

# Priorisoi datan laatu

Datan lähteet, rajapinnat ja laatu

Pk-yrityksen menestyksessä



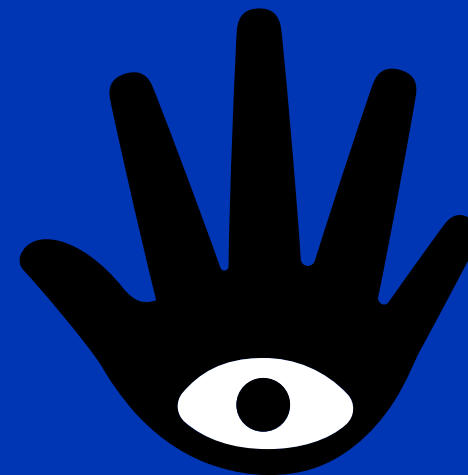
# Jani Nurmi

- **CEO, Partner @ Hubble Oy** 2017/04-
  - Software Architect / Product Owner @ Roima Intelligence
  - Software Architect / Product Owner @ Modultek Oy
  - Senior Software Architect @ Enersize Oyj
  - Development Manager @ Staart Oy
  - CEO @ Staart Oy
- 
- ISO 8000 Master Data Quality Manager
- 
- SAMK, Master of Business Administration, Business Management and Entrepreneurship 2013 –
  - SAMK, Bachelor, Business Information Systems 2000 – 2005



Hubble

hubble.fi



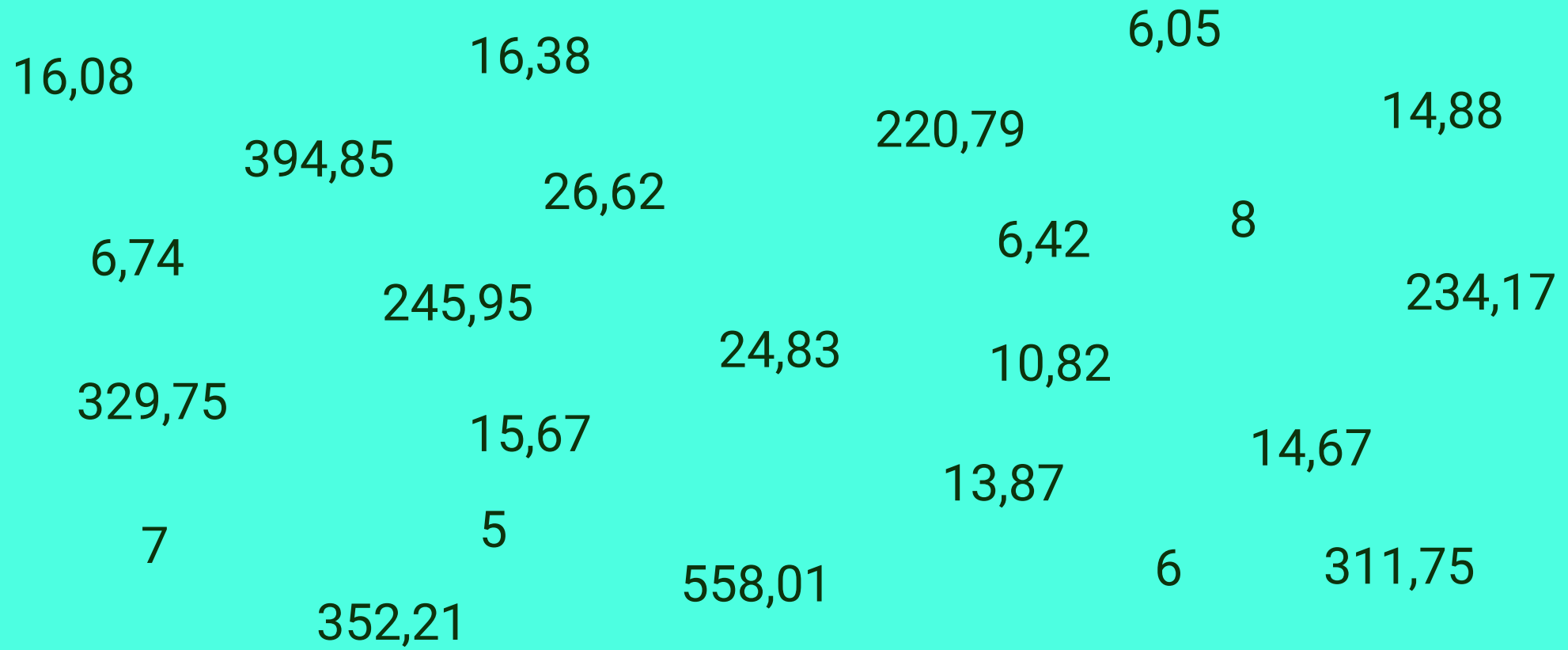
SILMA.IO

# Mitä odotettavissa?

- Herättää ajatuksia
  - Apua tunnistamaan datalähteitä
  - Ymmärrystä rajapinnoista
  - Tietoa datan laadusta
- 
- Ei tarvitse tehdä muistiinpanoja kuin omista ajatuksista.  
Materiaali tulee jakoon.

# Data, mutu ja totuus

# Mitä data on?



# Mitä data on?

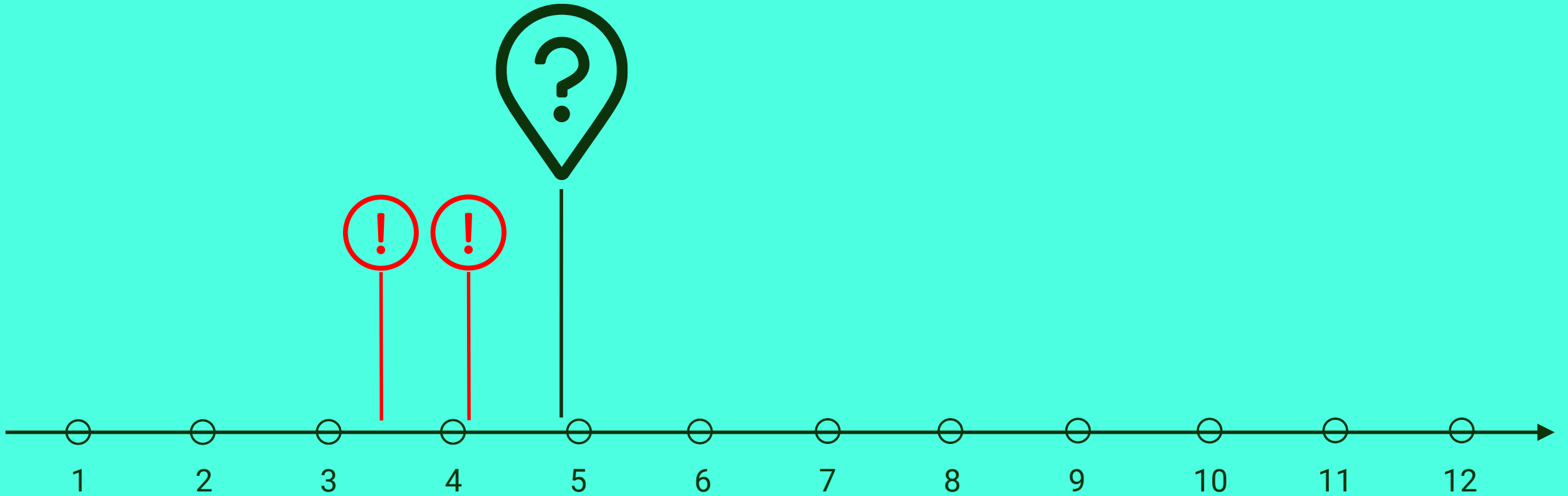
5	558,01	24,83	394,85	10,82	26,62
6	352,21	15,67	234,17	6,42	16,38
7	329,75	14,67	245,95	6,74	16,08
8	311,75	13,87	220,79	6,05	14,88

# Mitä data on?

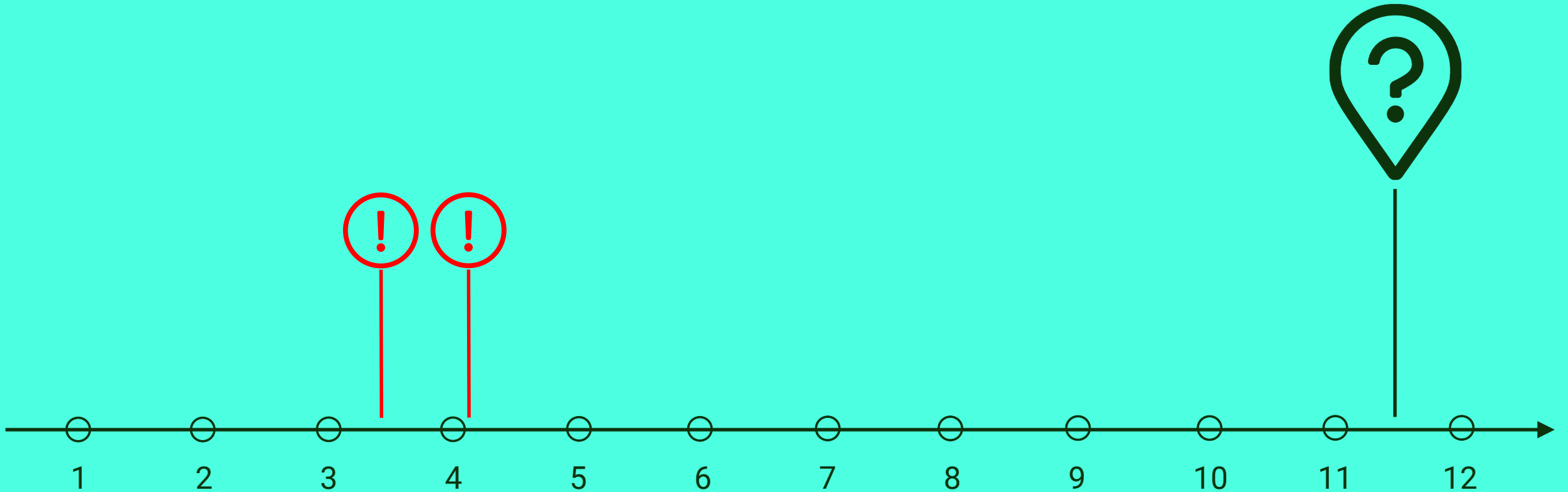
Kk	Päiväsiirto	Ps hinta	Yösiirto	Ys hinta	Sähkövero
5	558,01 kWh	24,83 €	394,85 kWh	10,82 €	26,62 €
6	352,21 kWh	15,67 €	234,17 kWh	6,42 €	16,38 €
7	329,75 kWh	14,67 €	245,95 kWh	6,74 €	16,08 €
8	311,75 kWh	13,87 €	220,79 kWh	6,05 €	14,88 €



# Mutu vai totuus



# Mutu vai totuus



# Datan lähteet

# Järjestelmät ja palvelut

- Toiminnanohjaus
- Taloushallinta
- Asiakkuudenhallinta
- Tuotannonohjaus
- Prosessinohjaus
- Varastonhallinta
- IoT-järjestelmät
- Palautejärjestelmät
- Verkkosivuston analytiikka

# Sovellukset ja tiedostot

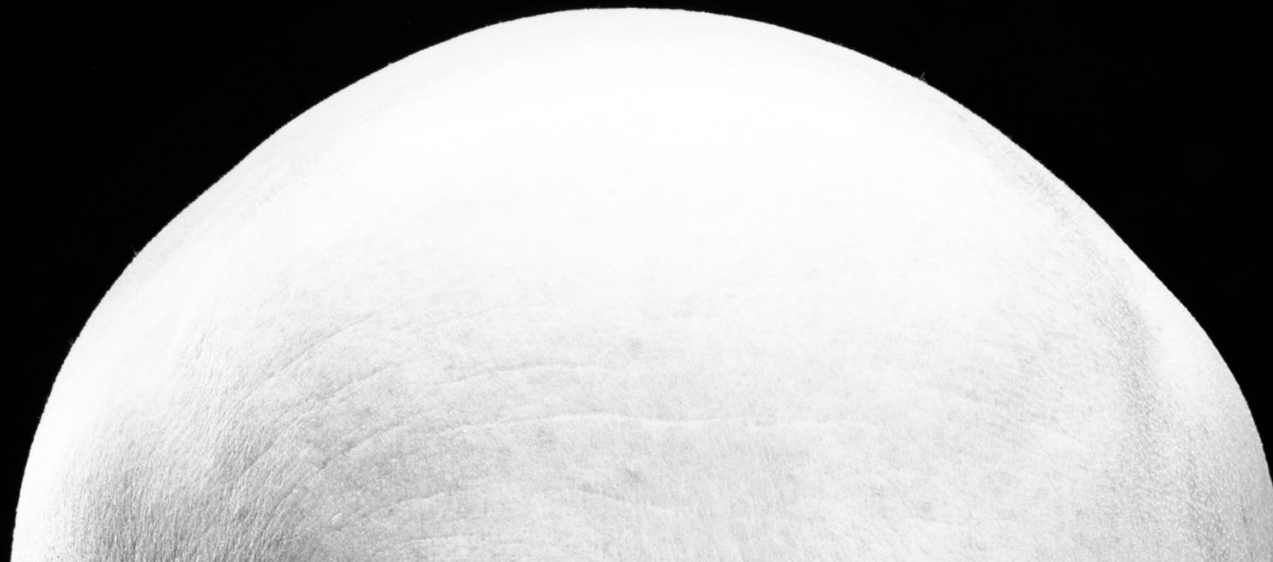
- Suunnitteluohjelmat
- Laskentaohjelmat
- Excelit
- Wordit
- Pdf-tiedostot
- Muut raportit

# Avoim data

- [avoindata.fi](http://avoindata.fi)
- [www.ilmatieteenlaitos.fi/avoim-data](http://www.ilmatieteenlaitos.fi/avoim-data)
- [www.pori.fi/kaupunki-ja-hallinto/pori-tieto/avoim-data](http://www.pori.fi/kaupunki-ja-hallinto/pori-tieto/avoim-data)
- [avoindata.prh.fi](http://avoindata.prh.fi)
- [data.europa.eu/en](http://data.europa.eu/en)
- [datasetsearch.research.google.com](http://datasetsearch.research.google.com)

# Vaikeimmat

- Muistiinpanot, vihot, merkinnät
- Ajo- ja käyttöpäiväkirjat
- Nahkakantiset, eli hiljainen tieto



# Rajapinnat



# Rajapinta, integraatio, API (Application Programming Interface)

- Mahdollistaa datan sujuvan siirtymisen järjestelmästä toiseen tai analytiikkaratkaisuun ilman manuaalisia prosesseja
- Saattaa maksaa lisähintaa normaalien käyttökustannusten päälle
- Ei välttämättä pääsyä kaikkeen järjestelmän dataan
- Julkinen dokumentaatio yleistymässä, mutta voi joutua pyytämään
- Laajuus, dokumentaatio ja hinnoittelu tarkistettava aina viimeistään ennen järjestelmän tai palvelun hankintaa
- Esimerkit

# Datan laatu

# Datan laadun ulottuvuudet

1. **Tarkkuus** (Accuracy)
2. **Täydellisyys** (Completeness)
3. **Johdonmukaisuus** (Consistency)
4. **Yksilöllisyys** (Uniqueness)
5. **Ajantasaisuus** (Timeliness)
6. **Pätevyys** (Validity)

Datakohtaisesti  
ulottuvuuksien  
tärkeydet  
vaihtelevat

# Tarkkuus

- Ulottuvuus viittaa datan oikeellisuuteen ja tarkkuuteen
- Data on tarkkaa, jos se oikein heijastaa todellisen maailman arvoja tai olosuhteita virheettömästi.
- Toinen tapa mitata tarkkuutta on yksinkertaisesti vertaamalla todellisia arvoja luotettavan lähteen tarjoamiin standardi- tai viitearvoihin. Esim. luotettava tilastodata

# Täydellisyys

- Täydellisyys tarkoittaa sitä, että kaikki tarvittava tieto on läsnä.
- Datasetti on täydellinen
  - Kaikki tietueet löytyy.
  - Tietueiden kaikki tarvittavat attribuutit (kentät) löytyy.
  - Tietueiden tarvittavien attribuuttien arvot ovat kokonaisia

# Johdonmukaisuus

- Data ei ole ristiriidassa itsensä kanssa eri järjestelmissä tai tietueissa
- Data noudattaa samoja formaatteja ja standardeja eri järjestelmissä
- Data on yhtenevää eri aihealueiden välillä
- Aikasarjadata on johdonmukaista

# Yksilöllisyys

- Datassa ei ole duplikaatteja
- Voi muodostua, jos on täysin sama tietue kahteen kertaan
- Voi muodostua, jos on kaksi eri tietuetta, mutta ne ovat oikeasti sama asia. Esim. sama kirjoitettu vähän eri tavoilla

# Ajantasaisuus

- Data on ajan tasalla
- Data saadaan käyttöön oikea-aikaisesti



# Pätevyys

- Data täyttää sille määritellyt säännöt
- Numerodatan tarkistukset
- Päivämäärien formaatit
- Aikaleimojen aikavyöhykkeet
- Numerodatan arvojen ”haarukat”
- Tapahtumien järjestys


Perustiedot

# Tyypillisesti tärkeimmät perustiedot

- Asiakastiedot
- Tuotetiedot
- Toimittajatiedot
  
- Määrittelyt perustietojen laadulle
- Mikä on minkäkin perustiedon master-järjestelmä?

# Perustietojen luokitukset

- Erittäin tärkeä tiedolla johtamisen kannalta
- Yhdellä luokituksella kuvataan hyvin rajattua asiaa
- Luokittelun asiat tulisi olla samantasoisia henkilöauto vs. farmariauto
- Vältetään ”-muut”-tyyppisiä luokituksia
- Jos on vaikeaa päättää mihin luokitukseen tietue kuuluu, on luokittelu huono
- Hyvän luokittelun tunnusmerkit
  - Oikean luokan valinta on helppoa
  - Tietuetta ei saa sovitettua kuin yhteen luokkaan
- Käytetään usein raportoinnin rajauksina → johdetaanko hyvällä tiedolla

The background of the image consists of a dense, repeating pattern of green numbers (0-9) of varying sizes and orientations, creating a digital or data-like aesthetic. The numbers are semi-transparent and overlap each other, giving a sense of depth and movement.

**Miten huono datan  
laatu näkyy?**

# Laatukustannuksia

- Virheelliset oivallukset
  - Huonolaatuisesta datasta johdetut päätökset voivat olla virheellisiä.
- Taloudelliset menetykset
  - Huonolaatuisen datan pohjalta tehdyt virheelliset investoinnit.
- Organisaation tehokkuus
  - Huonolaatuisen datan pohjalta syntyvät virheelliset hankintaehdotukset tai resursointi.
- Yleiset pullonkaulat toiminnan kehityksessä
  - Huono datan laatu vaikuttaa negatiivisesti toiminnan digitalisoituessa. Uudet järjestelmät tuottavat uusia laatuvaatimuksia datalle.

Missä kura syntyy,  
kuka siivoaa?

# Kuraa syntyy

- Ylivoimaisesti suurin syy on käyttäjät itse
  - Ei ole määritelty laadukasta dataa → eri käyttäjät tuottaa erilaista
  - Järjestelmien pakollisuudet tyypillisesti vain teknisen toimivuuden kannalta
- Epäluotettava lähde esim. ulkopuolisessa datassa
- Siirron tekniset ongelmat
- Vialliset anturit jne.



# Puhdistus

- Käsipelillä kaikkien läpikäynti
  - Esim. Excelillä järjesteele ja tutkii löytyykö puutteita tai virheitä
  - Suoraan järjestelmässä
  - Jatkuva parantaminen
- Isot massat
  - Kumppani apuun, jolla osaaminen ja työkalut

# Puhdistus

- Perustiedot kuntoon heti, koska niitä tarvitaan varmasti
- Muu data tarpeen vaatiessa
- **Tärkeää!** Mahdollisessa järjestelmävaihdossa älä ainakaan kaada vanhoja perustietoja tutkimatta uuteen.

The background of the image consists of a dense, repeating pattern of green numbers (0-9) of varying sizes and orientations, creating a digital or data-like aesthetic. The numbers are scattered across the entire frame, with some appearing larger and more prominent than others.

# Kotiläksy työpajaan osallistuville

# Datan lähteet

- Listaa yrityksesi datan lähteet
  - Voit ottaa tämä esityksen avuksi pohdintaan
  - Älä jumiudu vielä miettimään mitä dataa lähteessä on ja mitä sillä tekisit
  - Älä jumiudu miettimään kuinka sen saisi hyötykäyttöön sieltä
- Bonus: Mieti kaksi kysymystä, joihin haluaisit datasta vastauksen. Unohda tässä vaiheessa mitä lähteitä sait listalle.



# Yhteystiedot

Hubble Oy

Eteläkauppatori 4 A (5. krs), 28100 Pori

Jani Nurmi

050 626 92

[jani.nurmi@hubble.fi](mailto:jani.nurmi@hubble.fi)

