

TIEDON LAATU DATAN HALLINNASSA – HUS TIETOHALLINTO

09.12.2022 Timo Hakala, VTL, tietoarkkitehti

JOHDANTO – KÄSITTEET JA TAVOITTEET

TIETOJOHTAMISEN KESKEISET KÄSITTEET DATAN LAADUNHALLINNAN PERUSTANA

Tietojohtamisella tarkoitetaan yleisesti organisaation kykyä luoda arvoa tiedolla ja osaamisella. Tietojohtaminen jaotellaan usein käsitteellisesti tiedolla johtamiseen ja tiedon johtamiseen.

Tiedolla johtamisessa on kyse tietoperusteisesta päätöksenteosta ja sen edistämisestä. Tiedolla johtamisen edellytyksenä on laadukas tieto ja sen ympärille rakentuvat päätöksentekomekanismit. Organisaation tiedolla johtamisen kulttuuri on ratkaisevaa tiedolla johtamisen onnistumisessa.

Tiedon johtamisessa edistetään tiedon hallinnan, tietoarkkitehtuurin ja teknologioiden sekä datan laadunhallinnan viitekehyksien avulla tiedon hyödyntämisen edellytyksiä. Tiedon johtamisella kehitetään ja ylläpidetään mm. datan laatua osana tietotuotantoa. Keskeinen osa tätä työtä on raportoida datan laadusta.

Datan laadun näkökulmat ovat datan tekninen, käsitteellinen ja käytännöllinen laatu. Näiden näkökulmien huomioiminen varmistaa datan korkean laadun.

DATAN LAADUN HALLINNAN YLEISET TAVOITTEET

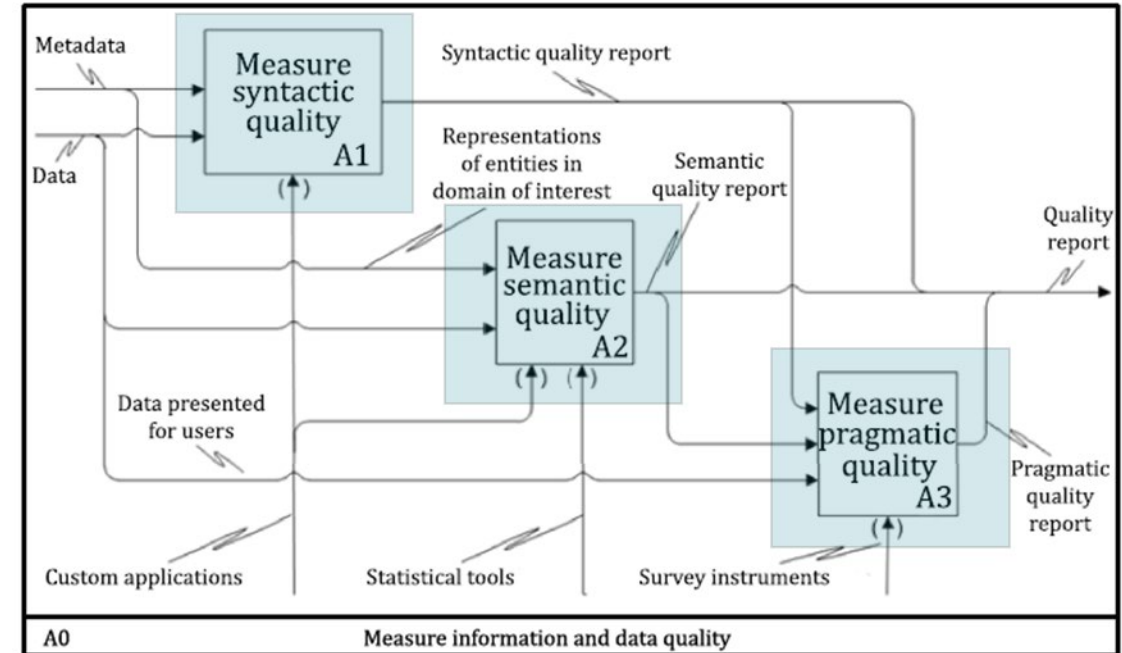
- 1. Tavoitteena** on luoda HUS Tietohallintoon kansainväliseen ISO-standardiin perustuva, geneerisesti sovellettavissa oleva datan laadun hallinnan viitekehys ja toimintamalli. Mallin avulla varmistetaan ja osoitetaan datan luotettavuus tietotuotannossa.
- 2. Tavoitteen toteuttaminen** edellyttää systemaattista datan laadun arviointiprosessia.
- 3. Datan laadun kehittämisprosessissa** huomioidaan mm. johtamiseen, rooleihin, datan laatukulttuuriin, datan laadun mittaamiseen, prosesseihin ja raportointiin liittyvät tekijät.

DATAN LAADUN KEHITTÄMISPROSESSI ISO 8000-8 MUKAAN

Datan laadun kehittäminen kohdistuu seuraaviin asiakokonaisuuksiin:

1. Johtaminen, organisaatio ja roolit
2. Ihmiset ja datan laatukulttuuri
3. Datan laadun mittaaminen, mittarit ja mittaustyökalut
4. Prosessit ja prosessien tehokkuus
5. Riskienhallintaprosessit
6. Raportointi

Kuhunkin yllä mainittuun asiakokonaisuuteen liittyy joukko toimenpiteitä. ISO 8000-8 standardissa laadun kehittämisen toimenpiteet ovat joko syntaktisia, semanttisia tai pragmaattisia.



TIETOTUOTANNON YLEISET PÄÄVAIHEET JA ISO 8000-8

Tietotuotannon päävaiheet:

Integraatiot - Tietoallas – Tietovarasto – Raportointi

ISO 8000-8 työssä on keskitytty

- tekniseen datan laadun hallintaan (kaikki data on saapunut tietoalalle, ei esim. kirjausten laatuun)
- käsitteelliseen datan laadun hallintaan (käsitteiden yhdenmukaisuus)
- pragmaattiseen datan laadun hallintaan (datan käytännöllinen sovellettavuus)
- datan laadun raportointiin

DATAN LAATUUN
VAIKUTTAVIA TAVOITTEITA
JA TREND EJÄ

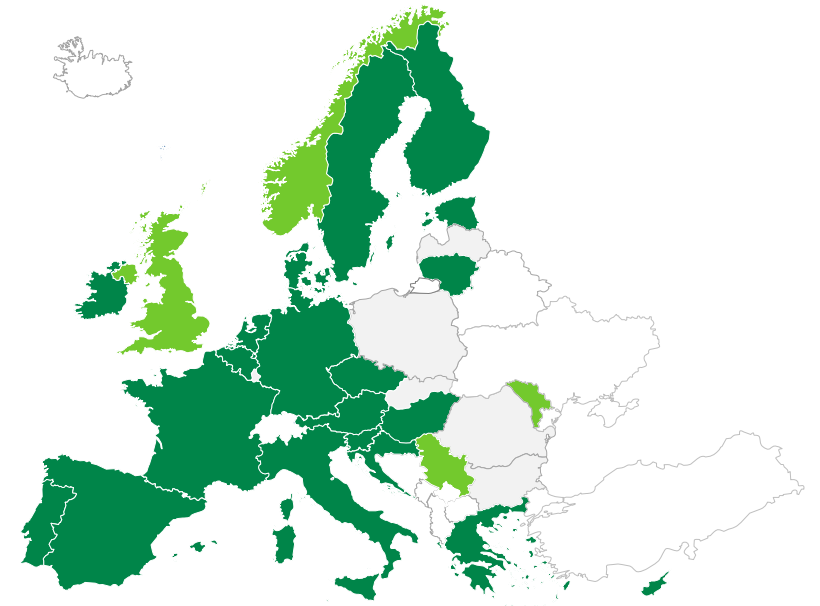
DATAAN LIITTYVIÄ TRENDEJÄ

1. Regulaatio lisääntyy tiedon saatavuuteen esimerkiksi:

- Yleinen tietosuoja-asetus 2016/679
- HILMO/AVOHILMO
- KANTA
- Erialaisten SOTE-tilannekuvien tietotuotanto
- Valmiuskeskusten tiedonsaanti
- Findata

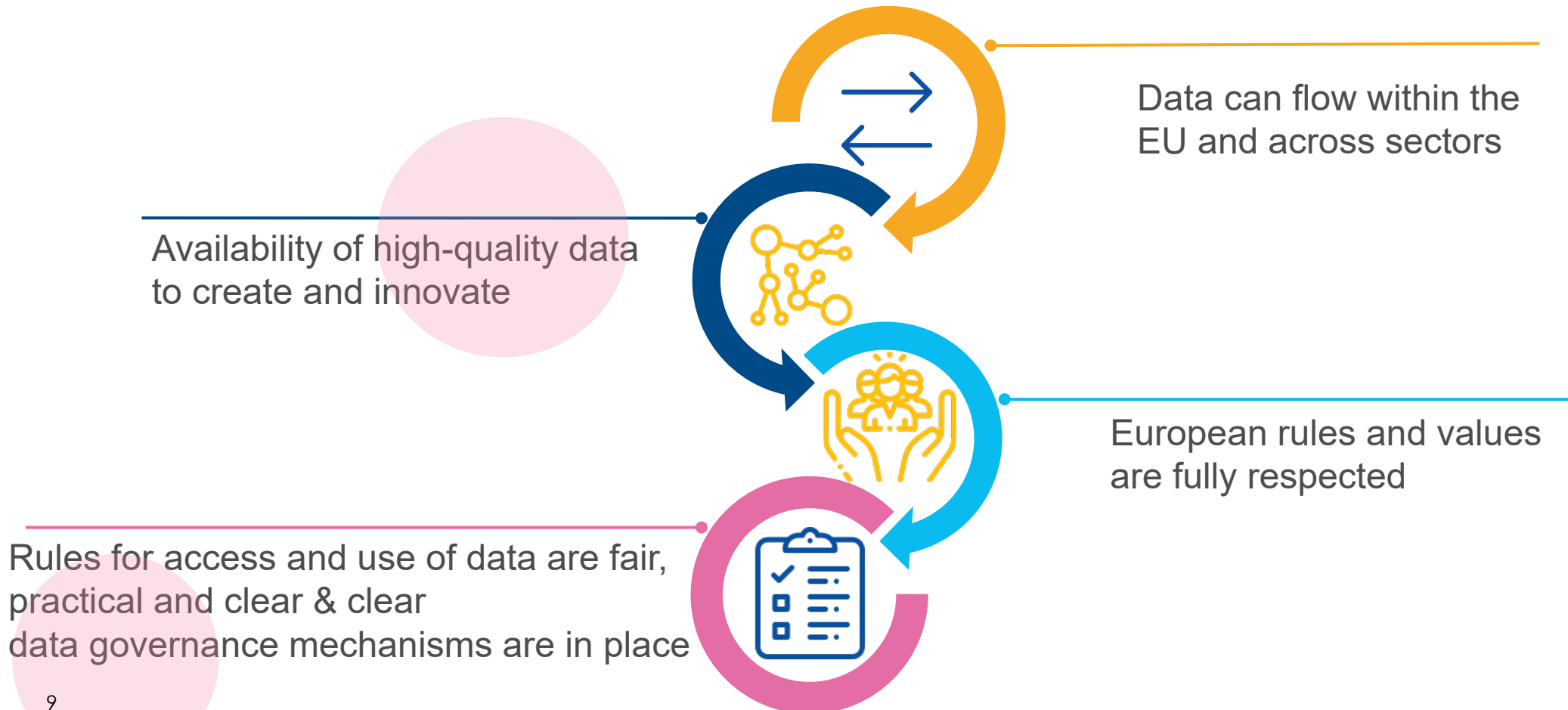
2. Regulaatio lisääntyy tiedon laatuun liittyen esimerkiksi:

- Kansallinen SOTE-tiedon keruu
- Eurooppalainen terveysdata-avaruus (COM(2022) 197/2)
- Tekoälyn opettaminen
- Automaattinen päätöksenteko (esim. EU:n tietosuojaryhmä WP251rev.01)
 - HE 145/2022: 53 e §. Asian automaattisen ratkaisemisen edellytykset
- "Data markkinat" – European Strategy for Data



EUROPEAN STRATEGY FOR DATA (2020)

A common European data space, a single market for data



European Health Data Space (EHDS)

OBJECTIVES

Effective use of health data

SCOPE & EXPECTED IMPACT

Use of health data
(primary,
MyHealth@EU)

- Empower individuals to control their data
- Standardization and mandatory certification of EHR systems
- Voluntary labelling of wellness apps
- European Electronic Health Record Exchange Format

Single market for health data, data protection, free movement of people, digital goods and services

Re-use of health data
(secondary,
HealthData@EU)

- Health data access bodies
- Purposes for use and forbidden use
- Data permits, secure environments, no identification

Facilitated Research & Innovation

Better Policy Making

MEANS

Legal / Governance

Quality of data

Infrastructure

Capacity building/digitalisation (MFF)

The needs, expectations and views of stakeholders on economic sustainability



Options for governance models for data-sharing at EU level, in particular networking for secondary use

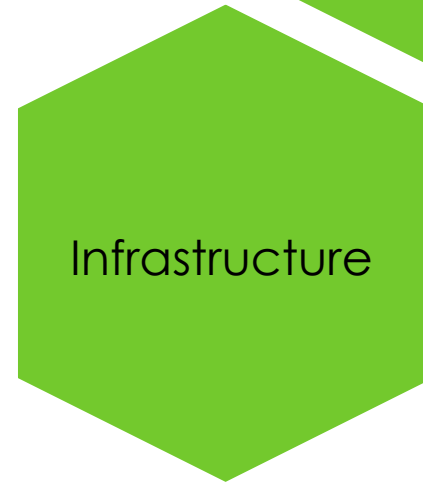
Definitions, good practices and use cases of GDPR-compliant data altruism



Secondary use of health data



Citizen perception of health data and data-sharing practices



Shared European data infrastructures across the EU (cf. eHDSI)

DATAN LAADUN TAUSTALLA ISO 8000-8 STANDARDI

HUS HYÖDYNTÄÄ KANSAINVÄLISIÄ STANDARDEJA HUS⁺

Laadunhallinta

Sertifikaatti
Nro 9942-01

Inspecta

Sertifikaatti
Nro 10349-01

kiwa

Tietoturvallisuuden hallinta

Sertifikaatti
nro 11560-01

kiwa

Inspecta Sertifiointi Oy on myöntänyt tämän sertifikaatin,
joka varmentaa, että organisaation

Helsingin ja Uudenmaan Sairaanhoidopiirin kuntayhtymä HUS-Tietohallinto Helsinki

laatu järjestelmä täyttää seuraavan standardin vaatimukset

ISO 9001:2015

Sertifiointiin sisältyvä toiminta

ICT-palvelut. Toimipaikat liitteessä.

Sertifikaatti on myönnetty 2018-02-13

Sertifikaatti on voimassa 2021-02-13 asti.



Mikko Törmänen, toimitusjohtaja

Sertifikaatti on voimassa edellyttäen, että organisaation laatu järjestelmä
täyttää jatkuvasti edellä mainitun standardin ja yleisen ohjeen ABC 200 vaatimukset.
Sertifikaatin voimassaolon voi tarkistaa osoitteesta www.inspecta.fi

Inspecta Sertifiointi Oy on myöntänyt tämän sertifikaatin,
joka varmentaa, että organisaation

Helsingin ja Uudenmaan sairaanhoidopiirin kuntayhtymä HUS Tietohallinto Virtuaalitiimi Helsinki

laatu järjestelmä täyttää seuraavan standardin vaatimukset

ISO 13485:2016

Sertifiointiin sisältyvä toiminta

**Algoritmeja, keinoälyä, koneoppimista ja data-analytiikkaa hyödyntävät
sovellukset ja järjestelmät sosiaali- ja terveydenhuollossa.**

**Main Technical Area: Active Medical Devices
Technical Area: General active medical devices
Product Categories Covered by the Technical Area: Software**

Sertifikaatti on myönnetty 2018-06-20.
Sertifikaatti on voimassa 2021-06-20 asti.



Mikko Törmänen, toimitusjohtaja

Sertifikaatti on voimassa edellyttäen, että organisaation laatu järjestelmä
täyttää jatkuvasti edellä mainitun standardin ja yleisen ohjeen ABC 200 vaatimukset.
Sertifikaatin voimassaolon voi tarkistaa osoitteesta www.inspecta.fi

Inspecta Sertifiointi Oy on myöntänyt tämän sertifikaatin, joka varmentaa,
että organisaation

Helsingin ja Uudenmaan Sairaanhoidopiirin kuntayhtymä HUS Tietohallinto Helsinki

tietoturvallisuuden hallintajärjestelmä täyttää seuraavan standardin vaatimukset

ISO/IEC 27001:2013

Sertifiointiin sisältyvä toiminta

**ICT-palvelut. Toimipaikat liitteessä.
Soveltuvuuslausunnon versio 1.00 (2020-12-02).**

Sertifikaatti on myönnetty 2021-04-26

Sertifikaatti on voimassa 2024-04-26 asti.



Maija Vanttaja, toimitusjohtaja

Sertifikaatti on voimassa edellyttäen, että organisaation tietoturvallisuuden hallintajärjestelmä täyttää edellä
mainitun standardin ja yleisen ohjeen ABC 200 vaatimukset. Sertifikaatin voimassaolon voi tarkistaa
osoitteesta www.kiwa.com/fi.

ISO STANDARDIEN LOGIIKKA JA ISO 8000-8



Sertifikaatti
Nro 9942-01

Inspecta

INTERNATIONAL
STANDARD

ISO
8000-8

First edition
2015-11-15

Inspecta Sertifiointi Oy on myöntänyt tämän sertifikaatin,
joka varmentaa, että organisaation

Helsingin ja Uudenmaan Sairaanhoidopiirin kuntayhtymä
HUS-Tietohallinto
Helsinki

Harmonisoidaan osaksi
ISO 9001
Laadunhallintajärjestelmää

laatujärjestelmä täyttää seuraavan standardin vaatimukset

ISO 9001:2015

Sertifiointiin sisältyvä toiminta

ICT-palvelut. Toimipaikat liitteessä.

Sertifikaatti on myönnetty 2018-02-13

Sertifikaatti on voimassa 2021-02-13 asti.

Mikko Törmänen, toimitusjohtaja

Sertifikaatti on voimassa edellyttäen, että organisaation laatujärjestelmä
täyttää jatkuvasti edellä mainitun standardin ja yleisen ohjeen ABC 200 vaatimukset.
Sertifikaatin voimassaolon voi tarkistaa osoitteesta www.inspecta.fi

Data quality —

Part 8:
**Information and data quality:
Concepts and measuring**

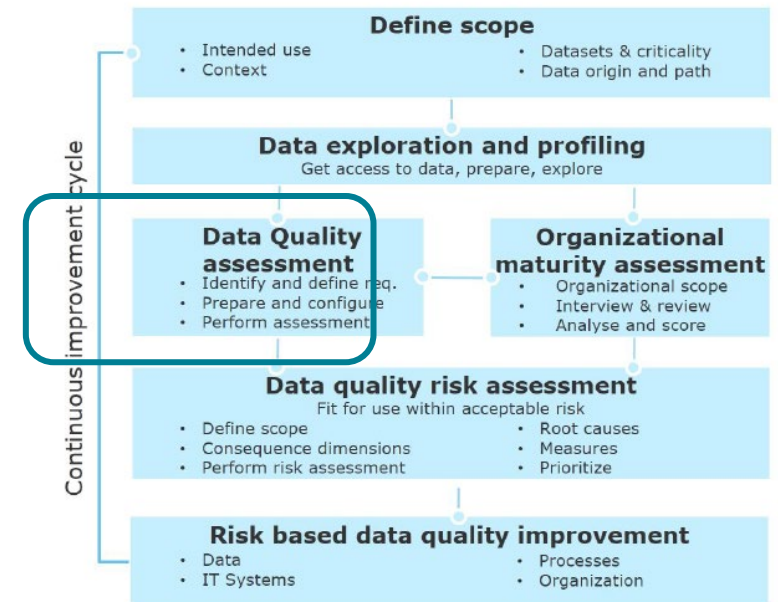
Qualité des données —

Partie 8: Informations et qualité des données: Concepts et mesurage

DATAN LAADUN HALLINNAN PÄÄLOTTUVUUDET

- Datun laadun arvioinnissa käytetään kolmea eri ulottuvuutta eli tarkastuspistettä
 - Syntaktinen ulottuvuus
 - Semanttinen ulottuvuus
 - Pragmaattinen ulottuvuus
- **Syntaktiset ja semanttiset tarkastuspisteet** ovat datun laadun verifiointia eli **todentavaa** vertailua tavoitearvoihin nähden,
- **Pragmaattinen tarkastuspiste** on datun validointia eli käyttötarkoituksen mukaista datun **soveltuvuuden hyväksymistä** käyttäjän näkökulmasta.
- Syntaktinen tarkastelu on yleensä täysin automatisoitua ja kohdistuu kaikkeen tai lähes kaikkeen dataan.
- Semanttinen tarkastelu sen sijaan toteutetaan vain valitulle joukolle dataa.
- Pragmaattinen datun laatutarkastelu toteutetaan datun käyttäjän näkökulmasta.

Det Norske Veritas-tulkinta standardista



... ESIMERKKEJÄ

Syntaktinen ulottuvuus (Tekninen todentaminen eli verifiointi):

Datan, mallinnuksen ja raportointitarpeen tekninen tarkistuspiste. Tällöin data vastaa metadatasia esitettyjä rakenteellisia ja muotosäännöllisiä määrittelyjä.

- Esimerkki 1: potilaalla ei voi olla kahta tunnistetta (identifioijatus).
- Esimerkki 2: Tietokentässä ei ole tyhjiä arvoja.
- Esimerkki 3: Viikonnumeroa nn ei voi laittaa kenttään, jossa edellytetään pp.kk.vvvv –arvoa

Semanttinen ulottuvuus (Asiasisällöllisen oikeellisuuden todentaminen eli verifiointi):

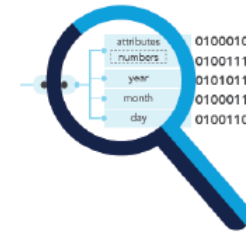
Tarkistetaan datan ja raportointitarpeen sisällön välinen oikeellisuus.

- Esimerkki 1: Käsittemalli kuvaa käsitteiden ja niihin liittyvän datan suhteita (käsitteellinen oikeellisuus).
- Esimerkki 2: Potilalla on tietokannan osoitetietueessa oikea osoite eikä väärää osoitetta (asiasisällöllinen oikeellisuus).
- Esimerkki 3: Data ja sen representaatio ovat verifioitavissa eli todennettavissa (esim. referenssidatajoukon avulla lähdejärjestelmästä tulevan datan vastaanoton hyväksymiskriteerinä).

Pragmaattinen ulottuvuus (Käyttötarkoituksen mukainen hyväksyminen eli validointi):

Tarkistetaan, että tuotettu raportointikokonaisuus vastaa raportointitarvetta ja että data on ymmärrettävää ja käyttötarkoitukseensa soveltuvaa ja että datan välittämään tietoon (informaatioon) voidaan luottaa.

- Esimerkki 1: Raportoitava tieto vastaa toiminnallisia tarpeita.
- Esimerkki 2: Metadata on olemassa.
- Esimerkki 3: Datan ajantasaisuus on asianmukainen käyttötarpeeseen nähden.



Syntactic

Verify according to technical and business requirement



Semantic

Verify if data correspond to real world / what it represents



Pragmatic

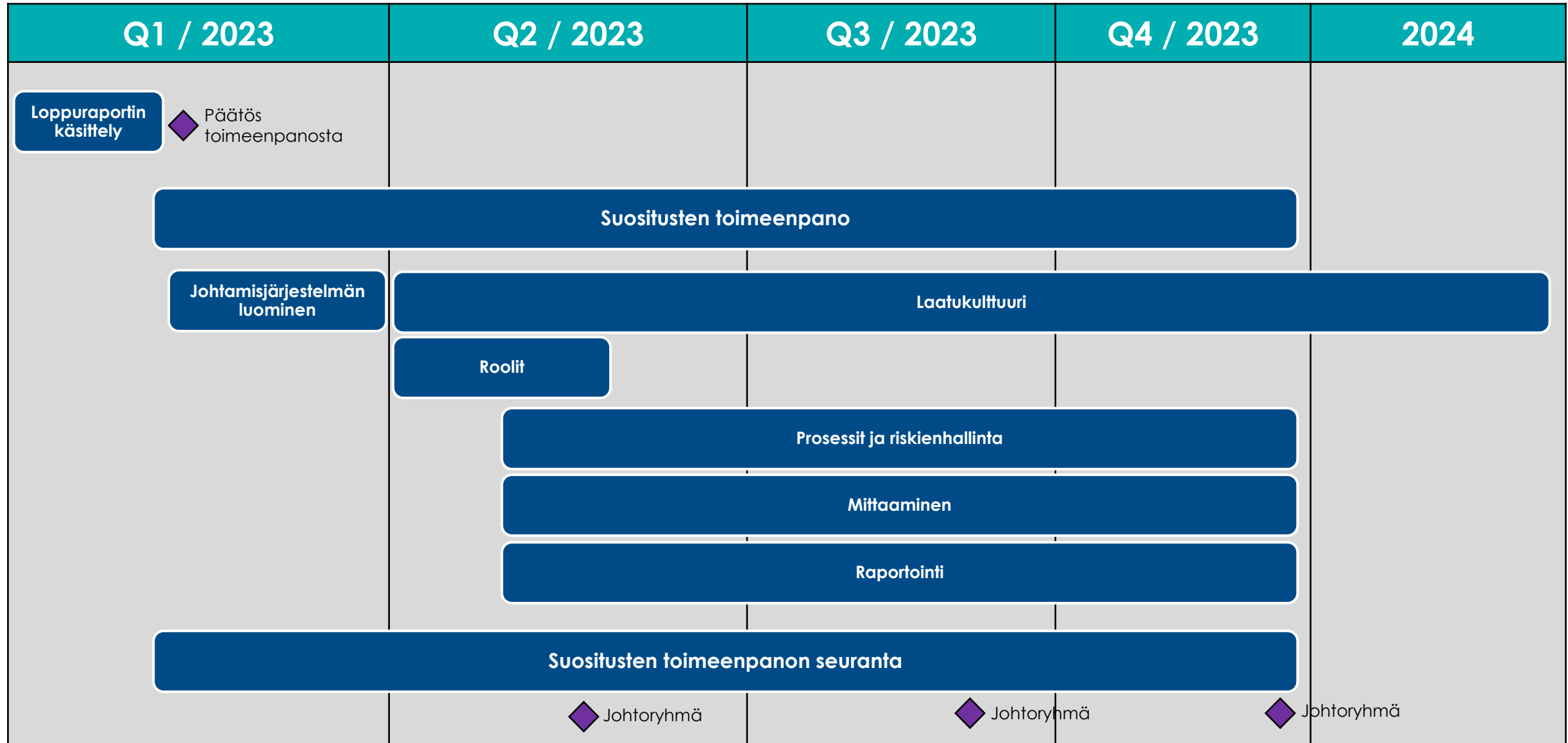
Validate if you have the data needed for intended use

Käsitteitä:

- Verify = todentaminen
- Validate = kelpuutus
- Intended use = käyttötarkoitus
- Metadata = tietoa tiedosta. Se mahdollistaa tiedon jalostamisen erilaisiin käyttötarkoituksiin. Se sisältää erilaisia tiedon hallinnan parametreja ja attribuutteja (ulottuvuuksia).

DATAN LAADUN HALLINNAN KEHITÄMISPOLKU HUSISSA

SUOSITUKSET – DATAN LAADUN (JAKTO)TIEKARTTA **HUS***



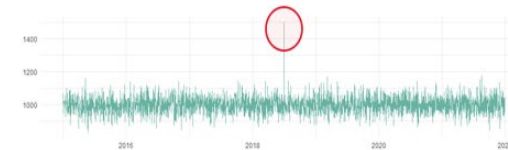
SYNTAKTINEN LAATU –
TAPAUSESIMERKKI:
ANOMALIOIDEN JÄLJITTÄMINEN
AIKASARJOISTA

SYITÄ SYNTAKTISEN LAADUN HALLINNALLE

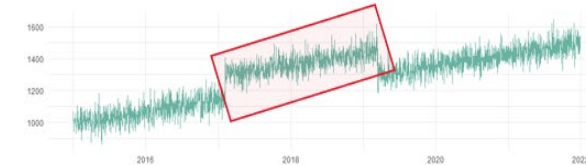
- Aineistojen laadun haasteet: aineistoihin voi syntyä virheitä monin eri tavoin (aineiston kirjaamishaasteet, ongelmat integraatioissa ja aineistoputkissa, väärin määritellyt logiikat aineistotuotannossa, ...)
- Aineiston käyttöön vaikuttaa yli ajan tapahtuvat muutokset, mm. kirjaamiskäytännöissä, järjestelmävaihdoksissa
- Suuri osa (sairaala)maailman aineiston suureista seuraavat ennakoitavaa trendiä
 - Ennakoimattomat aineistopisteet tai äkkinäiset muutokset aikasarjoissa voivat indikoida ongelmia aineiston laadussa
 - Nämä huomioitava käytössä

ANOMALIOIDEN LUOKKIA

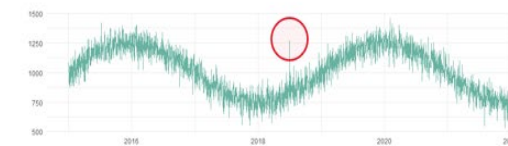
- *Pisteanomalia*



- *Kollektiivinen anomalia*



- *Kontekstuaalinen anomalia*



ANOMALIA TUNNISTAMISEN PROSESSI

1. Anomalioiden tunnistaminen

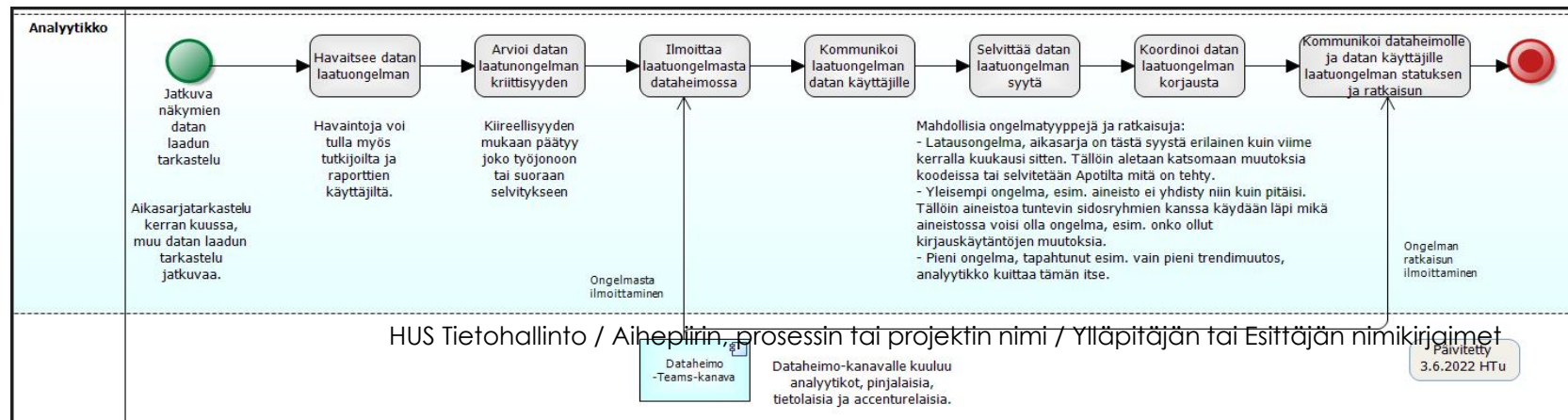
- Kuukausittainen aikasarjamonitorointi hyödyntäen PELT-algoritmia
- Muut huomiot

2. Selvittelytyö

- Aikasarjan kvalitatiivinen analysoiminen
- Yhdistellään kaikkia mahdollisia tietolähteitä, ml. klinikoiden/muiden tiedon käyttäjien kontaktointi

3. Toimenpiteet

- Tiedon kerääminen/tallentaminen dokumentaatioon ja aineistojalosteiden logiikkaan ja kehittämistyöt

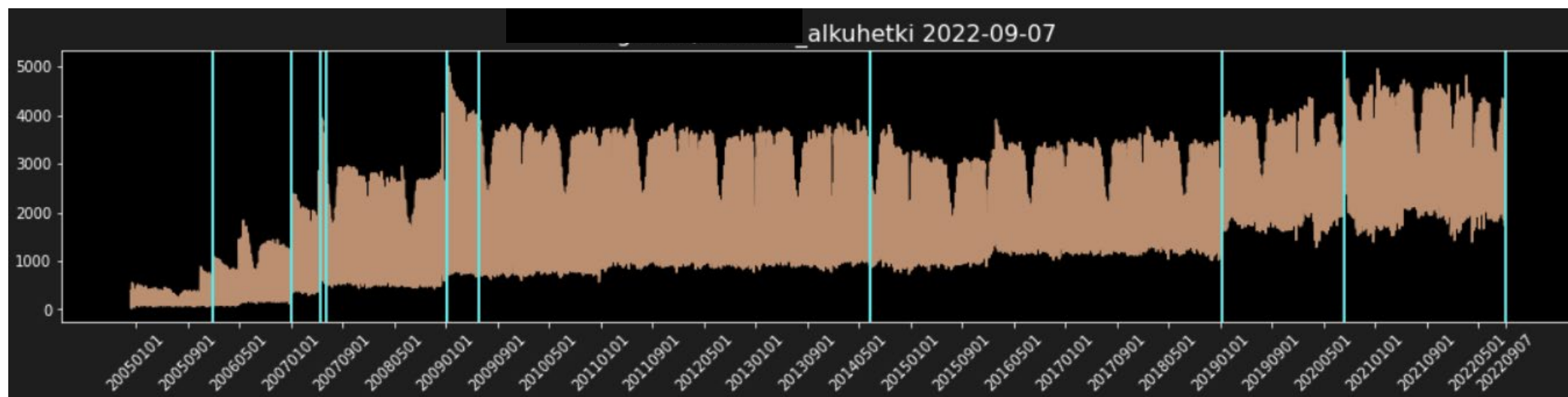


MENETELMÄN KUVAUS: PELT + REGRESSIO + HEURISTIIKKA 1

1. Etsitään muutospisteet PELT (Pruned Exact Linear Time) metodilla

- Python paketti ruptures
- $\mathcal{O}(n)$

$$\sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1): \tau_i}) + \beta]$$



MENETELMÄN KUVAUS: PELT + REGRESSIO + HEURISTIIKKA 2

2. Sovitetaan regressiosuorat aikaväleille $\tau_1 : \tau_m$ ja lasketaan

- Kulmakertoimet $\beta_{\tau_1} : \beta_{\tau_m}$
- Absoluuttisten residuaalien keskihajonnat jokaiselle aikavälille
- Keskiarvot
- Regressiomallin mukainen ennuste seuraavan välin keskiarvolle

$$\frac{1}{N(y_{(\tau_i+1): \tau_{i+1}})} \sum \hat{y}_{(\tau_i+1): \tau_{i+1}}$$

MENETELMÄN KUVAUS: PELT + REGRESSIO + HEURISTIIKKA 3

3. Verrataan perättäisten aikavälien
 - Kulmakertoimia
 - Keskihajontaa trendin ympärillä
 - Ennustetun ja toteutuneen keskiarvon erotusta

$$\frac{1}{N(y_{(\tau_i+1): \tau_{i+1}})} \sum y_{(\tau_i+1): \tau_{i+1}} - \frac{1}{N(\hat{y}_{(\tau_i+1): \tau_{i+1}})} \sum \hat{y}_{(\tau_i+1): \tau_{i+1}}$$

... ja päätellään heuristisesti, onko tulokset merkitseviä (eroavatko kulmakertoimet, keskihajonnat tai ennustettujen ja toteutuneiden keskiarvojen suhteelliset arvot enemmän kuin asetetut raja-arvot)

RAPORTOINTI

- Aikasarja-analyysin tuloksia raportoidaan loppukäyttäjille
 - ISO 8000-8 vaatimus
 - HUS Intra
 - Raportti päivitetään kaksi kertaa vuodessa
- Tavoite
 - Läpinäkyvyys aineiston aggregaattitason laadusta
 - Lisää informaatiota aineiston kvalitatiivisista muutoksista
 - Raportointi lisää kannusteita aineiston laadun kehittämiseen ja ylläpitoon
 - Mahdollistaa ulkoisen raportoinnin

Mallin esimerkkituloste:

```

Model 6 and model 7
Coefficient condition: False
* Model 6:
  [0.03710245]
* Model 7:
  [0.15427271]
Deviation condition: True
* Model 6:
  0.1958520204576159
* Model 7:
  0.1671738019067703
Mean condition: True
* Expected mean:
  2565.5421058535676
* Actual mean:
  2382.7588908981315
  
```

Sanallinen selite:

Muutospisteet: 01.01.2006, 01.01.20
10.08.2020

Selvitetyt: 01.01.2006, 01.01.2007 ja

SEMANTTINEN DATAN LAATU - KÄSITE- JA TEKNISET MALLIT HUSISSA

KANSALLISET KÄSITEMALLIT SEMANTTISEN YHTEENTOIMIVUUDEN VARMISTAJANA ->

- Virta-hankkeessa DigiFinlandin ja alueiden yhteisenä tavoitteena on mahdollistaa hyvinvointialueiden johtaminen luotettavalla ja ajantasaisella tiedolla.
- DigiFinlandin Virta-hankkeessa tuotetaan kansallisiin määrittämiin perustuvia käsitelmalleja, joissa kuvataan keskeisiä sote-järjestäjän tietojohdamisessa tarvittavia kansallisia käsitteitä, käsitteiden välisiä suhteita sekä käsitteiden lisätietoja.
- Yhtenäisillä käsite- ja tietomalleilla on tarkoitus varmistaa kansallisesti vertailukelpoiset tietosisällöt.
- Virta-hankkeessa tehtävän käsitelmallityön tavoitteena on helpottaa hyvinvointialueiden työtä tietojohdamisen ratkaisujen toteuttamisessa

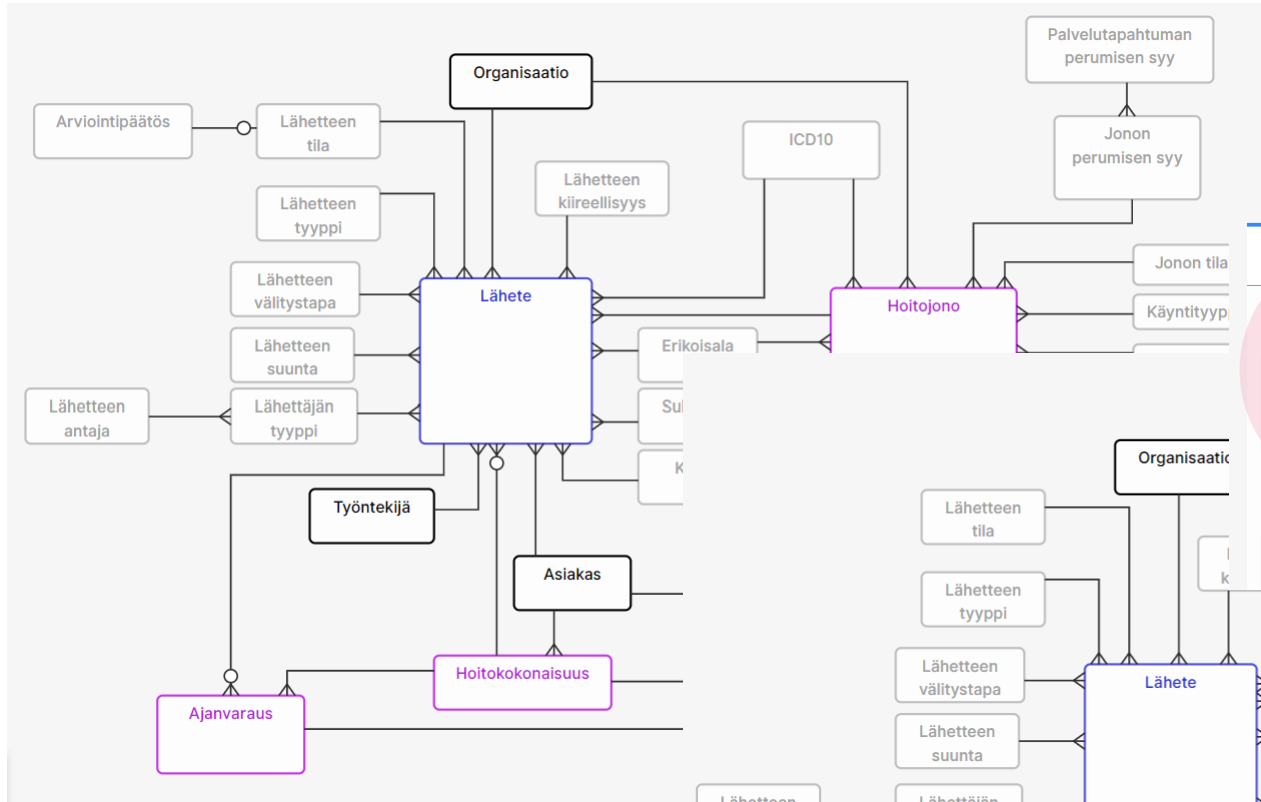
Semanttinen laatu: asiasisällöllisen oikeellisuuden todentaminen eli verifiointi

-> MALLINTAJAN JA DATAVAULTAAJAN LÄHDEAINEISTO

- Virta-kuvaukset
 - Toiminnan työkulujen kuvaukset (jos saatavilla)
 - Lähdejärjestelmäkuvaus ja -asiantuntijatt
 - Ratkaisukuvaus
 - Rajapintakuvaus: esimerkkinä Henkilöstösuunnittelu- ja laskenta (SuLa)
 - Taulun nimi
 - Sarakkeiden nimi
 - Kerättävä tieto (attribuuttitieto, satelliitti)
 - Muoto (muuttujan tyyppi esim. char, datetime)
 - Mahdolliset lisätiedot
 - Vastaavuudet toisien järjestelmien tietosisältöihin
 - Fyysiset taulukuvaukset (jos saatavilla)
- HUS: Henkilöstöhallinto / Ainehallinto, prosessin tai projektin nimi / Ylläpitäjän tai Esittäjän nimikirjaimet

SEMANTTISEN LAADUN VARMISTUS HUSIN TIETOVARASTOKEHITYKSESSÄ

- HUS:n nykyiset käsittemallit käydään läpi suhteessa DigiFinlandin kansallisiin käsitemaleihin
- Käsittemallit käydään läpi yhdessä liiketoiminnan kanssa ja peilataan nykyiset käsitteet suhteessa kansallisiin käsitteisiin → mallintajilla tulee olla yhteinen näkemys liiketoiminnan kanssa
- Käsittemallit ylläpidetään Ellie-työkalussa, johon samalla kullekin käsitteelle kirjataan kuvaukset → Dokumentointi alkaa Ellie -työvälineessä
- Liiketoiminnan käsitemallista muodostetaan tekninen käsitemalli eli jätetään vain ne asiakokonaisuuteen liittyvät käsitteet, joihin on ladattavissa dataa tietolähteistä ja jotka eivät esiinny muissa käsitemaleissa
- Tekninen käsitemalli ohjaa tietovarastokehitystä eli jokaisesta käsitteestä muodostuu entiteetit tietovaraston Raaka Data Vault –kerrokseen.



Kuvaukset ja määrittelyt saa tulostettua

Entity type: Contract

Hoitojono

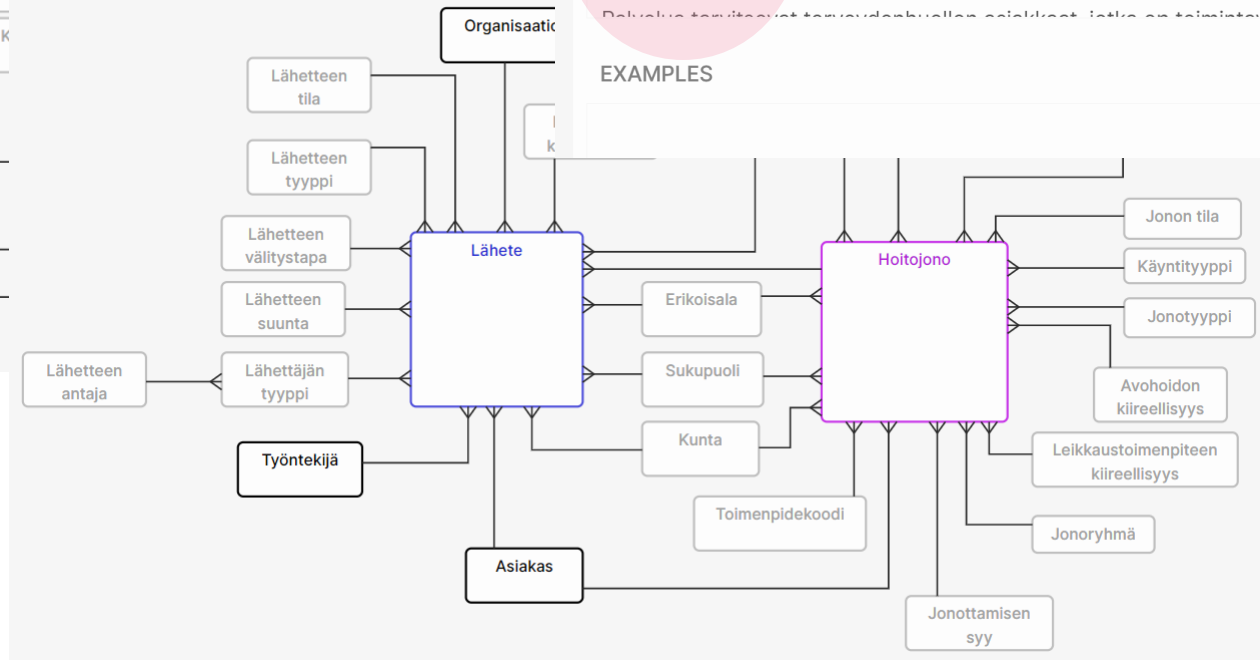
- ABOUT
- ATTRIBUTES
- RELATIONSHIPS
- LOGICAL

DESCRIPTION

Kiireettömään hoitoon pääsyä kiireellisyysjärjestyksessä odottavien henkilöiden joukko. (sotesanastot.thl.fi)

EXAMPLES

SYNONYMS



SEMANTTISEN LAADUN VARMISTUS HUS:N TIETOVARASTOKEHITYKSESSÄ

- Tietovarastokehitys tehdään tiukasti käsittemallipohjaisesti eli tietovaraston Raaka Data Vault –kerrokseen voi tulla ainoastaan entiteettejä, jotka perustuvat teknisten käsittemallien käsitteisiin
- Tekninen käsittemalli vietään WhereScape 3D –työvälineeseen Ellien ja 3D:n integraation avulla. WhereScape 3D:n konversiosäännöstöllä käsittemallista konvertoidaan looginen ja Raaka Data Vaultin fyysinen tietokantamalli. Mallinnetut fyysiset tietokantarakenteet siirretään WhereScape RED -välineeseen, jossa automatiikka generoi latauskriptit tietojen lataamista varten.
- Liiketoiminnan tunnistamat ja kansallisesti yhteiset termit siirtyvät suoraan tietovarastokerrokseen jolloin niiden hyödyntäminen ja tulkinta on ristiriidatonta

TYÖVÄLINEITÄ DATAN LAADUN MITTAAMISEN PRIORISOINTIIN

KYSELYT PRAGMAATTISEN LAADUN TARKASTELUSSA

Tavoite: Saada näyttöä raportin / raporttien soveltuvuudesta käyttötarkoitukseensa tiedon käyttäjän näkökulmasta

Miten suunnitella ja toteuttaa kysely?

- ❑ Mitä kysytään ja keneltä?
 - ⇔ *Case: HUSin PowerBI (otos, 15 eri raporttia)*
- ❑ Kyselyn reunaehdot ⇔ Ankkurointi raportteihin
 - ✓ Määrämuotoisuus, kvalitatiivisuus ja kvantitatiivisuus
 - ✓ Luotettavuus, edustavuus, uusittavuus, monistettavuus
 - ✓ Tilastollinen käsittely
- ❑ Muita aihepiirirajauksiin vaikuttavia tekijöitä
 - Aiempien kyselyjen hyödyntäminen
 - Tämän kyselyn kohderyhmä (mm. osoitetiedot)

Vastaaajien taustatiedot:

- Ammattiryhmä ⇔ käyttäjäryhmä
- Raportin käyttöiheys
- Organisaatiotieto ⇔ kohdealue
- Raportin aihepiirit (+ käyttötarkoitus)

❑ Määrämuotoisten kysymysten asteikko:

- Täysin eri mieltä
- Jokseenkin eri mieltä
- Ei eri eikä samaa mieltä
- Jokseenkin samaa mieltä
- Täysin samaa mieltä
- En osaa sanoa (EOS)

❑ Vapaat kysymykset => Ryhmittely

PRAGMAATTINEN LAATU: Loppukäyttäjäkyselyn toteutus



Aihepiiri 1: (subjekttiivinen arvio)

- Raportit ovat helposti saatavilla
- Raporttien käyttäminen on sujuvaa
- Käyttämäni raportit ovat helposti opittavissa
- Raportit ovat helposti muokattavissa
- Raportit vastaavat käyttötarpeitani
- Voin suositella raporttia myös kollegoilleni

Aihepiiri 2: Käyttökokemuksesi perusteella käyttämäsi raportin laatu on

- virheetöntä*
- johdonmukaista*
- tarpeellista
- luotettavaa
- ajantasaista*
- lisäarvoa tuottavaa
- vertailun mahdollistavaa
- ymmärrettävää
- kattavaa*

*-merkityt Gartnerin jaottelussa ovat objektiivisia datan laadun ulottuvuuksia

Aihepiiri 3: Valitse raporttien ominaisuuksista tärkeimmät, toiseksi tärkeimmät ja kolmanneksi tärkeimmät attribuutit

- Virheettömyys (*accuracy*)
- Johdonmukaisuus (*consistency*)
- Tarpeellisuus (*relevancy*)
- Luotettavuus (*believability / reliability*)
- Ajantasaisuus (*timeliness, current*)
- Lisäarvoa tuottava (*value-added*)
- Vertailtavuus (*comparability*)
- Ymmärrettävyys (*understandability*)
- Kattavuus (*completeness, representativeness*)

Vapaa kysymys:

- Datun laatuun ja sisältöön liittyvät kehittämissuhteukset (Raportin nimeäminen, jos kohteena tietty raportti)

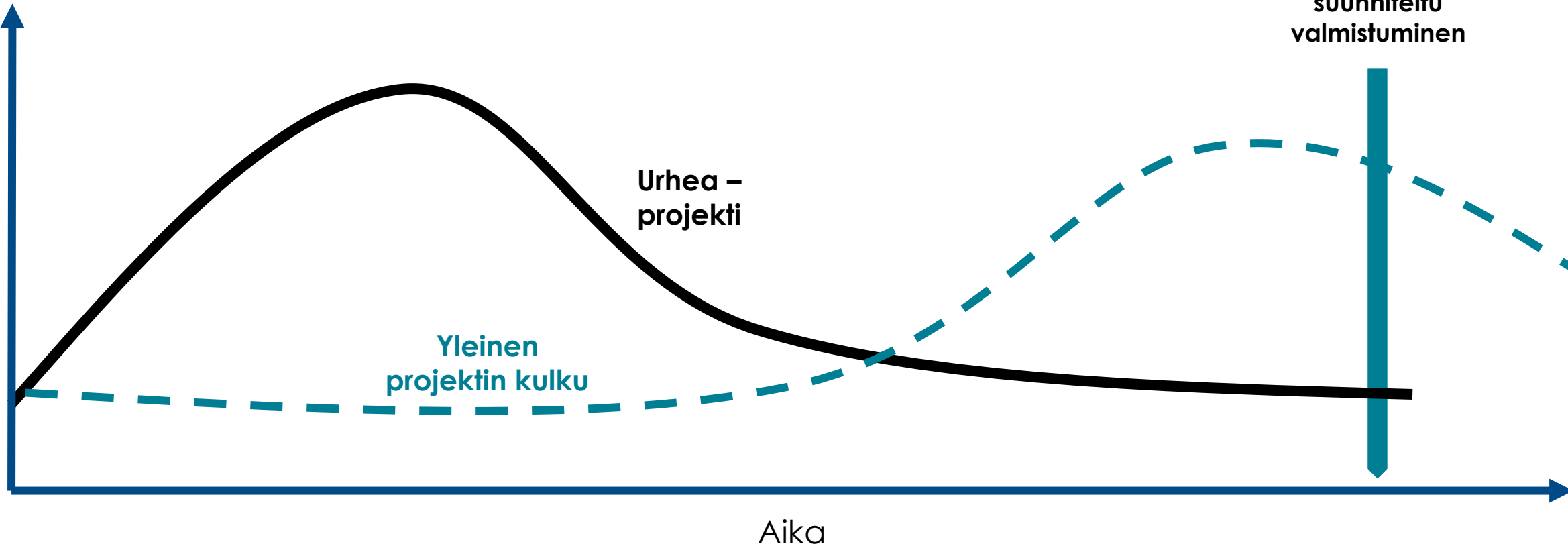
Kyselyn vastausprosentti oli
n. 22% (40/180)



KEHITTEILLÄ HUSISSA.....

DATAN LAADUN HALLINNAN LÄHTÖKOHTA PROJEKTEISSA

Datan
laatuun
liittyvä
työmäärä



DATAN LAADUN MITTARIT

Mittarit	Kuvaus	Esimerkkejä käytännössä	Miten Urheassa voitaisiin tehdä?
Accessibility	Data on saatavilla, helposti haettavissa ja/tai liitettyä liiketoiminnan prosessiin.	Oikeat raportit ja/tai tietomallit Power BI:ssä prosessin omassa Team kanavassa.	Voisiko verrata käsitemalliin? Prosesseittain käsitemallinäkökulma?
Accuracy	Datassa olevat arvot/luvut vastaa tarkoituksen mukaisesti oikean elämän tapahtumia ja asioita.	Osaston henkilömäärä tietyssä ajan hetkenä vastaa raportilla olevia arvoja.	Tietovaraston/raportoinnin näkökulmasta tämä olisi lukuja oikeellisuus tai esim. ettei ole mitään keksittyjä henkilöitä (testitapauksia)
Completeness	Tietueista ei puutu kenttiä eikä datakokonaisuuksista ei puutu tietuejoukkoja.	Tärkeä kenttä puuttuu taulusta tai jonkun päivän csv -tiedosto puuttuu.	Tämä on ehkä enemmän tietoaletta juttu. Eli jos nämä puuttuu niin jää kiinni jo ennen tietovarastoa. Voisi olla myös sitä, että puuttuu kenttiä tai ekstrakteja vs. data katalogin tiedot.
Consistency	Sama tieto useassa järjestelmässä/paikassa näyttää samalta ja on samassa muodossa.	henkilön työntekijätunnus jokaisessa järjestelmässä samassa muodossa.	Esimerkiksi verrataan kaikkien järjestelmien työntekijätunnuksia toisiinsa -> löytyykö tuplia?
Existence	Liiketoiminnalle tärkeä data on olemassa järjestelmissä.	esineen tilavuus ei löydykään järjestelmistä vaikka sitä tarvitaan päätöksenteossa.	Voisiko verrata käsitemalliin? Prosesseittain käsitemallinäkökulma?
Integrity	Oletuksen mukaan yhdistyvät datat eri järjestelmistä pystytään oikeasti yhdistämään toisiinsa.	Kahden järjestelmän saman käsitteen datat eivät sisällä tarkkaa linkkiä.	Voidaan tehdä tarkistus että kuinka moni rivi osuu eri järjestelmistä toisiinsa. (palvelujakso?)
Precision	Datat kirjataan liiketoimintaprosessin vaatimuksen mukaisella tarkkuudella.	Vaadittuun projektin numero kenttään on kirjoitettu jotain muuta kuin vaadittu numero.	Tämä voidaan tehdä. Vaatii, että liiketoiminta määrittelee säännöt, mitä vasten tarkistetaan. Hyvä tietovarastoon (kunhan ei haukata resursseja liikaa).
Reliability	Data tulee tunnetusta ja luotettavasta lähteestä.	Organisaation virallisesti käyttämät prosessin järjestelmät.	DataVault ainakin määrittää, että lähdetieto löytyy. Käsitemalliin vertailu?
Timeliness	Data päivitetään niin usein kuin liiketoiminnan vaatimukset määräävät.	Kerran yössä, 15 min välein...	Onko tämä sitä, että toteutuuko oikeasti vai että on mekanismi käytössä jolla refresh tehdään?
Uniqueness	Useassa paikassa sijaitsevalla datalla on sama arvo/luku.	Henkilöllä ei ole eri työntekijäID:tä eri järjestelmissä.	Tämä voidaan tehdä. Vaatii, että liiketoiminta määrittelee säännöt, mitä vasten tarkistetaan. Hyvä tietovarastoon (kunhan ei haukata resursseja liikaa).
Validity	Data mukautuu määritettyihin liiketoiminnan sääntöihin ja kuuluu sallittuihin parametreihin.	Osasto löytyy organisaation virallisesta osastojen listasta.	Tämä voidaan tehdä. Vaatii, että liiketoiminta määrittelee säännöt, mitä vasten tarkistetaan. Hyvä tietovarastoon (kunhan ei haukata resursseja liikaa).

TIETYN AJAN HETKEN NÄKYMÄ DATAN LAATUUN - Esimerkkinä HR-data

	Työntekijä		Palvelussuhde		Palvelujakso		Henk. määrä (mittari)
Accessibility	67%	33%	33%	33%	33%	N/A	
Accuracy	99%	99%	99%	99%	99%	99%	
Completeness	N/A	N/A	N/A	N/A	N/A	N/A	
Consistency	100%	N/A	65%	99%	99%	99%	
Existence	N/A	N/A	N/A	N/A	N/A	N/A	
Integrity	99%	96%	95%	95%	95%	N/A	
Precision	99%	98%	97%	97%	97%	98%	
Reliability	75%	50%	50%	50%	50%	50%	
Timeliness	95%	95%	95%	95%	95%	95%	
Uniqueness	99%	75%	86%	86%	86%	N/A	
Validity	N/A	N/A	N/A	N/A	N/A	N/A	

Kiitos mielenkiinnosta

Timo Hakala

Tietoarkkitehti

HUS Tietohallinto

Osoite: PL 760, 00029 HUS

Käyntiosoite: Paciuksenkatu 25, Helsinki

Sähköposti: timo.hakala@hus.fi

Puh.: +358 40 537 8828