Data-Driven Approaches to Healthcare Research during the COVID-19 Pandemic

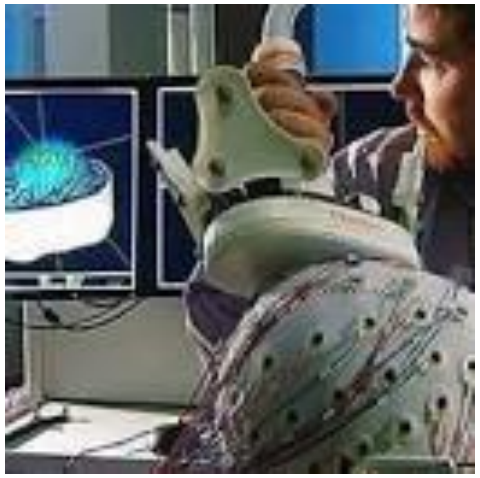
Miika Leminen

Development manager,
Head of Analytics and AI Development
at Helsinki University Hospital (HUS)
June 25, 2020

CS-E4002: Human-Centred Research and Design in Crisis

Aalto University

MIIKA LEMINEN



- **M.Sc. (technology), Aalto university**
  - **Bioelectronics, Medical Engineering, Work psychology & Leadership**
- **M.A. (psychology), Licensed psychologist, University of Helsinki**
- **Finishing PhD studies: neuropsychology and cognitive neuroscience in children with language development difficulties**
- **Laboratory engineer (Brain research), Researcher, Data scientist**
- **26 peer-reviewer scientific papers, Google Scholar h-index 13**
- **Teached over 12 courses and 20+ lectures in HY, Aalto, China, and Denmark**

AGENDA

1. AI & Analytics development @ HUS University Hospital
2. Clinical and other Data in Covid-19 pandemic
   - Also collaboration, data ownership, and legal issues
3. ML application development in health care
4. HUS Data Lake
5. Scenario models in Covid-19

AI AND ANALYTICS DEVELOPMENT / HUS IT

HUS Data Lake

Data service

ML application DEV

ML application PROD

Custom analytics

IT for research and teaching

HUS THANKS ALSO PRO BONO TROOPS

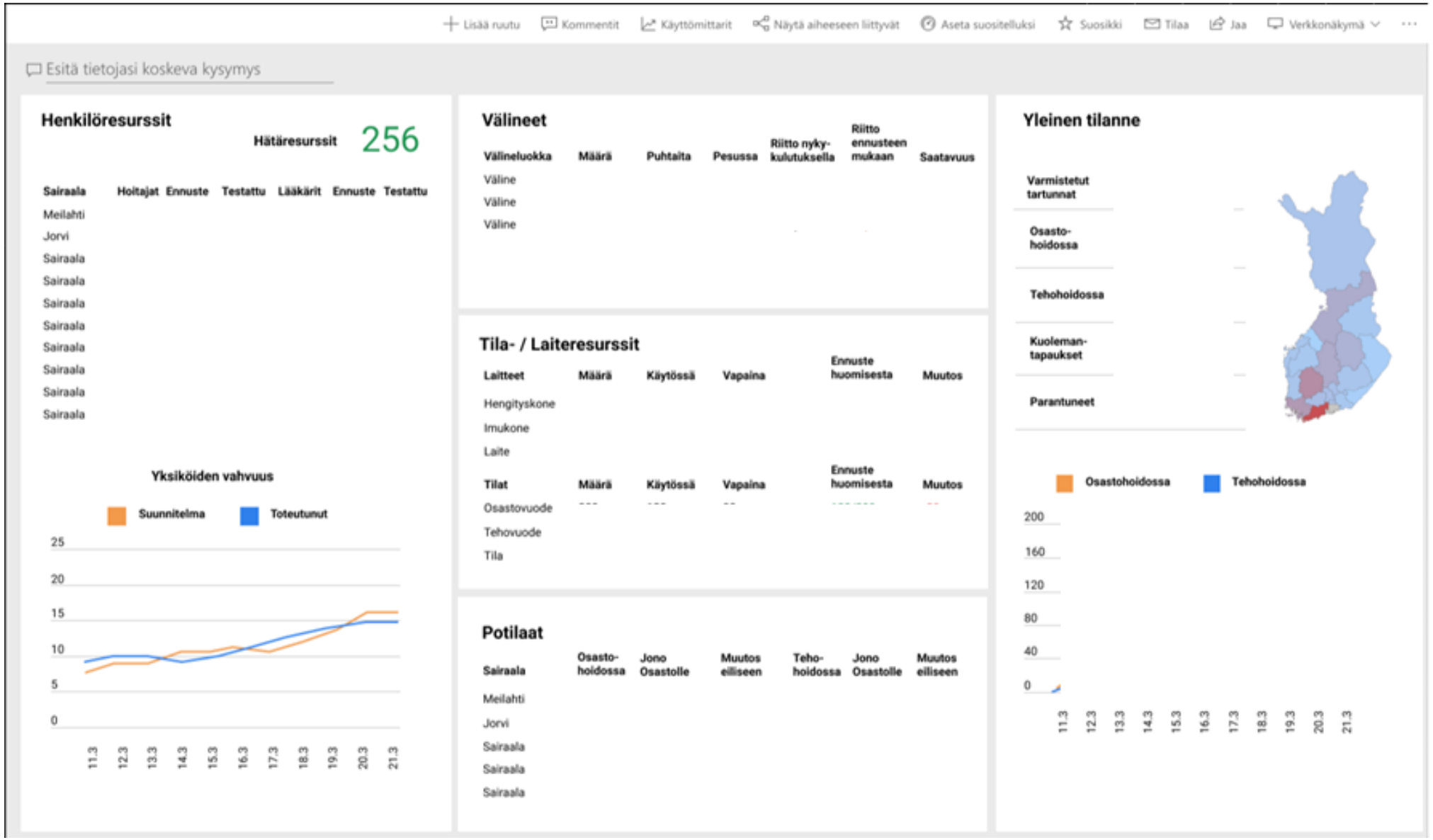
BCG BOSTON CONSULTING GROUP

CLINICAL DATA IN COVID-19 PANDEMIC

- **Diagnoses**: challenge especially in the beginning of the pandemia
- **Laboratory results**: how to take delay in the results into account, and what number is reliable when the test population varies on daily basis
- **Bed count**: define the ICU, define the confirmed case criteria
- **Weak signals**: e.g., contacts in päivystys-apu and other early medical support channels, chatbot activity
- Analyses of effects from other countries (**public data sources**)

CASE MOBILITY: DID YOU STAY HOME

- **Elisa provided mobility data from it's mobile networks**
- **Aggregated, population level data, based on contracts**
- **NOT GPS data from anywhere, but based on visits to link towers**
  - **E.g. in post code level, where population in one specific area usually visit/travel whithin Helsinki (might indicate areas with higher pandemic risk)**
  - **E.g. in population level, how quickly traveling to/from Uusimaa returnes to baseline level after the lock-up**
- **These measures are less usefull when pandemic is not "wildly distributing in population" anymore**
- **Similar information available publicly, e.g. Apple or Google map use**


Lisää ruutu
Kommentit
Käyttömittarit
Näytä aiheeseen liittyvät
Aseta suositelluksi
Suosikki
Tilaa
Jaa
Verkkonäkymä
Esitä tietojasi koskeva kysymys
Henkilöresurssit
Hätäresurssit
256
Sairaala
Hoitajat
Ennuste
Testattu
Lääkärit
Ennuste
Testattu
Meilahti
Jorvi
Sairaala
Yksiköiden vahvuus
Suunnitelma
Toteutunut
25
20
15
10
5
0
11.3
12.3
13.3
14.3
15.3
16.3
17.3
18.3
19.3
20.3
21.3
Välineet
Välineluokka
Määrä
Puhtaita
Pesussa
Riitto nyky-kulutuksella
Riitto ennusteen mukaan
Saatavuus
Väline
Tila- / Laiteresurssit
Laitteet
Määrä
Käytössä
Vapaina
Ennuste huomisesta
Muutos
Hengityskone
Imukone
Laite
Tilat
Määrä
Käytössä
Vapaina
Ennuste huomisesta
Muutos
Osastovuode
Tehovuode
Tila
Potilaat
Sairaala
Osasto-hoidossa
Jono Osastolle
Muutos eiliseen
Teho-hoidossa
Jono Osastolle
Muutos eiliseen
Meilahti
Jorvi
Sairaala
Yleinen tilanne
Varmistetut tartunnat
Osasto-hoidossa
Tehohoidossa
Kuoleman-tapaukset
Parantuneet
Osastohoidossa
Tehohoidossa
200
160
120
80
40
0
11.3
12.3
13.3
14.3
15.3
16.3
17.3
18.3
19.3
20.3
21.3

COLLABORATION AND PERMISSIONS

COLLABORATION & DATA

Secondary use of data & pandemic

Who can use what data and under what law? What is the line between **primary** and **secondary** use?

Collaboration with City of Helsinki and other HUS area municipalities in covid-19

Daily collaboration, ready-made integrations between datalakes, collaboration in scenario modeling

Always do things right!

WHEN IS CONSENT NEEDED

Primary use

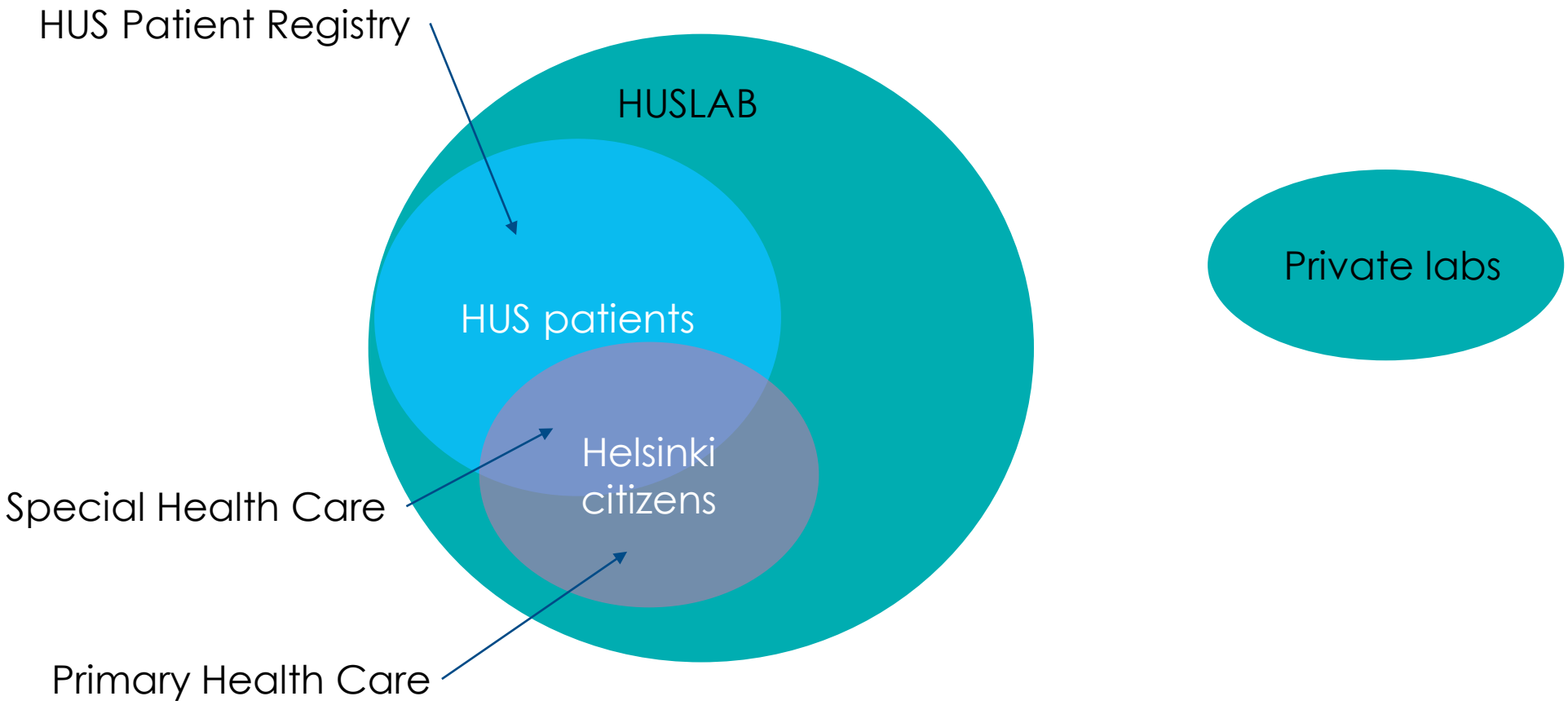
- **Clinical care of individual patient: all data can be used**
- **Data based management of healthcare**

Secondary use of healthcare data (Ns. toisiolaki)

- **new law gives right for registry owner to give a data permit without a consent**
- **Research, teaching, development…**

Clinical research: patient is contacted and/or study affects to care and/or new data (only because of current study) is being collected

HEALTH REGISTRY: WHO OWNS THE DATA



HOW TO GET ACCESS TO DATA (SIMPLIFIED)

If HUS data is enough

- **Info about data in Data/information service: https://www.hus.fi/tutkijalle/tietoallas**
- **If you work at HUS and/or research is being made in HUS: https://www.hus.fi/tutkijalle/tutkimuslupa**
- **If you work outside of HUS and/or other use than research: ”Tietolupa” is enough, contact Data service**

If you want to combine registries

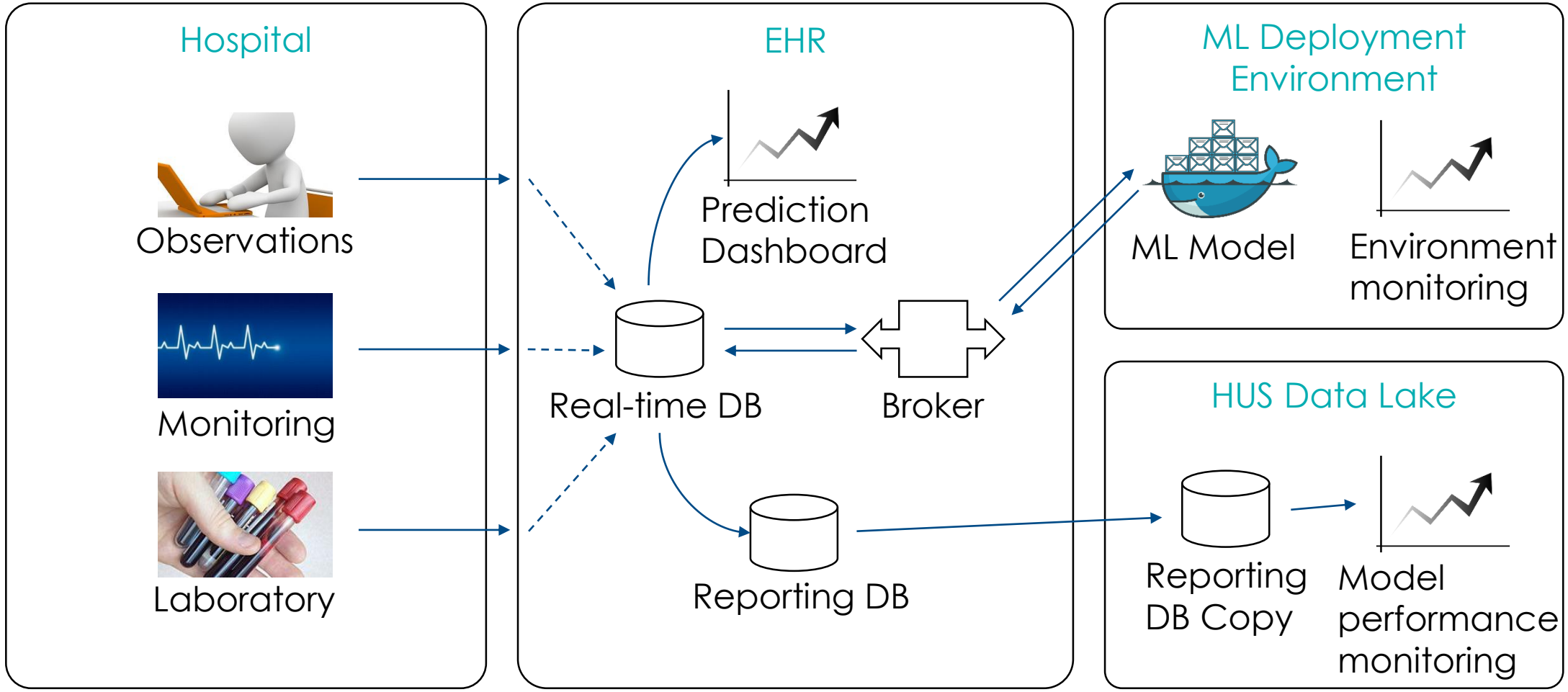
- **Findata/THL: https://www.findata.fi/**

COLLABORATION WITHIN HUS

- **Data Scientists, Data Analysts, BI developers, ML engineering**
  - **Medical images, EHR specialities, monitor signals, business insight**
- **IT specialists**
  - **Application architecture and development (back and front end)**
  - **Data architect, data integration specialists, security, cloud services**
- **Project managers**
- **Physicians, Nurses, Clinical Managers**

ML APPLICATION DEVELOPMENT IN HEALTH CARE

SAMPLE MODEL DEPLOYMENT



WHERE IT ALL BEGINS

Clinical process

What would benefit most?

Where, when, what, and how?

Existing knowledge

Data

Evaluation matrix
& prioritizing

MORE APPLICATION RELATED QUESTIONS

Related applications, messaging architecture, data pipelines?

What event should trigger the application?

What kind of information would be most useful? What is the most time consuming phase?
What cases should be detected and taken to the short list?
How new information could change/improve the clinical pipeline?

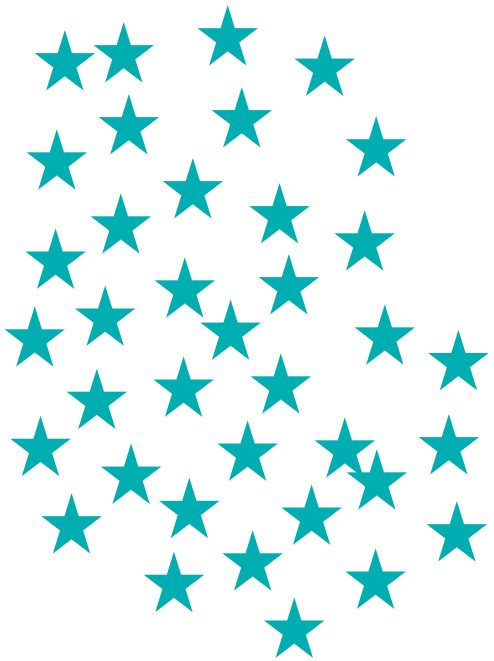
What kind of information would be needed for techincal/clinical validation?

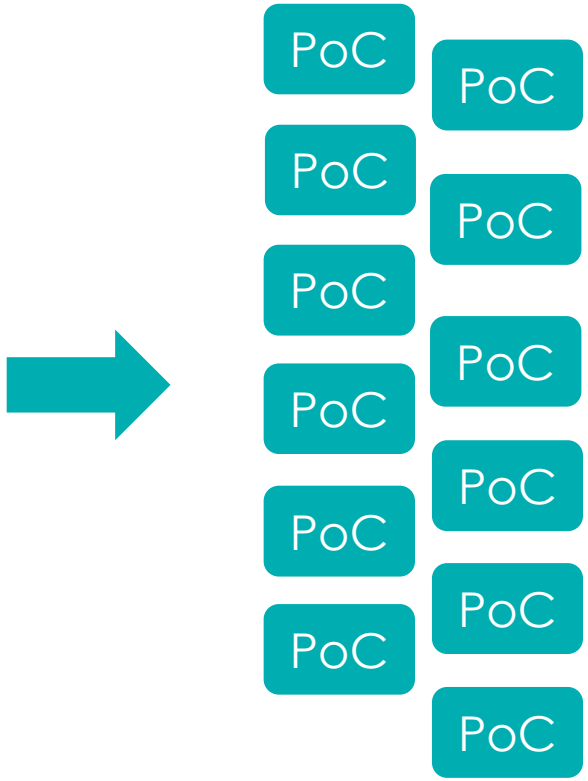
Criteria for clinical acceptance? Risks?

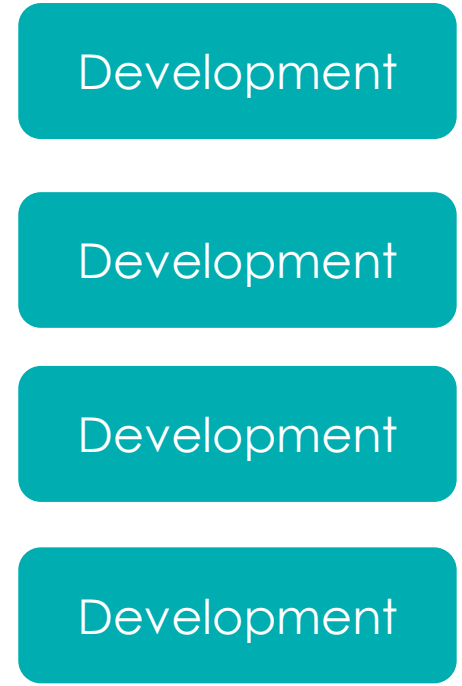
PROCESS

Development

PROCESS




PoC
PoC
PoC
PoC
PoC
PoC
PoC
PoC
PoC
PoC
PoC
PoC


Development
Development
Development
Development

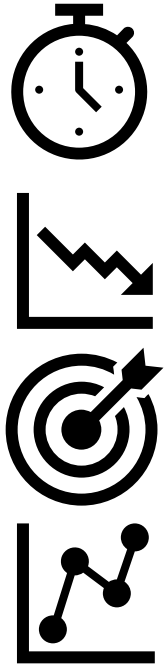
ESTIMATING CLINICAL PROFIT

Saved working time

Reducing complications

Resources can be focused on high risk group

New/additional/earlier info



constellationkidney.com

EVALUATION MATRIX

Clinical personnel and their resources: IT people can not work alone

Availability of technical tricks / technical risks

Cost benefit analysis

Benchmarking / research knowledge

Clinical need / motivation

CLINICAL VALIDATION AND CE PROCESS

Creating proper validation methods

Standard risk evaluations

Multi-professional clinical team

constellationkidney.com

Sertifikaatti
Nro 10349-01

kiwa

CERTIFICATE

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

Helsingin ja Uudenmaan sairaanhoitopiirin kuntayhtymä
HUS Tietohallinto
Virtuaalitiimi
Helsinki

laatujärjestelmä 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 laatujärjestelmä
täyttää jatkuvasti edellä mainitun standardin ja yleisen ohjeen ABC 200 vaatimukset.
Sertifikaatin voimassaolon voi tarkistaa osoitteesta www.inspecta.fi

Inspecta Sertifiointi Oy
P.O. Box 1000,
Dömölstenkatu 2
FI-00581 Helsinki,
Finland
Tel. +358 10 521 600

IQNet

kiwa certified ISO 13485

Inspecta Sertifiointi Oy

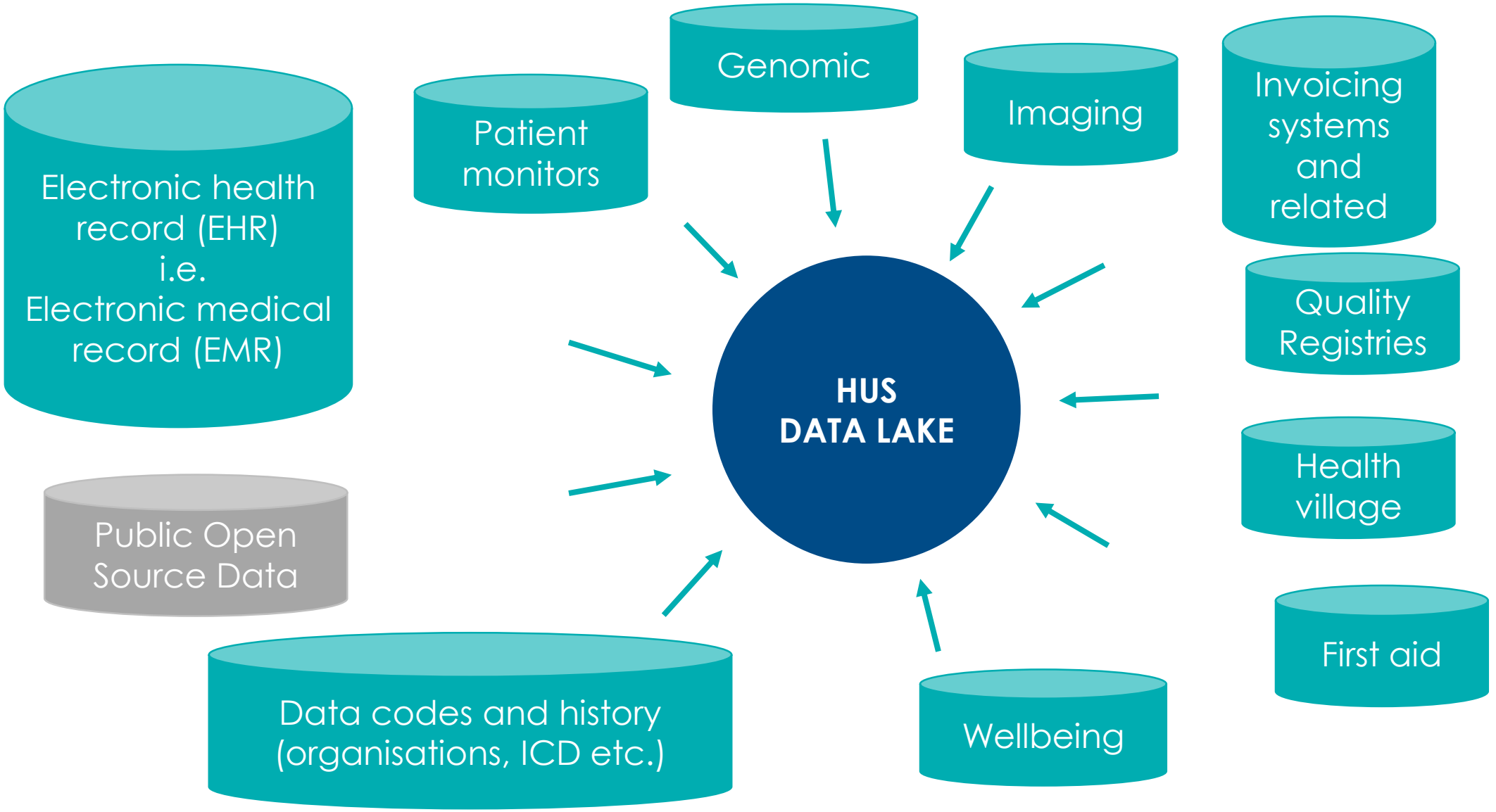
HUS DATA LAKE

DATA LAKE: PURPOSE AND BACKGROUND

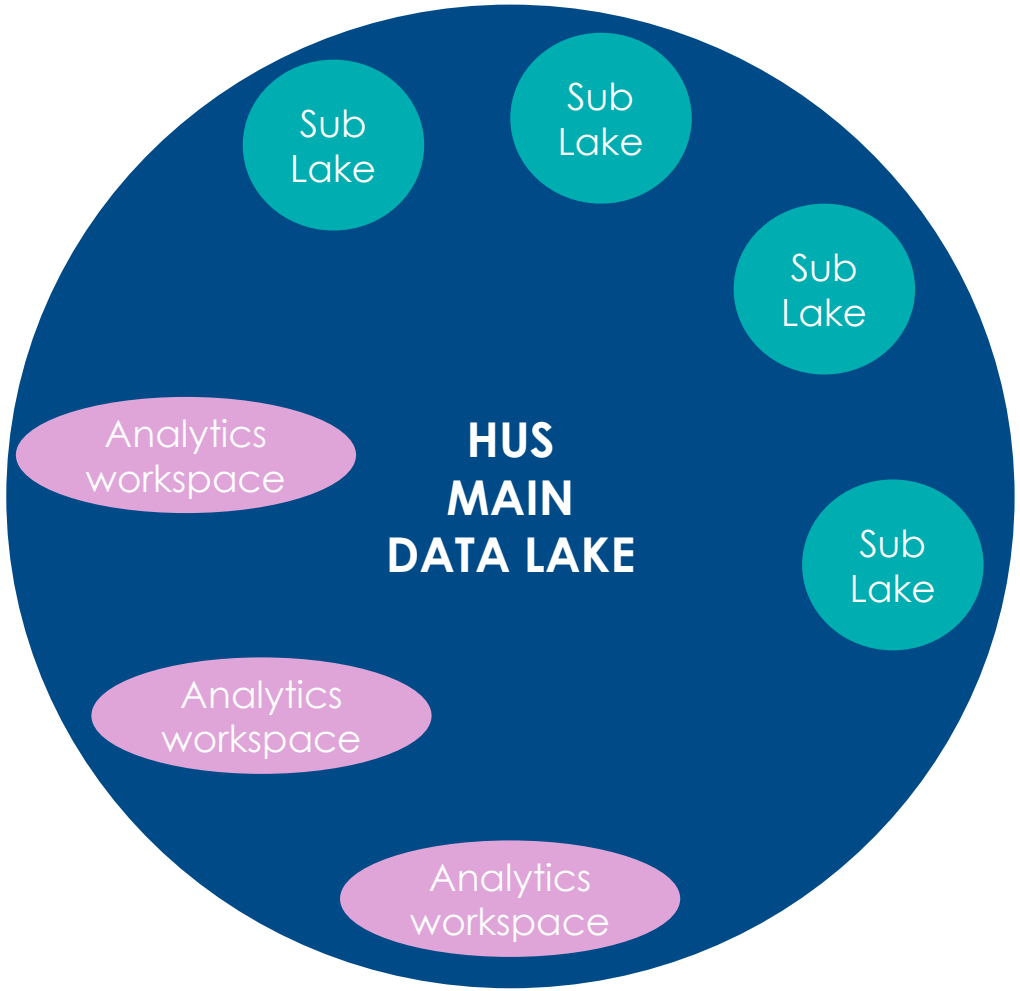
- Project started in 2015 (Sitra's ISAACUS project). Production 2018.
- Aimed to enhance secure usage of the Finnish health data for various purposes:
  - Clinical use (data availability and sharing for applications)
  - Research
  - Statistics
  - Authority/official guidance and surveillance
  - Knowledge management
  - Business intelligence & intellectual capital management
  - Development and Innovation functions
  - Staging area for data warehouse

HUS DATA LAKE - DATA SOURCES

Real-time integrations, overnight batch loads, historical one-time.



HUS DATA LAKE ECOSYSTEM



- **Main data lake**
  - Takes care of data integration and common data enrichment processes and permanent storage.
  - AI team working area.

- **'Sub Lakes'** and **'Analytics Workspaces'** are for specific research projects (permission based, for limited time, and with independent pseudonymization):
  - HUS-users and external users are separated to their own network segments.
  - Access is role based: depends on research permit.
  - Access for external users via VPN or direct connection.
  - No data is transferred outside HUS network.

- HUS Data Lake is based on Azure, which offers a modern scalable minicluster environment
- Analytics workspaces available also in GCP, AWS
- CE-certified (class 1) Medical Device

DATA VOLUMES, EXAMPLES

Annually

Unique patients 500 000
Hospital visist (käynnit) 3 000 000
Inpatient / Ward visits (osastohoitojaksot) 250 000
Babies delivered 16 000
Surgeries 92 000
Personnel 25 000

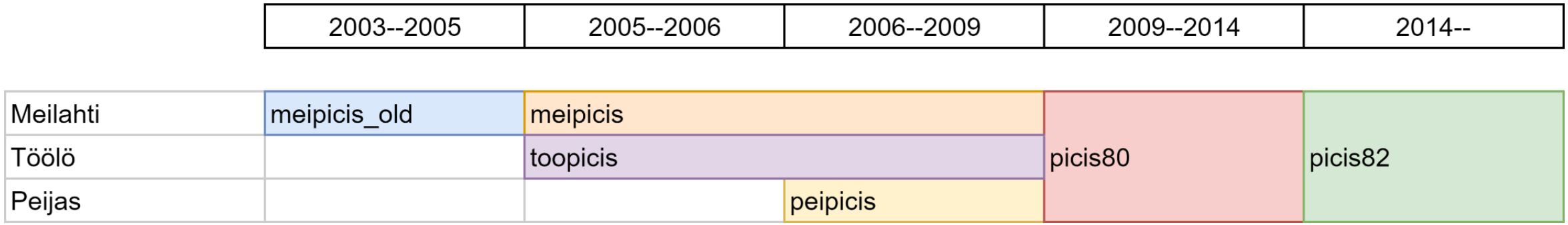
Total

Labotary tests 780 000 000
Medical case records (Potilaskertomus) 80 000 000

EHR DATA COMPLEXITY EXAMPLE: INTENSIVE CARE

- At present, two systems are used simultaneously: Caresuite & Clinisoft
- In future, both are replaced with Apotti.
- Looking just Caresuite (=Picis), there has been multiple versions of the product.

→ data discontinuity, terminology, data models…

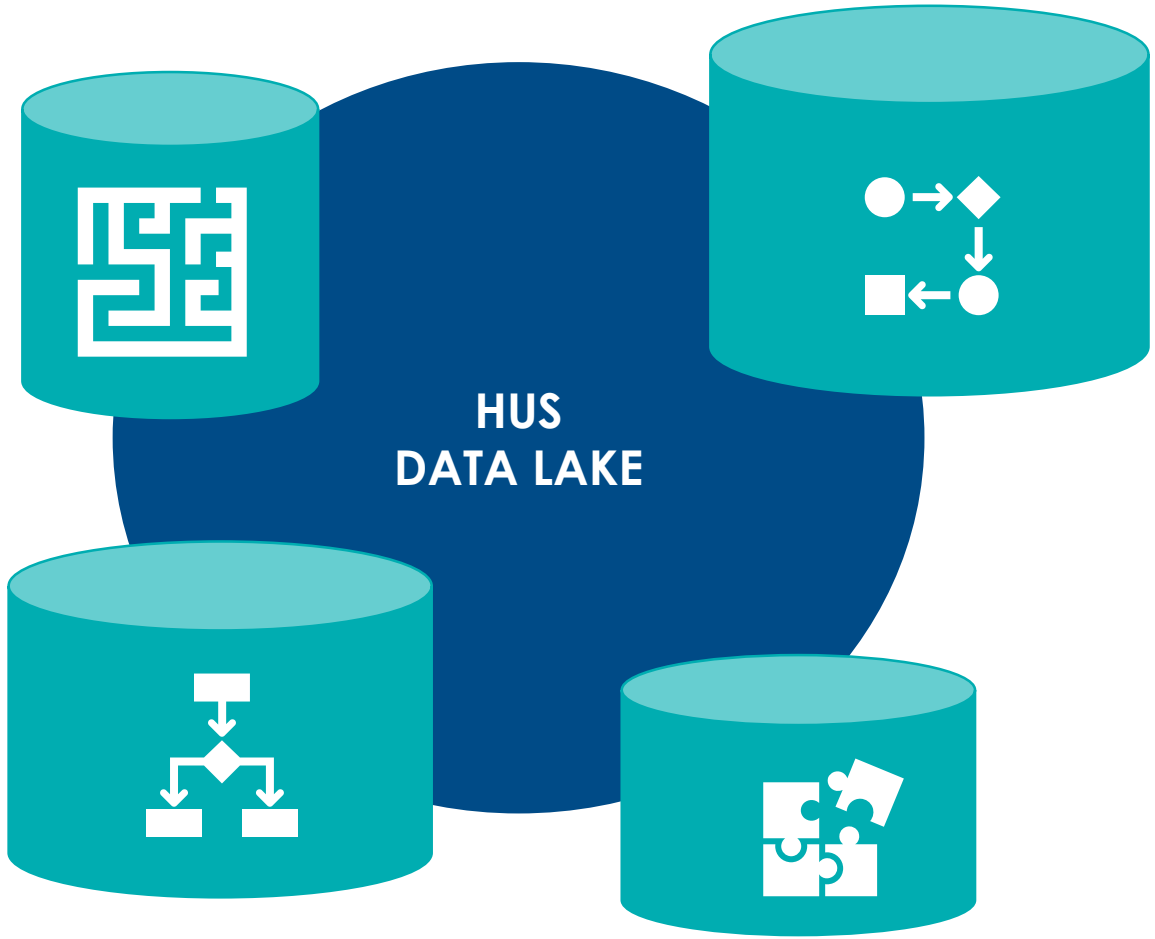


DATA AVAILABILITY BY SOURCE

Data Source	Data Content	Type	Availability
Lab and pathology	Labs, pathology, orders and results	Lab mostly structured, Path results text	1980's →
Disease-specific quality registers	Data specific to each disease, incl. PROMs	Mostly structured	Implementation of each register varies, earliest data late 1990's
EHR (Uranus)	PAS and EHR, medications etc.	Structured + text	Current EHR 2006/2010 →, previous data in archive (Musti)
DW	HUS data warehouse, operative and financial data	Structured	2000 →
RIS	Radiology ERP, orders, metadata etc.	Structured + text (results)	Late 1990's →
PACS	Images retrieved from archiving system, based on metadata in DL	Images	Approx. 2003 →
Muse	Digital ECG archive	Images	2008 →
OR	Operating room system	Structured + text	2005 →
ED & ICU	Emergency department and ICU data	Structured + text	1999 (partial) →
Obgyn	Pregnancy and childbirth	Structured + text	2005 →
Genomics	Will be available from FinnGen-study	Genomics data	Available 12/2019

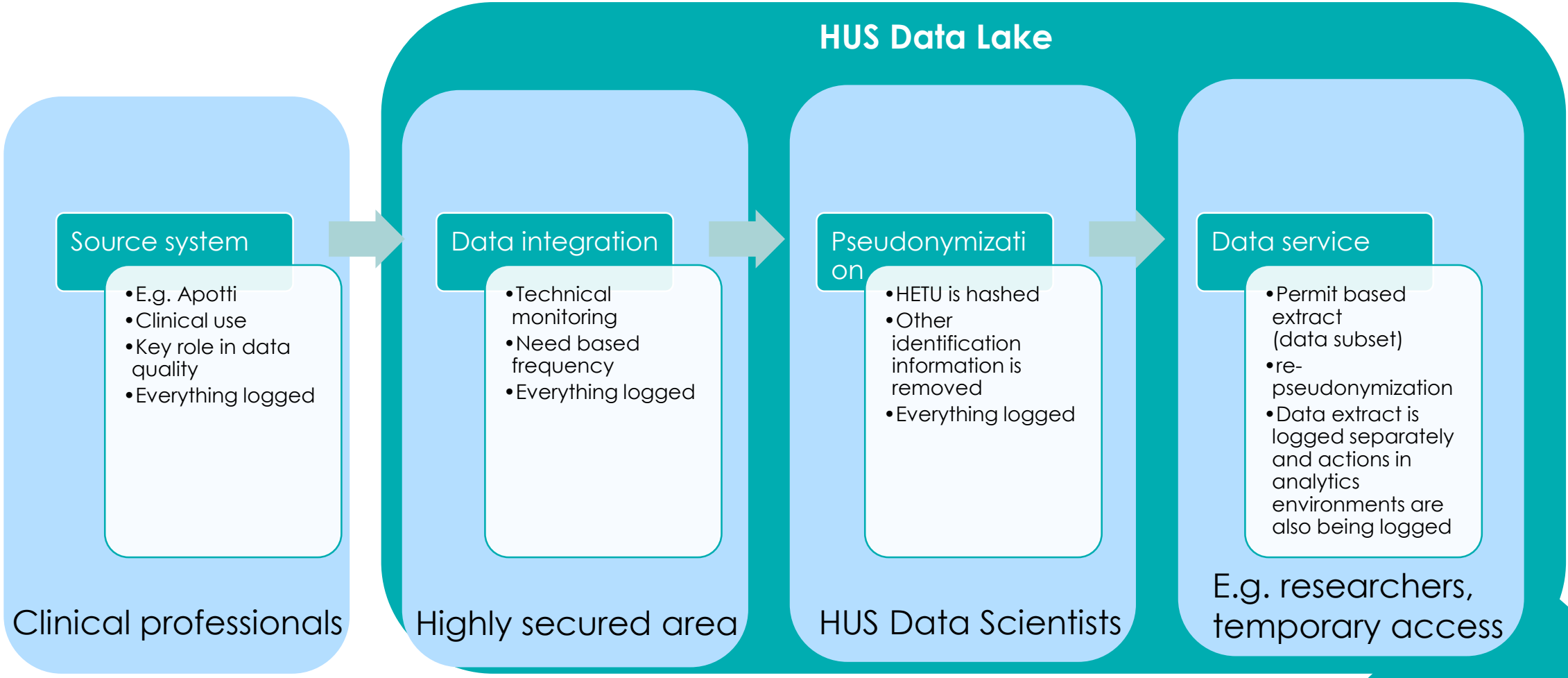
Tentative availability, coverage varies

DATA



- >100 Data bases, each very different with their data structure.
- 20-20 000 Tables per data base.
- Often original copy from source system (not for humans).

DATA PROTECTION PIPELINE



SCENARIO MODELS

SCENARIO MODELS

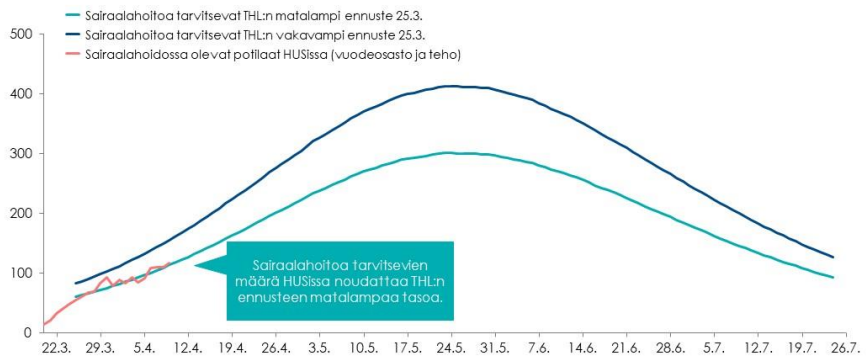
Bayesian SEIR:
Susceptible, Exposed,
Infectious, Recovered

Markov Chain: Home,
Dept, ICU, death.
Adapting to real situation

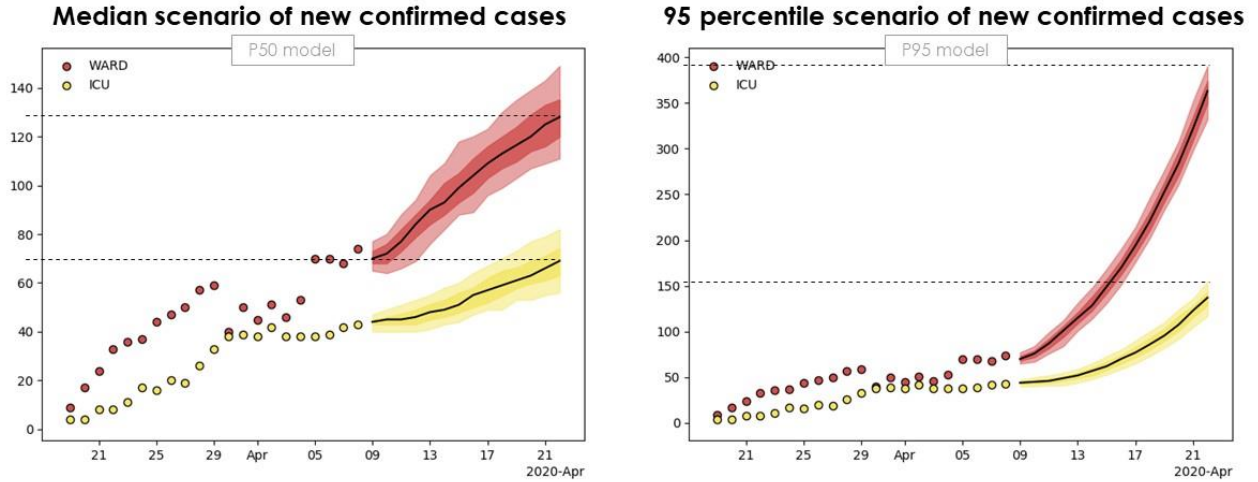
Time varying R
(Check e.g. an Oxford index for more information)

PRESS RELEASE APRIL 13TH

THL:N ENNUSTEET SAIRAALAHOIDON TARPEESTA JA TOTEUMA HUSISSA



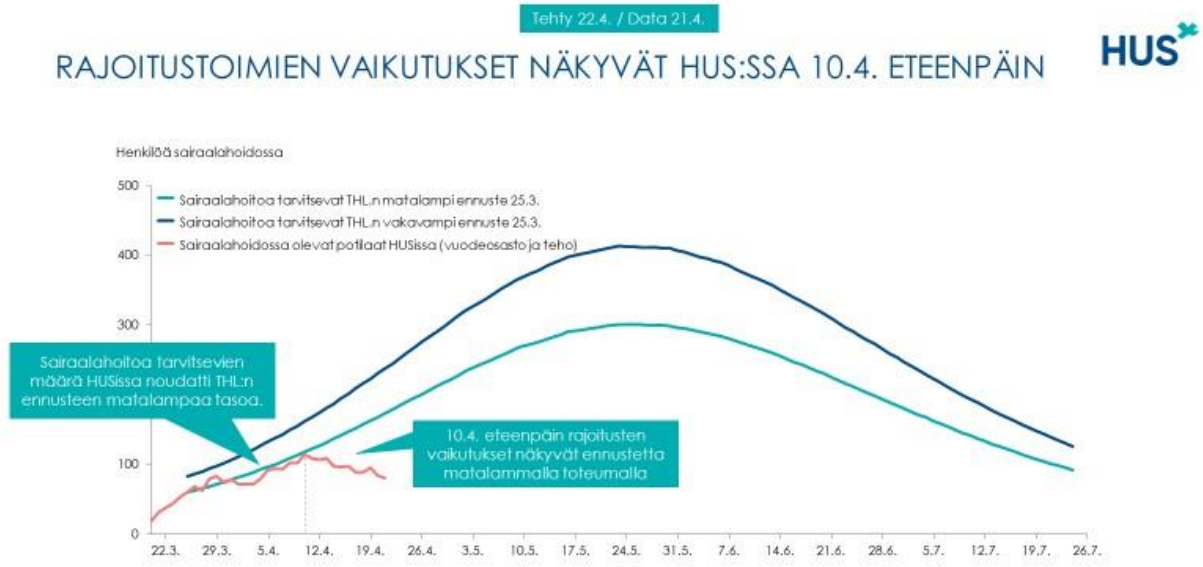
SAIRAALAHOITOA VAATIVIEN POTILAIDEN MÄÄRÄN ENNUSTETAAN KASVAVAN TASAISESTI

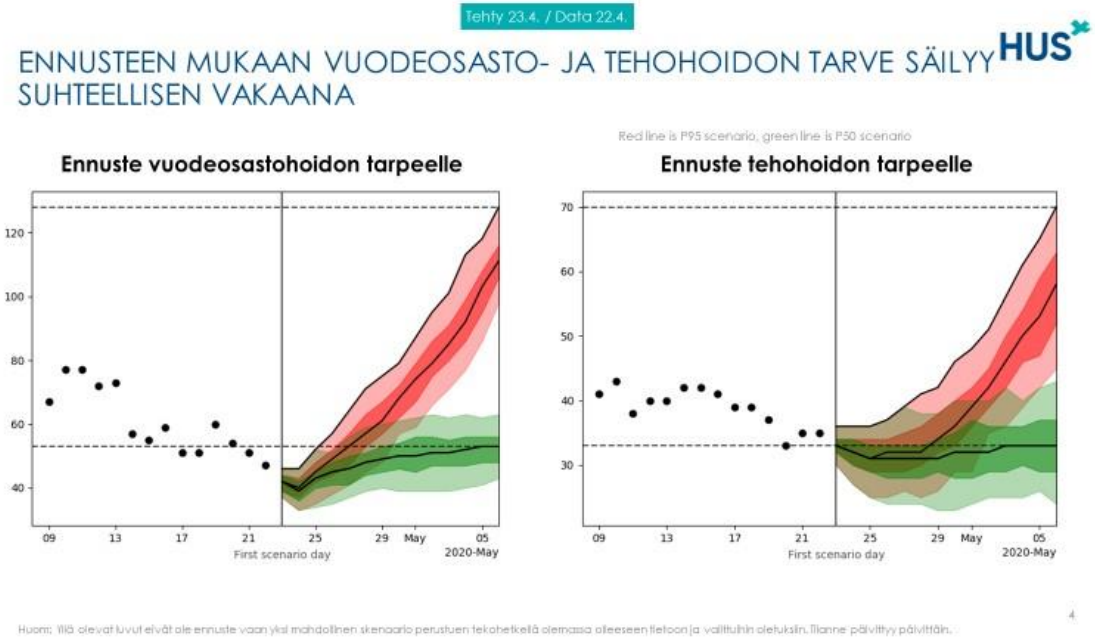


Pisteet ovat toteutuneita määriä, vaaleammalla värillä on merkitty 90 % vaihteluväli, tummemmalla värillä 50 % vaihteluväli.

https://www.hus.fi/hus-tietoa/uutishuone/Sivut/Mallinnuksia-hyödynnetään-koronapotilaiden-hoitoon-varautumisessa.aspx

PRESS RELEASE APRIL 24TH





https://www.hus.fi/hus-tietoa/uutishuone/Sivut/HUS-ennustemalli-potilaspaikat-riittävät-nykyisillä-rajoituksilla.aspx

TIPS FOR THE FUTURE

- **Get going, do not wait for complete and perfect data**
- **Fine-tune your model to be robust against error in estimated parameters**
- **Use parallel strategies (instead of one ”perfect model architecture”) and understand where their differences come from**
- **Improve your model while you get better data**
- **In practice: short term is more important!**

LESSONS LEARNED – READINESS FOR THE NEXT WAVES

Data pipelines and quality control	• From ”monthly reporting” to ”real time” • Weak signals, custom data streams
Collaboration networks	• Both human and digital. Law is the tool.
Scenario modeling strategies	• From long term (”more uncertainty”) to ”short term” (”operative”)
Choose your fight	• From an ”interesting AI study” to what is the ”most useful” for the organization/health

THANK YOU!

Miika Leminen

Development Manager, Head of Analytics and AI development services unit

miika.leminen@hus.fi

Participatory Workshop tomorrow
June 26, 10 - 12

Rethinking Critical Crisis Themes and Design Directions

Prof. Nitin Sawhney

Department of Computer Science, Aalto University

Dr. Salu Ylirisku

Design Teacher (Senior University Lecturer) at Aalto University

Aalto University
School of Science