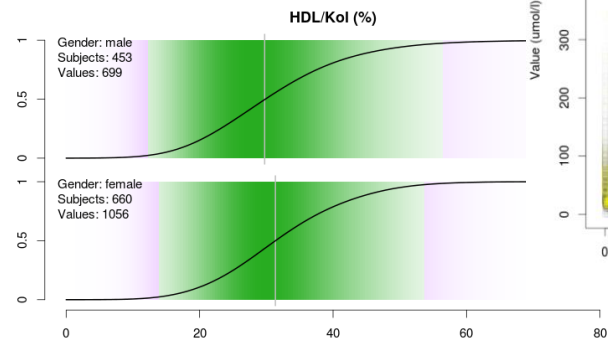
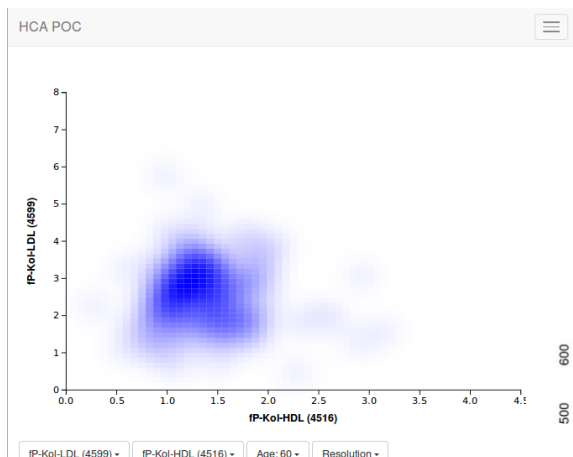


Näkymä: Maantieteelliset jakaumat





Näkymä: Populaatiotason sirontakuviot



Tiedon syöttö: Palautejärjestelmät ja kyselytutkimukset

Uro Forms

127.0.0.1:9292

EPIC-26

The Expanded Prostate Cancer Index Composite (EPIC)

Tämä kysely on tehty mittaamaan eturauhassyöpää sairastavien elämänlaatua. Oikean tuloksen saamiseksi on tärkeää, että vastaat kaikkiin kysymyksiin rehellisesti etkä jätä mitään kohtaa väliin.

Muista että vastauksesi tähän kyselyyn, kuten kaikki potilastiedot, ovat täysin luottamuksellisia.

Etunimi

Etunimi

Sukunimi

Sukunimi

Henkilötunnus

Henkilötunnus

Päivänmäärä

Päivänmäärä

1. Kuinka usein sinulla on ollut virtsankarkailua viimeisten 4 viikon aikana?

☒ Useammin kuin kerran vuorokaudessa

1

☐ Noin kerran päivässä

2

☐ Useammin kuin kerran viikossa

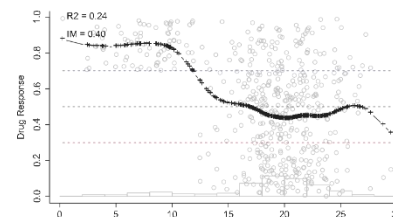
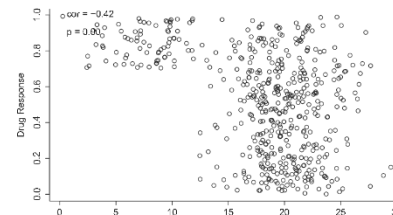
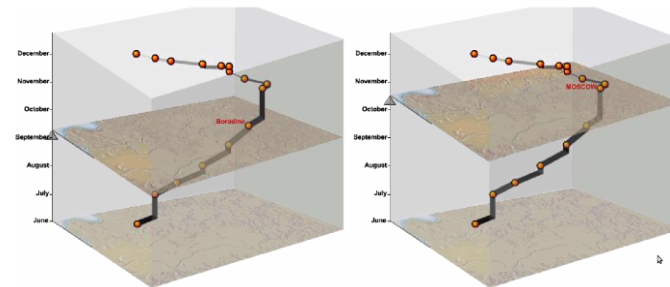
3

Näkymän rajoittaminen: Pseudonymisointi

- Ei sama kuin anonymisointi; data edelleen yksilöityä henkilötietoa
- Potilaan yksilöivät tiedot korvataan koodein, poistetaan tai luokitellaan
 - hetu, nimi, osoite, yms.
 - pieni kotikunta yhdistettynä muuhun tietoon yksilöi helposti
 - tarkat päivämäärät pois, esitetään esim. potilaan ikäluokka tapahtumahetkellä
- Samaa hetua vastaava pseudonyymi on aina eri tutkimuksissa erilainen
 - kullekin tutkimukselle on oma salt/suola, jonka perusteella funktio luo uniikin pseudonyymin jokaiselle hetulle
 - Hetu1 → Pseudonyymi1
 - Hetu1 → Pseudonyymi2
- Vapaasta tekstistä tunnisteet hyvin vaikea poistaa; teksti louhitaan ja rakenteistettu tulos luovutetaan tutkimuskäyttöön

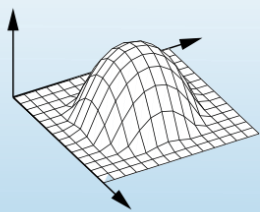
Data-analyysin kolme suuntaa

1. Visuaalinen analytiikka
2. Tilastollinen mallinnus
3. Mekanistinen (ilmiöihin perustuva) mallinnus

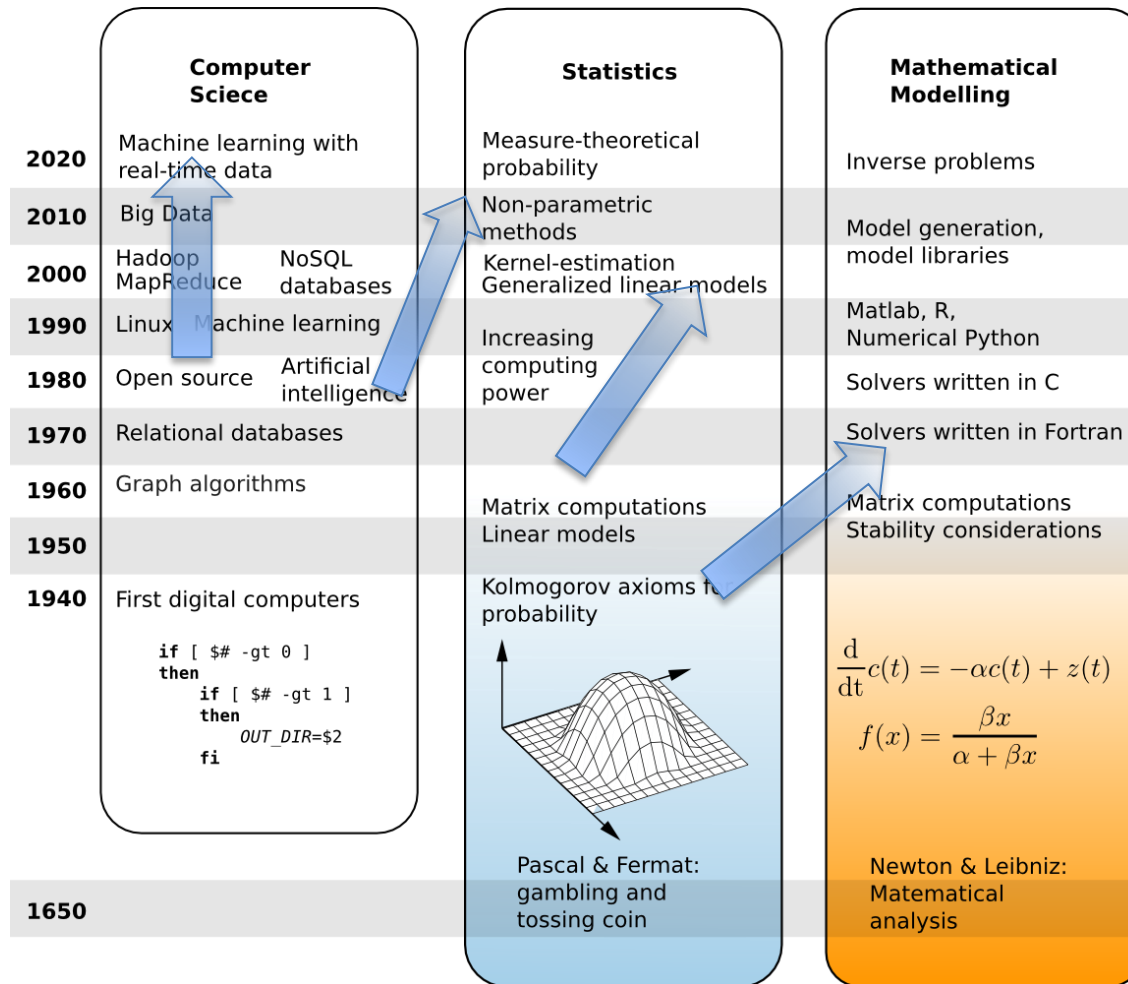


$$\frac{d}{dt}c(t) = -\alpha c(t) + z(t)$$

Mitä on Datatiede?

	Computer Science	Statistics	Mathematical Modelling
2020	Machine learning with real-time data	Measure-theoretical probability	Inverse problems
2010	Big Data	Non-parametric methods	Model generation, model libraries
2000	Hadoop NoSQL MapReduce databases	Kernel-estimation Generalized linear models	Matlab, R, Numerical Python
1990	Linux Machine learning	Increasing computing power	Solvers written in C
1980	Open source Artificial intelligence		Solvers written in Fortran
1970	Relational databases		
1960	Graph algorithms	Matrix computations Linear models	Matrix computations Stability considerations
1950			
1940	First digital computers <pre>if [\$# -gt 0] then if [\$# -gt 1] then OUT_DIR=\$2 fi</pre>	Kolmogorov axioms for probability 	$\frac{d}{dt}c(t) = -\alpha c(t) + z(t)$ $f(x) = \frac{\beta x}{\alpha + \beta x}$
1650		Pascal & Fermat: gambling and tossing coin	Newton & Leibniz: Mathematical analysis

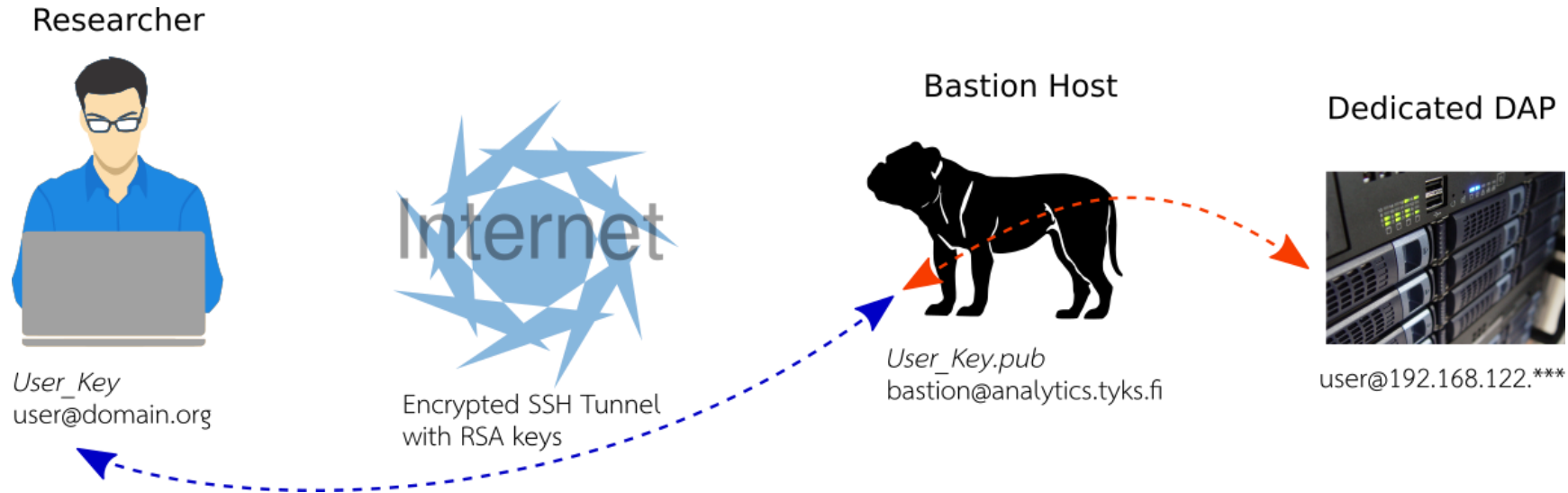
Mihin olemme menossa?



Tutkimuksen kansalliset tuulet

- Pyrkimystä terveystiedon hyödyntämiseen kansallisella tasolla
- Sähköisen potilaskertomuksen rakenteistaminen ja harmonisointi eri terveydenhuollon toimintayksiköiden välillä tärkeää
- E-Arkisto ei vielä hyödynnettävissä tutkimukseen, mutta helpottaa jo tiedon siirtymistä palveluntuottajien välillä
- Sitran Isaacus-palveluoperaattori –hanke (2015-2017)
 - pyrkii tarjoamaan ”yhden luukun” terveystietoon eri palveluntuottajilta ympäri Suomen
 - pyrkii yksinkertaistamaan ja nopeuttamaan lupakäytäntöjä
- Suomen biopankkitoiminta edistyksellistä ja kilpailuvaltti maailmalla
- Terveydenhuollon tietoon halutaan integroida muutakin tietoa laajemman näkökulman saamiseksi

Tutkijan Allas: Tutkimusyhteistyö



Ympäristön avaaminen

 Clone virtual machine

Create clone based on: xubuntu16_04_template

Name:

Networking: NAT (52:54:00:99:66:ad) Details...

Storage:  xubuntu16_04_template.qcow2

 - (Removable, Read Only)

Nothing to clone.

T287_runner Virtual Machine

File Virtual Machine View Send Key

- Overview
- Performance
- Processor
- Memory**
- Boot Options
- VirtIO

Memory

Total host memory: 257674 MiB

Current allocation: - MiB

Maximum allocation: - MiB

T287_runner Virtual Machine

File Virtual Machine View Send Key

- Overview
- Performance
- Processor**
- Memory
- Boot Options
- VirtIO Disk 1
- IDE CDROM 1
- NIC :1e:93:da
- Mouse
- Keyboard
- Display Spice
- Sound: ich6

CPUs

Logical host CPUs: 56

Current allocation: -

Maximum allocation: -

Configuration

☒ Copy host CPU configuration

▸ Topology

▸ Pinning

Etätyöpöydän käyttö

Session preferences - Top Research

Session Connection Input/Output Media Shared folders

Session name: Top Research

 << change icon

Path: / ...

Server

Host: 192.168.122.zzz <the machine address>

Login: <account name>

SSH port: 22

Use RSA/DSA key for ssh connection: ...

☐ Try auto login (via SSH Agent or default SSH key)
☐ Kerberos 5 (GSSAPI) authentication
☐ Delegation of GSSAPI credentials to the server
☒ Use Proxy server for SSH connection

Proxy server

Type: ☒ SSH ☐ HTTP

Same login as on X2Go Server ☐ Login: bastion

Same password as on X2Go Server ☐

Host: analytics.tyks.fi RSA/DSA key: <path to private key> ...

Port: 22

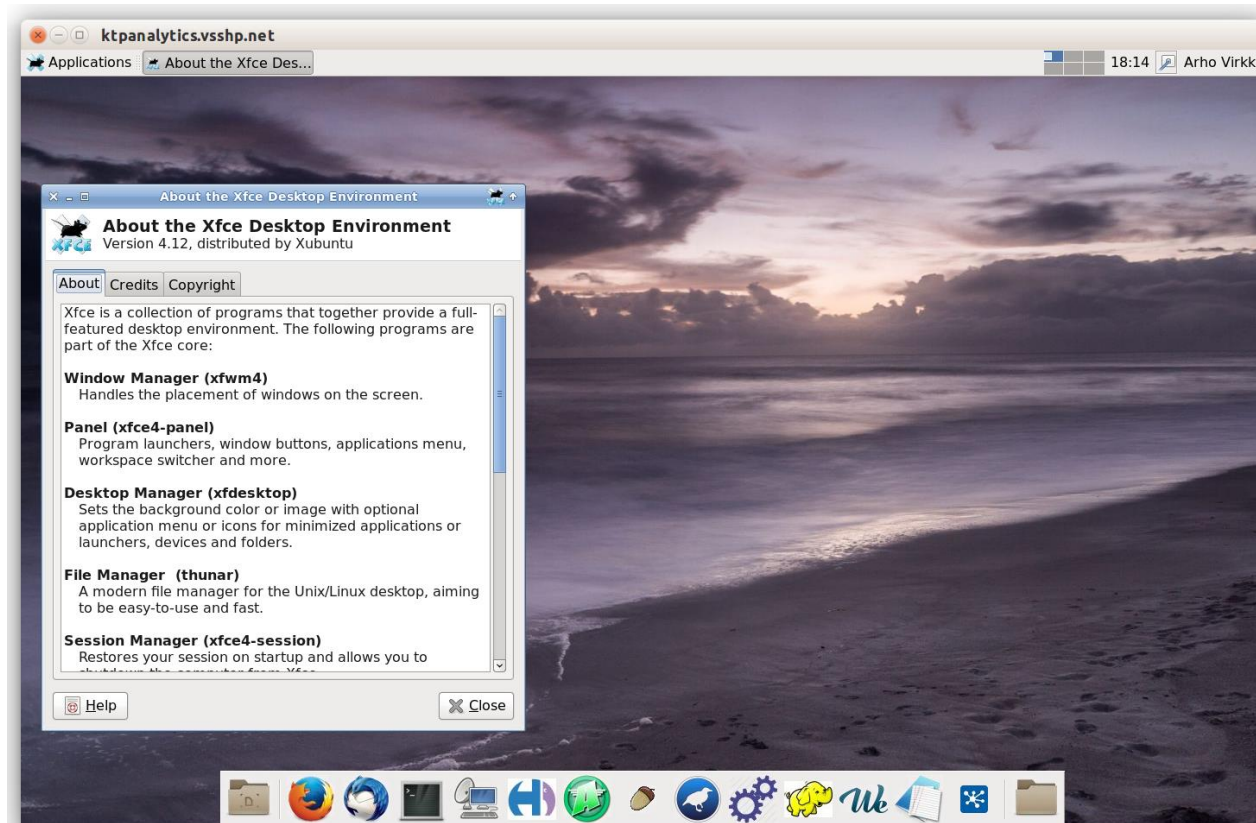
SSH Agent or default SSH key ☐
Kerberos 5 (GSSAPI) authentication ☐

Session type

XFCE Command: ...

OK Cancel Defaults

Etätyöpöydän käyttö



Etäresurssit

The screenshot displays an RStudio interface with the following components:

- Source Editor:** Contains the script `posgres_connection.R` with the following code:

```
1 library("RPostgreSQL")
2
3
4 if( !exists("con") ) {
5   drv <- dbDriver("PostgreSQL")
6   con <- dbConnect(drv, host="localhost", dbname="research",
7                     user="<username>", password="<password>")
8 }
9
10
11 asiakas <- dbReadTable(con, "asiakas")
12 sadesample <- dbGetQuery(con, "SELECT * FROM aria_sadehoito LIMIT 100")
13
14
```
- Environment:** Shows the global environment with two objects:
 - `asiakas`: 8493 obs. of 4 variables
 - `sadesample`: 100 obs. of 25 variables
- Values:** Shows the classes of the objects:
 - `con`: Formal class PostgreSQLConnection
 - `drv`: Formal class PostgreSQLDriver
- Console:** Shows the execution of the script:

```
> source("/opt/ktp/examples/posgres_connection.R")
>
```
- Terminal:** Shows the command to connect to the RStudio server via SSH:

```
arho@DESKTOP-M5E1QKK:~$ ssh -L 8787:localhost:8787 ktp@192.168.122.100 -o ProxyCommand="ssh bastion@analytics.tyks.fi -W %h:%p -i /mnt/c/Users/arhov/ssh/KTP_Key"
```
- Taskbar:** Shows the Windows taskbar with the Microsoft Bash Launcher icon highlighted.

Käyttösitoumus

Terms and Conditions for Using Centre of Clinical Informatics (CCI) Data Analysis Platform (DAP) at the Hospital District of Southwest Finland

Data Security

User names, passwords, and private keys for the asymmetric cryptography remote access are personal. You may not give your credentials to any other person (not even to the personnel at your research team) intentionally or unintentionally, or leave your computer unlocked while being connected to DAP. All traffic to DAP will be logged.

Even though the data is pseudonymized, it must be kept secret both during and after the end of research projects. In particular, it is forbidden to download unaggregated data to your personal computer or any media which resides outside the DAP. The data may be combined with other sensitive sources only with a prior permission from all the respective registrars or from National Institute for Health and Welfare (Terveyden ja hyvinvoinnin laitos, THL). In such cases, you will use remote procedure calls to minimize data transfer between different computing environments and consult CCI personnel to validate your algorithm for security before running your analyses.

Costs and Software Licenses

CCI will charge a calendar month-based fee for using the environment depending on the computational needs. The fee is based on the amount of central memory, disk space, number of CPU cores allocated and the number of users in the research team. The DAP is based on free and open sourced software, the usage of which incurs no separate fee.

The users are free to install additional open source and commercial software packages to their instance of DAP at their own expense. The network, storage or computational facilities are intended only for exploring the research data, and must not be used to run other resource consuming tasks or host unrelated services.

System Integrity and Security

The CCI takes no responsibility for damage due to technical or software errors or inappropriate use of the resources. CCI backs up the entire raw data and can make occasional backup copies of files, but the backup system does not cover all files or all versions of them. The general responsibility for making backups of scripts and using a version control system lies with the users. If the user discovers security lapses these should be reported to CCI. Such information may not be spread. Looking for security weaknesses is not prohibited, but it must aim at informing CCI and improving the data security of DAP. In all cases, all users will follow good research practices and the Finnish law.

I agree to the terms and conditions listed above.

Date: _____

Signature: _____

Clarification of signature: _____

After signing the form, please scan and send it to ktp@tyks.fi, or deliver in person at the CCI office at T. hospital, 6.th floor, office E664, Turku.

Palveluoperaattoridata

Palveluoperaattorirajapintaan mahdollisimman pitkälle harmonisoitua dataa (pääosa harmonisoinnista lähellä lähteitä, missä osaaminen ja ymmärryskin on)

- eri palveluntuottajien datassa käsitelmallit eroavat
- samojen lähdejärjestelmien data voi olla eri näköistä eri järjestelmäinstansseissa

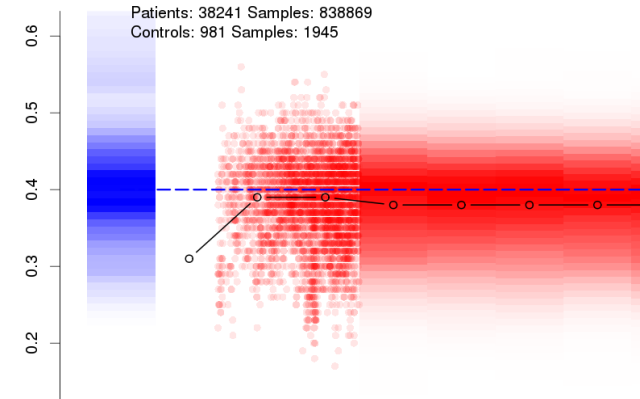
Harmonisointiin kansalliset ohjeet datan tuottajille

Kansalliset koodistot keskiössä

- erikoissalat, toimenpideluokitus, laboratoriotutkimusnimikkeistö
- mahdollisuuksien mukaan kansainvälisetkin koodistot (esim. Loinc, Snomed CT)

VSSH:n EHR4CR-hanke pilotoi ideaa

- saatavuustietokanta – potilasdatan ”näyteikkuna” kansainvälisille lääkeyrityksille
- helppo tapa arvioida datan saatavuutta
- koodistojen kääntäminen kansainvälisiksi
- ensimmäinen demo VSSH-datalla huomenna (Custodix)



Yhteystiedot:

Klininen tietopalvelu

ktp@tyks.fi