Services for Sensitive Data (TSD)

Gard Thomassen, PhD
Division for Research Computing Services
Assistant Director, University Centre for Information Technology (USIT)
University of Oslo

Todays topics ☺

- USIT
- Data – GDPR – Protection/security – Research
- What is TSD
- Building high security systems for multiple institutions
- TSD system design
- TSD organisation
- What do researchers do inside TSD
- Questions

USIT

- https://www.usit.uio.no/om/

- Student services
- Education support
- Research support
- Information and smaller (non credit) courses
- Administrative systems
- ~ 210 employees

Your relation to personal data?

1. Do you use social media?

Your relation to personal data?

2. Do you think about what you write?

Your relation to personal data?

3. Do you share pictures of yourself?

Your relation to personal data?

4. Do you share pictures of others?

Your relation to personal data?

5. Would I google you be before a job interview ?

Laws

- Lov av 14. april 2000 nr. 31 om behandling av personopplysninger (popplyl.)
- Forskrift av 15. desember 2000 nr. 1265 (personopplysningsforskriften)
- Lov av 18. mai 2001 nr. 24 om helseregistre og behandling av helseopplysninger (hlsregl.)

-> and then there was GDPR

SPECIAL CATEGORIES OF PERSONAL DATA

PERSONAL DATA REVEALING INFORMATION REGARDING RACIAL OR ETHNIC ORIGIN, POLITICAL OPINIONS, RELIGIOUS OR PHILOSOPHICAL BELIEFS, TRADE-UNION MEMBERSHIP, DATA CONCERNING HEALTH, SEX LIFE.

UiO – TSD

- Why?
- What
- Technical implementation

Why build TSD

- Offline laptops and desktops too small
- Increased focus on information security
- GDPR
- Visionary leaders at USIT

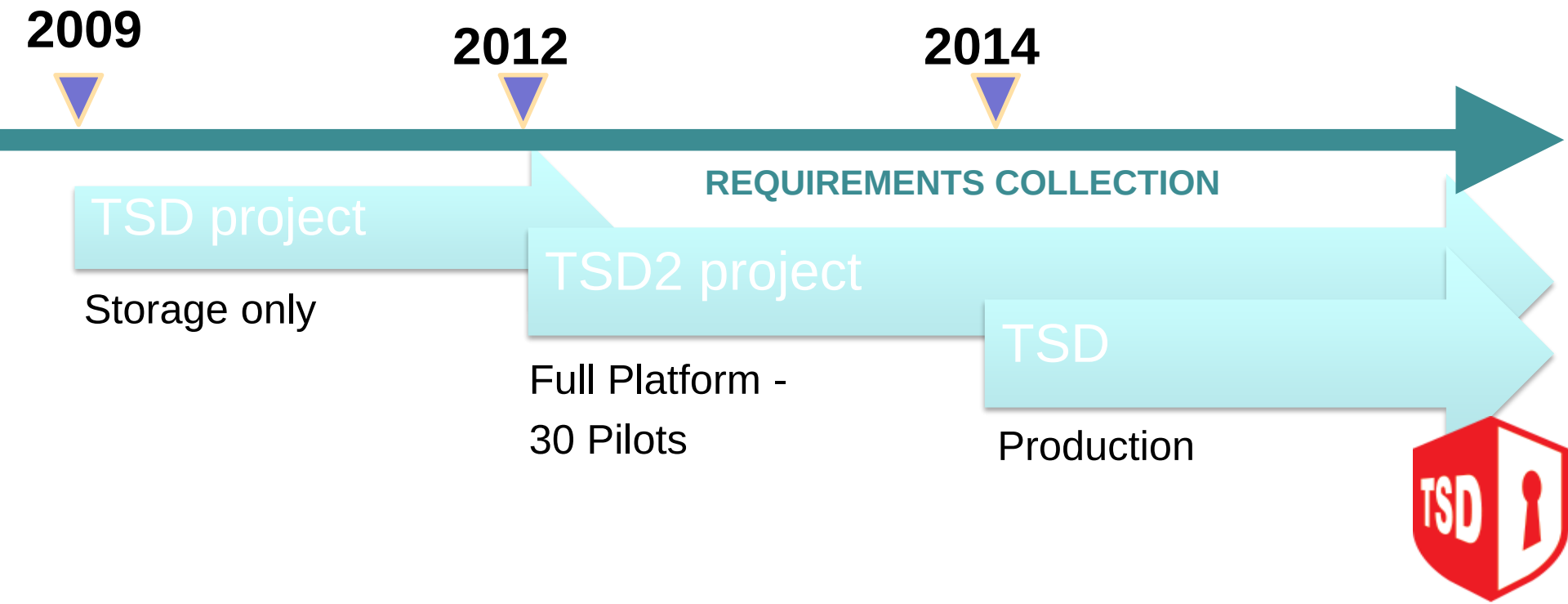
- Most important:

  To remained trusted by all research objects !!

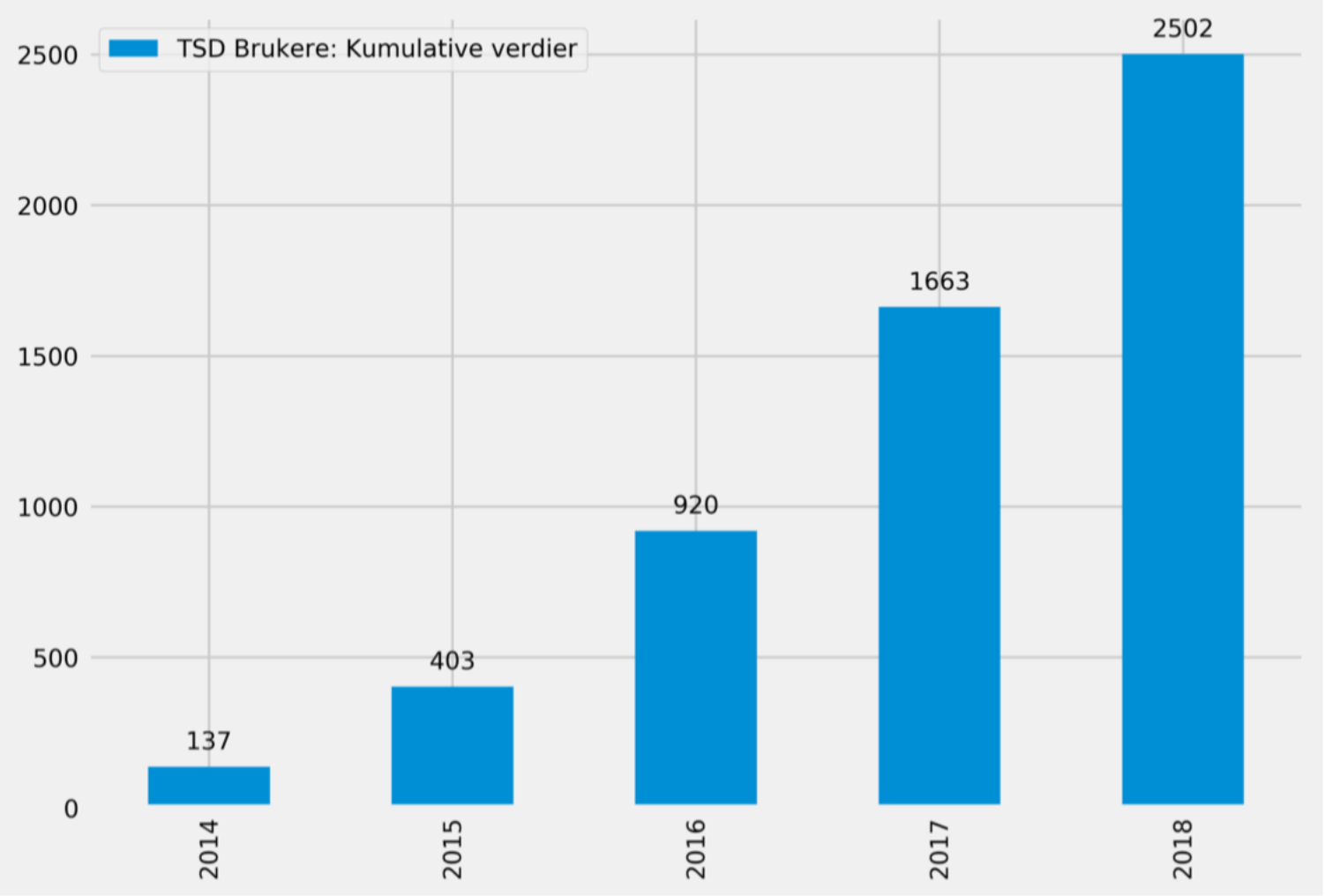
TSD was made for

- Directly identifiable data
- Pseudoanonymous data where there is a high risk of reidentification
- Data where the «data key» is known to the researcher
- Any other data that must be kept safe (armed forces, weapon industry, oil etc)

TSD – FROM PROJECT TO PRODUCTION SERVICE

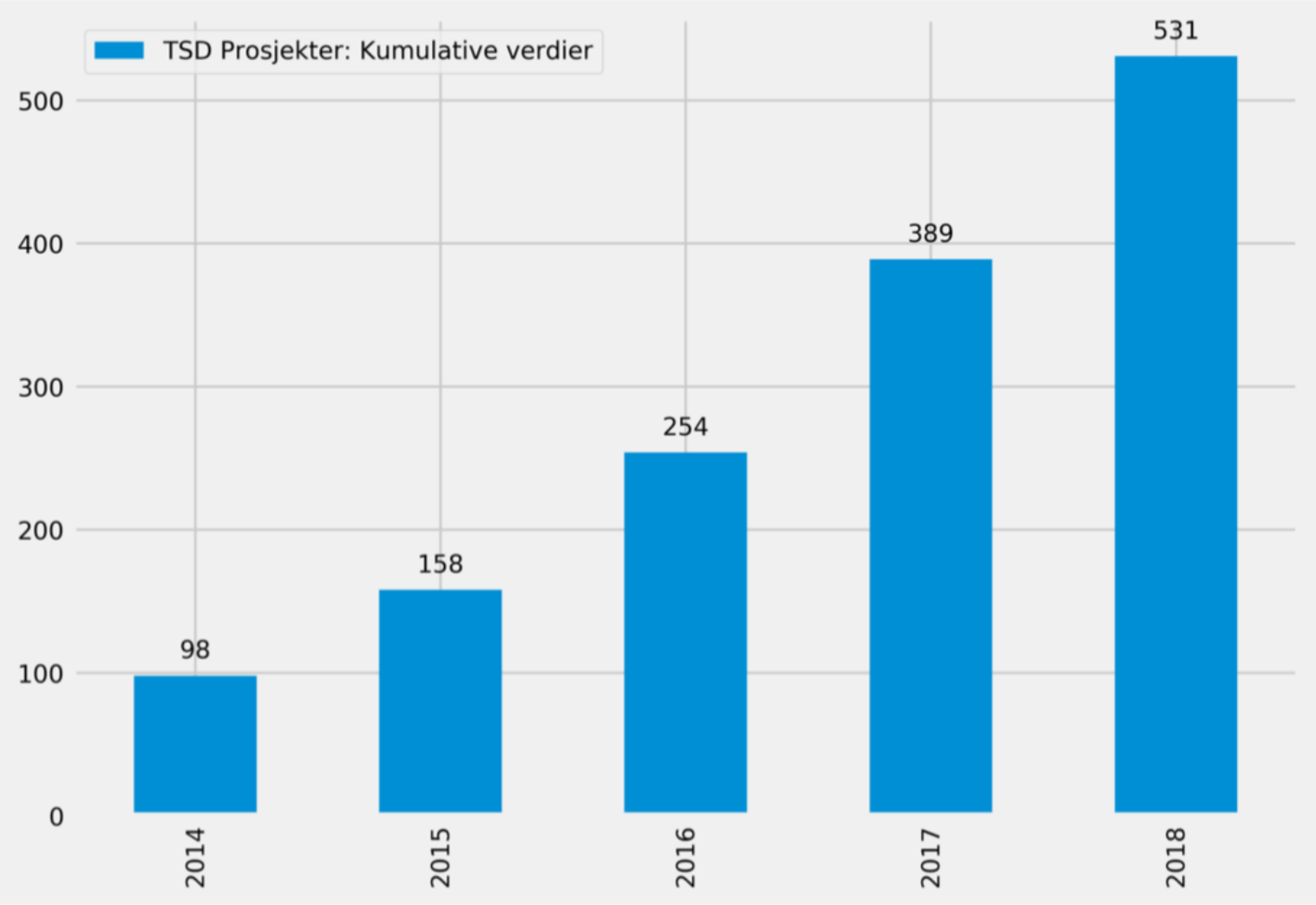


TSD



TSD i vekst

TSD



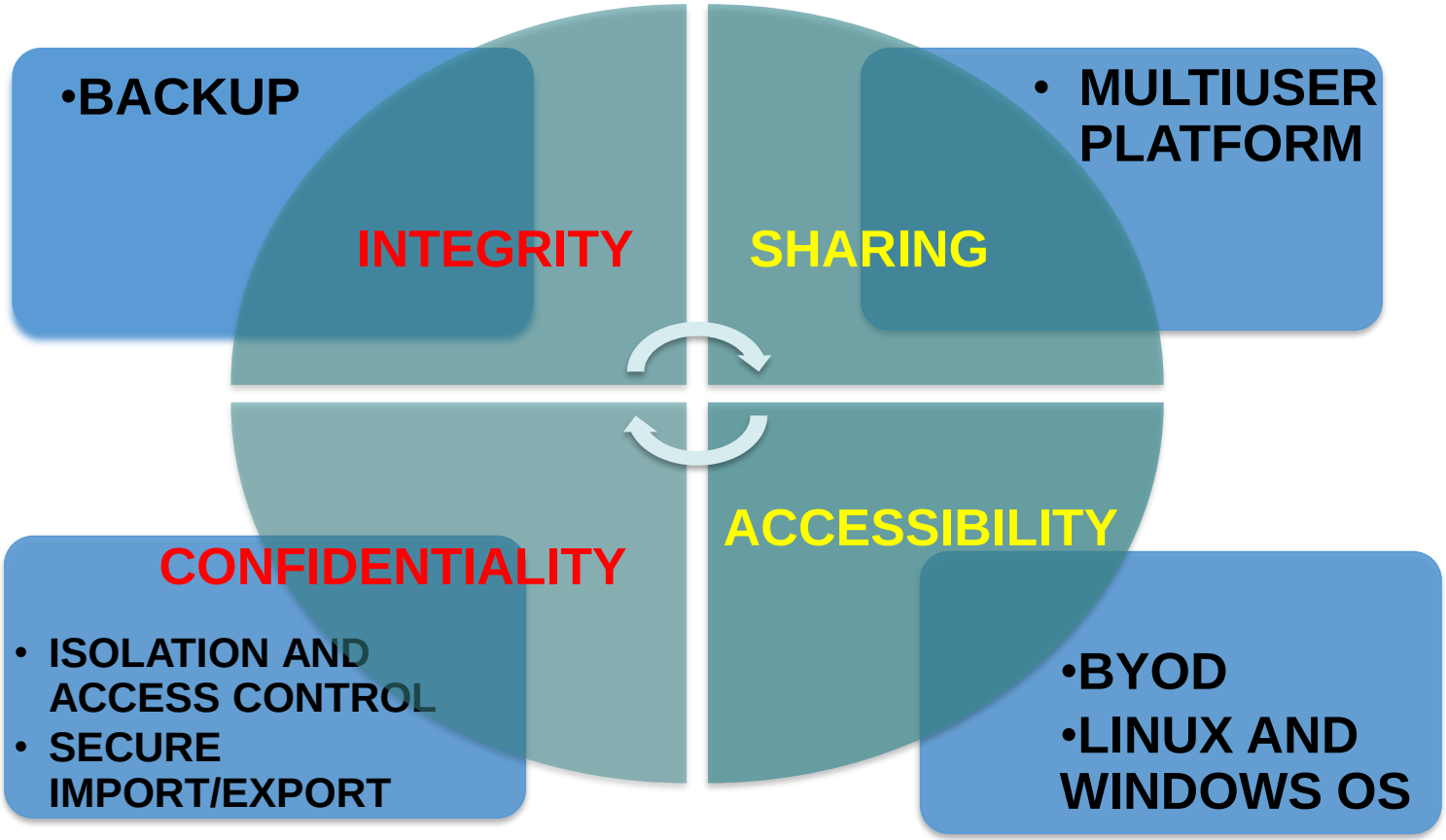
TSD i vekst

TSD tech development – how and why ?


Use-case driven design
TSD

Requirements

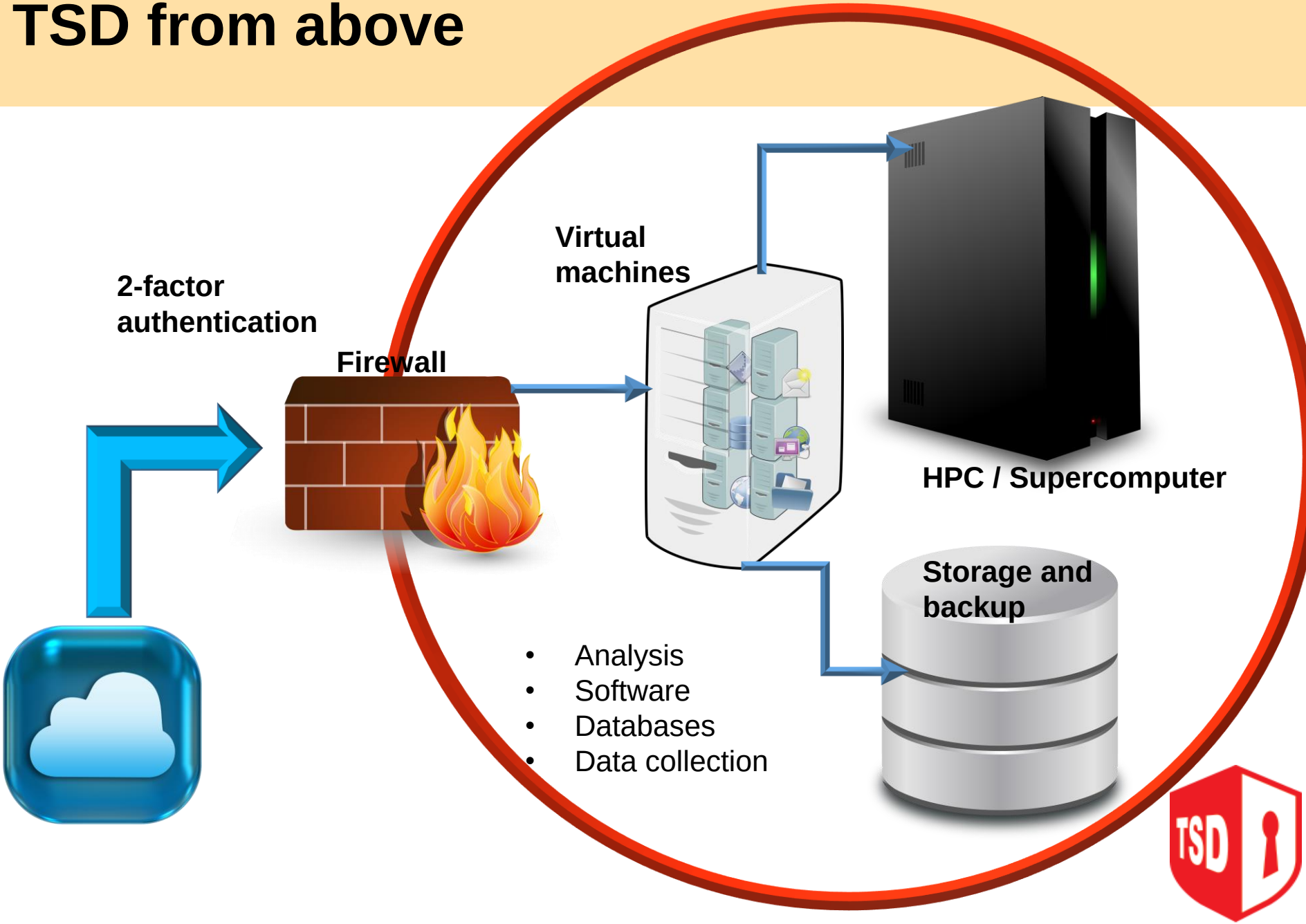
SYSTEM REQUIREMENTS



System requirements

- IT-security, isolation and access control as defined by law
- Unlimited storage
- Multi tenant, multi user environment
- Supercomputing capacity
- High bandwidth
- Affordable operations and maintenance
- Easy to use (including video and sound)
- Freedom in “user-space”
- BYOD – accessible from anywhere from any device
- Make software and reference data available
- Flexibility for new wishes and demands
- Windows and Linux support on the TSD side
- Data collection
- Data sharing, foreign users etc

TSD from above

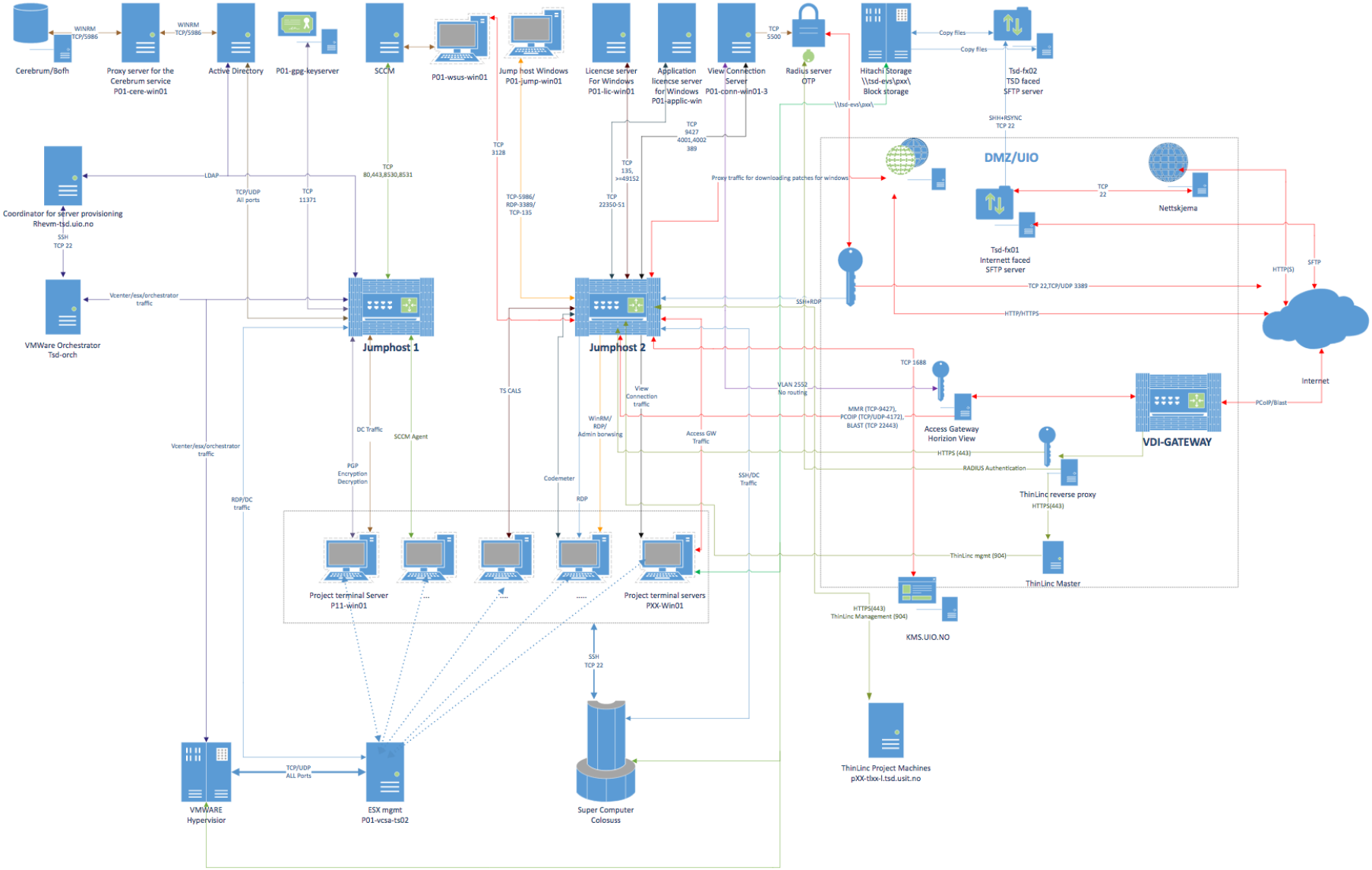


Listen to the end-users; help them!

- ½ PiB genomics data from Iceland – store it!
- We need to run containers – fix it!
- We want Dragen -> well… buy Dragen !
- When in trouble – admit it!

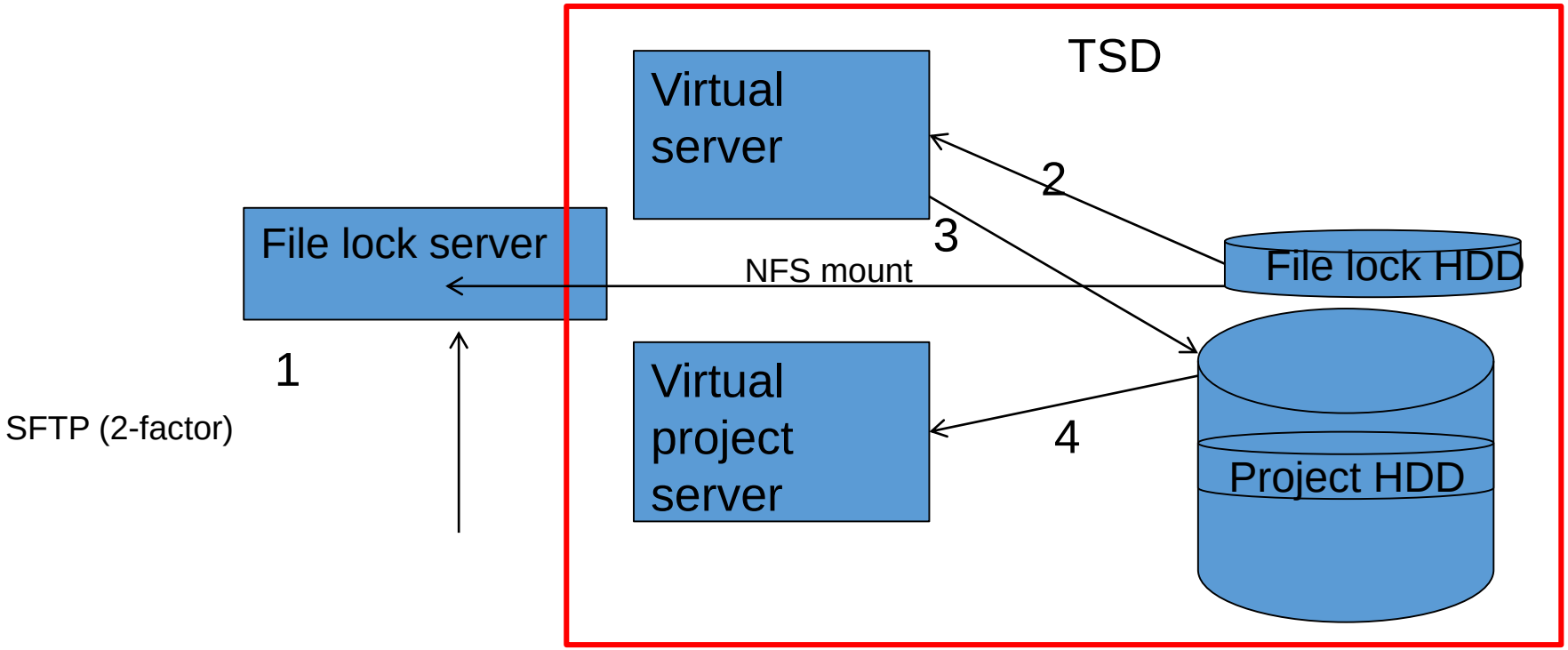
Basic architecture

- Central redundant BSD servers for routing and firewall
- VMware infrastructure for virtual machines
- VLAN-separation between research projects (moving towards microsegmentation)
- TSD provided user admin
- TSD provided everything, no dependencies outside of the firewall”
- Automated provisioning of servers, firewalls and software
- HPC system (Supercomputer and parallel storage)
- Storage (Hitachi Data Systems)


Cerebrum/Bofh
WINRM TCP/5986
Proxy server for the Cerebrum service P01-cere-win01
WINRM TCP/5986
Active Directory
P01-gpg-keyserver
SCCM
P01-wsus-win01
Jump host Windows P01-jump-win01
Licence server For Windows P01-lic-win01
Application licencse server for Windows P01-applic-win
View Connection Server P01-conn-win01-3
TCP 5500
Radius server OTP
Hitachi Storage \\tsd-evs\pxx\ Block storage
Copy files
Copy files
Tsd-fx02 TSD faced SFTP server
\\tsd-evs\pxx\
SHH+RSYNC TCP 22
DMZ/UIO
Nettskjema
TCP 22
Tsd-fx01 Internett faced SFTP server
HTTP(S)
SFTP
TCP 22,TCP/UDP 3389
HTTP/HTTPS
Internet
PCoIP/Blast
LDAP
TCP/UDP All ports
TCP 11371
TCP 80,443,8530,8531
Coordinator for server provisioning Rhevm-tsd.uio.no
SSH TCP 22
VMWare Orchestrator Tsd-orch
Vcenter/esx/orchestrator traffic
Jumphost 1
TCP 3128
TCP-5986/ RDP-3389/ TCP-135
TCP 135, >=49152
TCP 22350-51
TCP 9427 4001,4002 389
Proxy traffic for downloading patches for windows
Jumphost 2
SSH+RDP
TCP 1688
Access Gateway Horizion View
VDI-GATEWAY
VLAN 2552 No routing
View Connection traffic
MMR (TCP-9427), PCOIP (TCP/UDP-4172), BLAST (TCP 22443)
Access GW Traffic
HTTPS (443)
RADIUS Authentication
ThinLinc reverse proxy
HTTPS(443)
ThinLinc mgmt (904)
ThinLinc Master
TS CALS
WinRM/ RDP/ Admin borwsing
DC Traffic
SCCM Agent
PGP Encryption Decryption
Codemeter
RDP
SSH/DC Traffic
Vcenter/esx/orchestrator traffic
RDP/DC traffic
Project terminal Server P11-win01
Project terminal servers PXX-Win01
HTTPS(443) ThinLinc Management (904)
KMS.UIO.NO
SSH TCP 22
ThinLinc Project Machines pXX-tlxx-l.tsd.usit.no
TCP/UDP ALL Ports
VMWARE Hypervisior
ESX mgmt P01-vcsa-ts02
Super Computer Colosuss

Secure Import/Export

Data ~~import~~ and export



NB : Cut -> paste from client to TSD is allowed, not vice versa

Data import now

- REST-API based

https://data.tsd.usit.no

Linux and Windows virtual servers

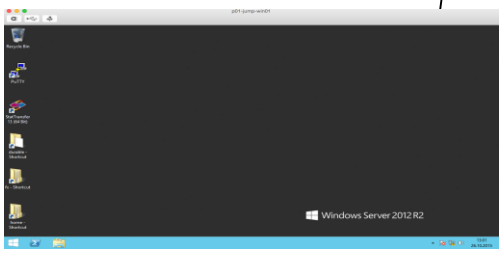
How to use TSD ?

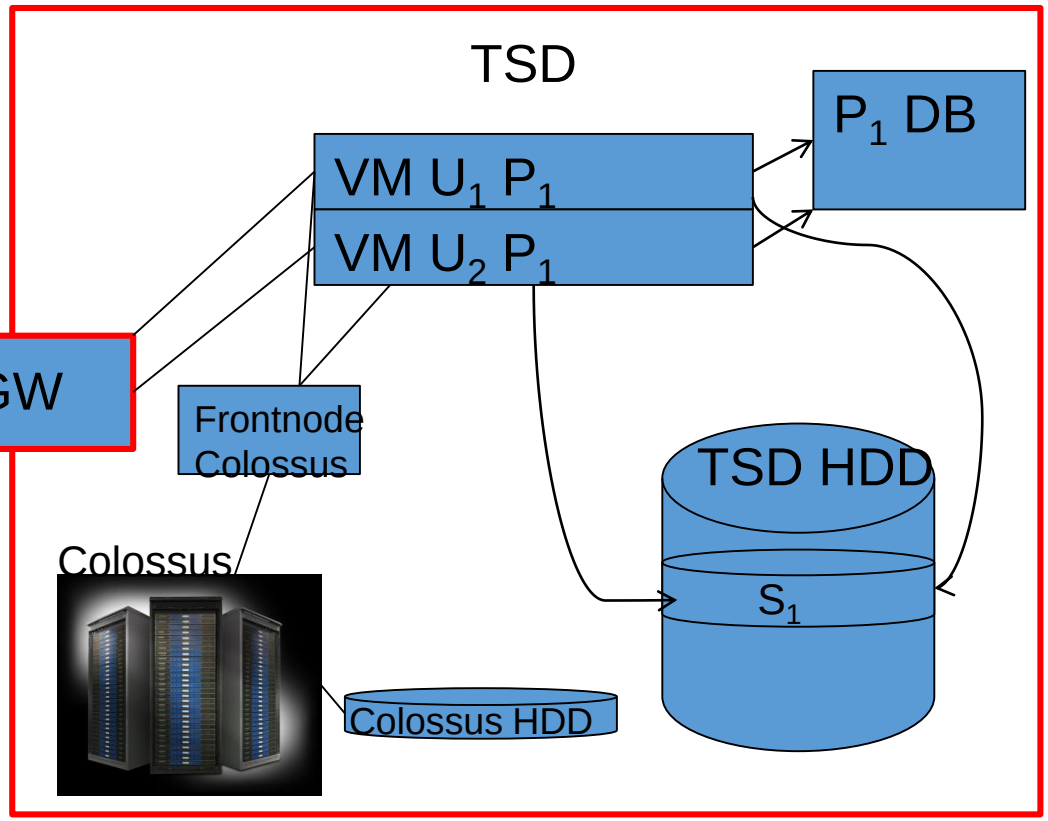
Libre Office
R
Module load
...

$User_1$ $Project_1$

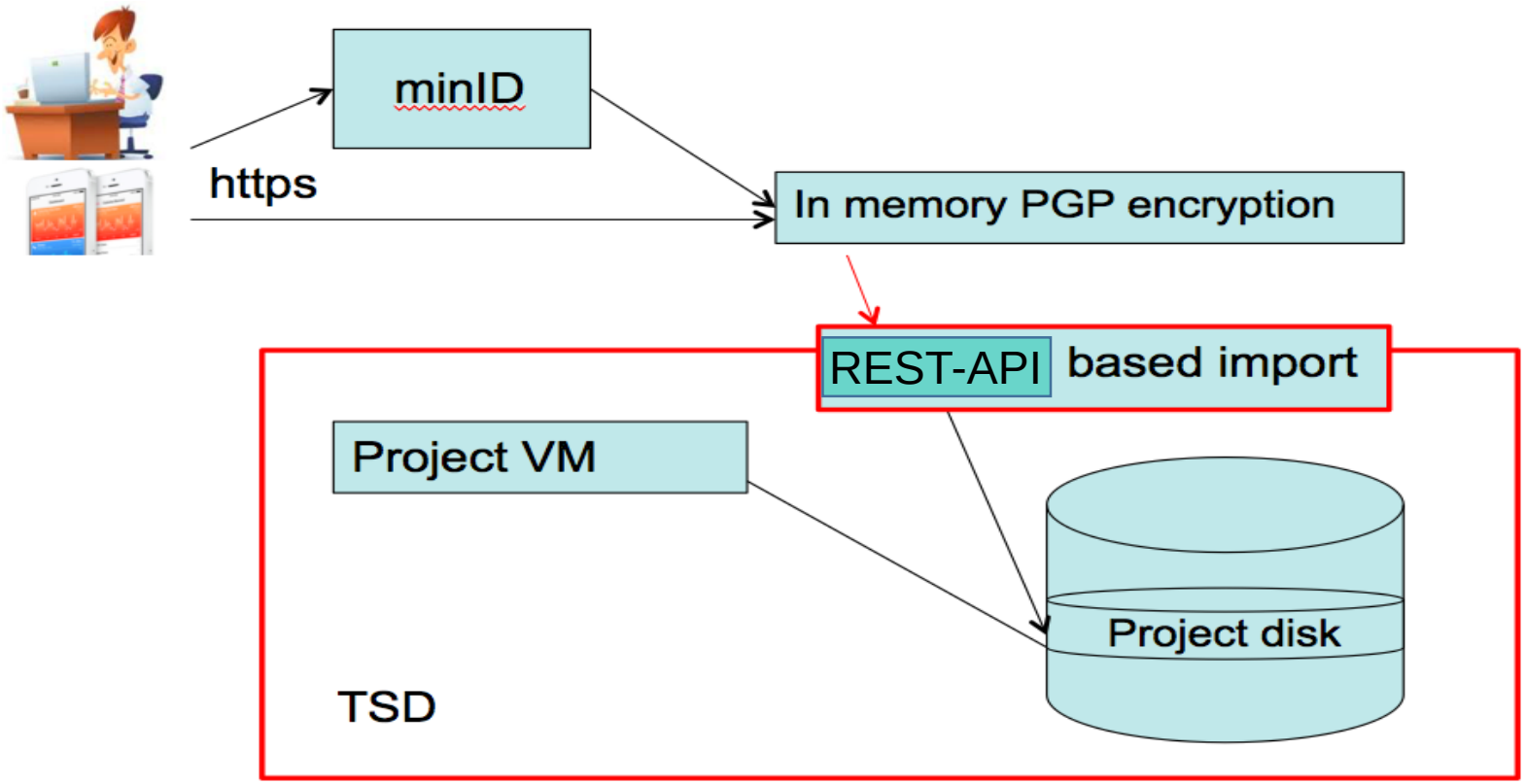
$User_2$ $Project_1$

SPSS
Office
Stata
SAS
R
Matlab
....





Data collection to TSD



Late 2018, 10-60.000 questionnaires per week

Data collection in TSD

- Nettskjema – www based self-service questionnaire
- API – import any datatypes from tablets and smartphones

HPC

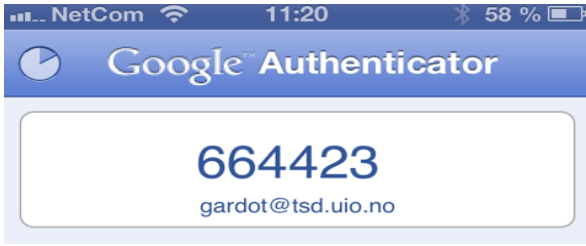
HPC – Colossus

- 1560 cores (~30 TFLOPs)
- 2 hugemem-nodes a 2 TiB RAM
- ~ 2,5 PiB BeeGFS HD
- No-login on the compute nodes
- SLURM queue system
- $SCRATCH is viped after each job
- Infiniband interconnect (56Gb/s Remote Direct Memory Access (RDMA))

Large Storage capability

Security details

- OATH TOTP 2-factor authentication
  - Smartphone or Yubikeys
- Data-export tightly regulated
- No open connections to the internet
- All administrations happens on the inside
- Strong separation between projects
- FreeBSD firewall and router
- Autogenerated firewall
- Encrypted backup, one key per research project
- Sys-admins are personal users as far as possible
- Sys-admin must use 2 factor for all admin tasks / login
- Sys-admin has granular access, root access =! storage access



TSD and end-user institutions?

or…
How to build such services in the real world

UiO : University of Oslo

eInfrastructure for research and clinical usage

Stakeholders
Stakeholders
Stakeholders

eInfrastructure for research and clinical usage

A good start :

- Stakeholders
- Use-case driven development
- Financing (long term)
- Skilled IT developers
- Skilled IT security
- Focused project management
- And throw in some lawyers

Vision for the clinical / research border

One patient == one research project

Many patients == big data analytics

Better health care – better treatment – better research

Reports on personalized medicine

Google estimated 5000+ results in Norway only

Anything happening in the clinics?

Personalized medicine -> precision medicine

Clinicians (at least some) really want to go there

There are someone responsible

The responsible can not do anything in practice

So…. not too much that is supported from above

Anything happening within research?

A lot – and way faster than in the clinic

- National Cancer Genome Consortium
- BigMed
- Norwegian researchers collaborate with Decode
- .. and a lot of smaller initiatives

So…. a lot happening... but not too much support for eInfrastructure from abov

BigMed

RCN funded project : $ 10.000.000 (2017-2019)

- Oslo University Hospital (OUS)
- University of Oslo
- Norwegian University of Science and Technology
- Sykehuspartner (Hospital IT)
- IBM
- PubGene
- And many more

Goal : Show the possibilities for big-data analytics and decision support at OUS

Organization

- Division Head – responsible for the service
- Group Leader – responsible for operations and development
  - Core development team
  - Core operations team
    - Dev and ops partly overlap and do perform DEVOPS
  - USIT basic operations teams support many TSD services
  - Houston (UiO) first line support
  - Officer on duty – one week per person from dev/ops teams
- Change advisory board – IT security have all power

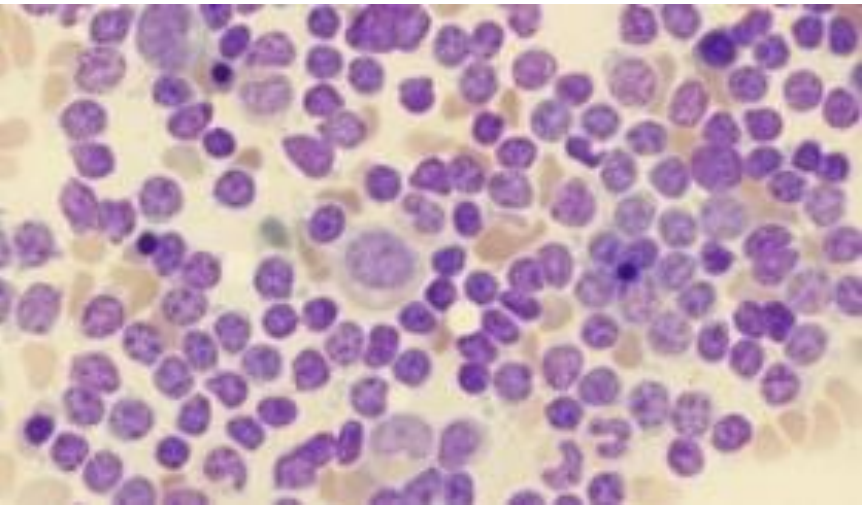
Mind-set

- Security
- Use case driven
- Any wishes must be made into something generic (i.e. for other users as well)
- Automatization
- Interactions
- Standards (JSON (XML), REST-API, S3)
- As few SPOFs as possible
- As few dependencies as possible
- Low time to market
- Modular development
- Self service

Capabilities Enabled in TSD

CANCER GENOMICS

Norwegian Cancer Genomics Consortium

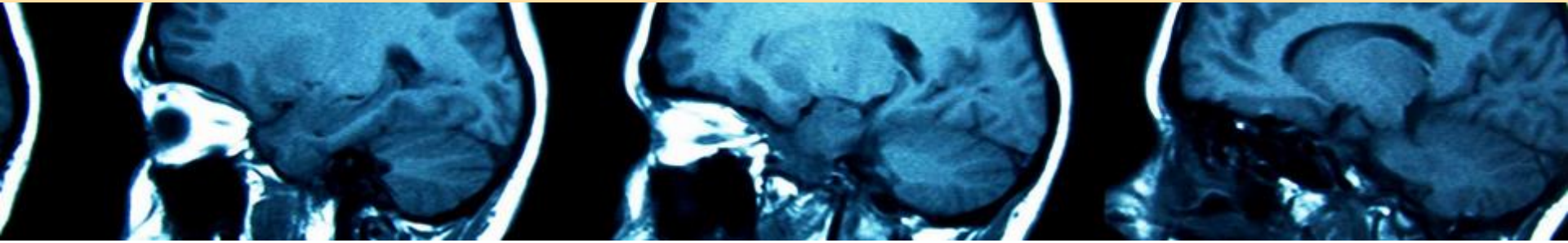


The largest project in Norway for Personalized Cancer Medicine

The project will establish the required general technology, methodology and interpretation procedures, which will be commonly shared, for utilizing the mutation profile of the tumor to guide the adaptation of cancer treatment to the individual patient.

TSD

COGNITIVE NEUROSCIENCE



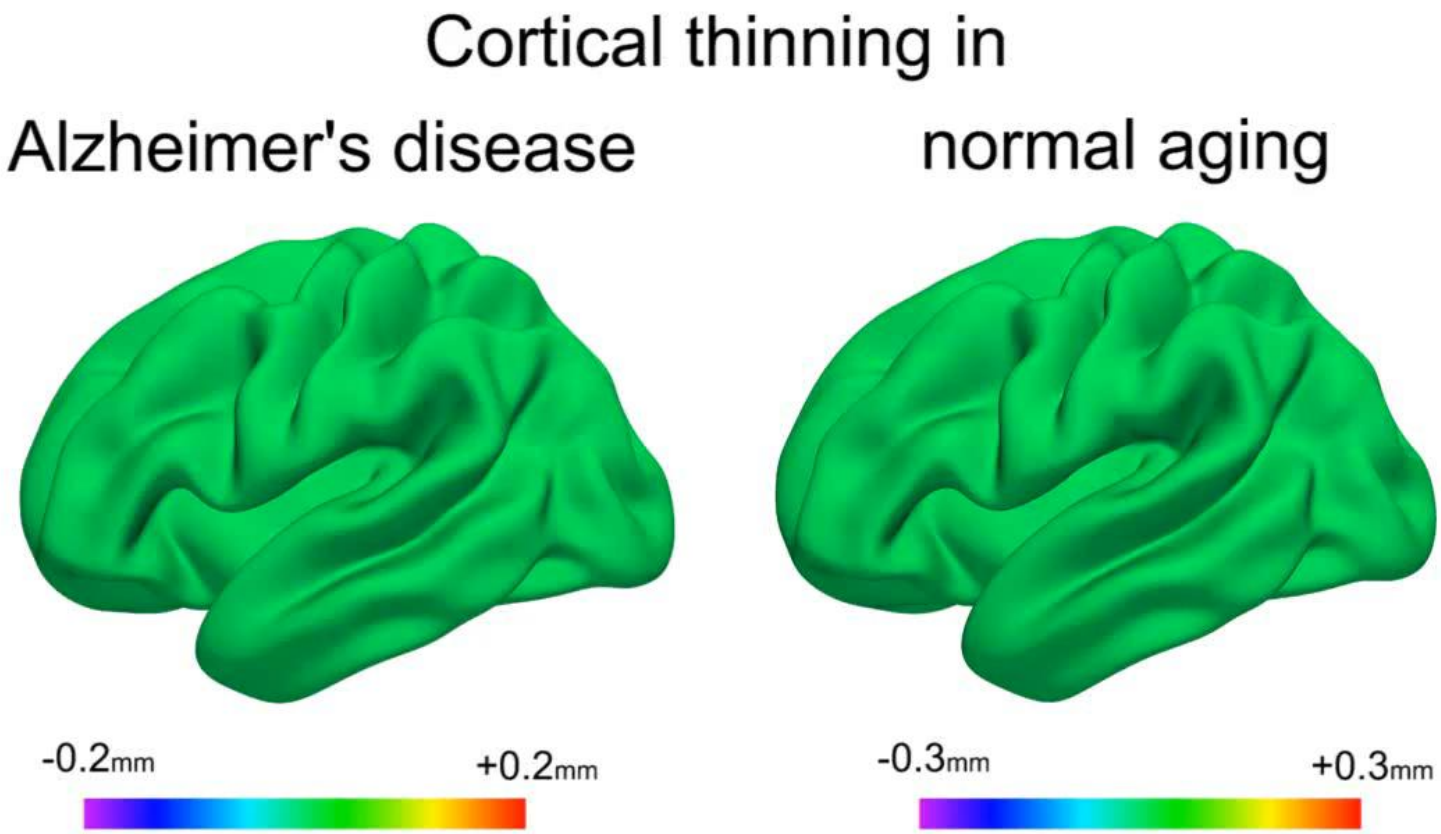
Life Brain – Breaking new ground in Lifespan Cognitive Neuroscience (*Prof.* Anders Fjell et co. PSI)

11 European Brain-Imaging Studies in 7 Countries will collect images of healthy brains and correlate the evolution of brain functions with the environmental conditions.

IMAGES COLLECTED FROM ALL OVER IN EUROPE AND STORED IN TSD FOR FURTHER ANALYSIS (XNAT)

What happens in TSD

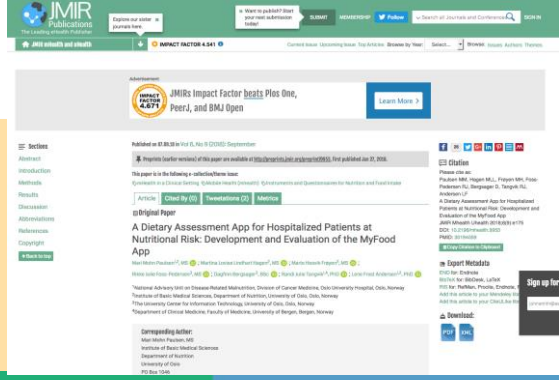
- Lifespan changes in the brain and cognition (LCBC)



Video courtesy of Inge Amlien, Post. Doc, UiO

UiO : University of Oslo

DIETARY RESEARCH



MinMat – an App to monitor nutritional habitudes in hospitalized patients and post-hospitalized persons (Prof. Lene Frost Andersen)

App that allows to collect easily nutritional data directly from the patient.

FREQUENT DATA INGESTIONS; NEED FOR A SECURE REPORT MODULE ACCESSIBLE FROM THE OUTSIDE OF TSD UPON SUCCESSFUL MIN-ID AUTENTICATION.

What happens in TSD – Kidney transplants

UiO : Realfag og teknologi

Titan.uio.no
...ny kunnskap hver dag!

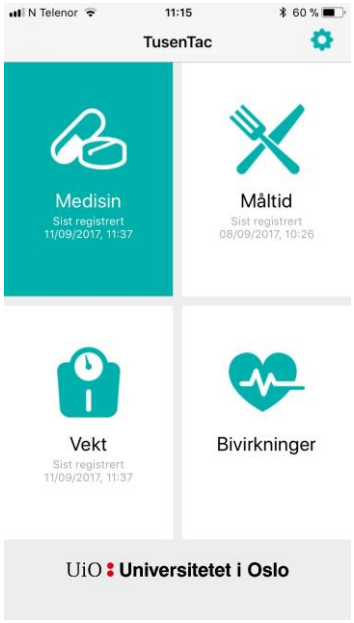
– Det er veldig individuelt hvor mye immundempende medisin ulike pasienter trenger. Derfor er det veldig viktig å finne den riktige dosen til hver pasient, forteller Anders Åsberg ved Farmasøytisk institutt. Bruk bildet.

App for transplanterte

permalenke

Folk som har fått transplantert nye organer må ta en nøyaktig dose immundempende medisiner to ganger daglig resten av livet. Farmasøytisk institutt og USIT skal nå utvikle en mobil-app som kan hjelpe pasientene med å huske på de livsviktige medisinene.

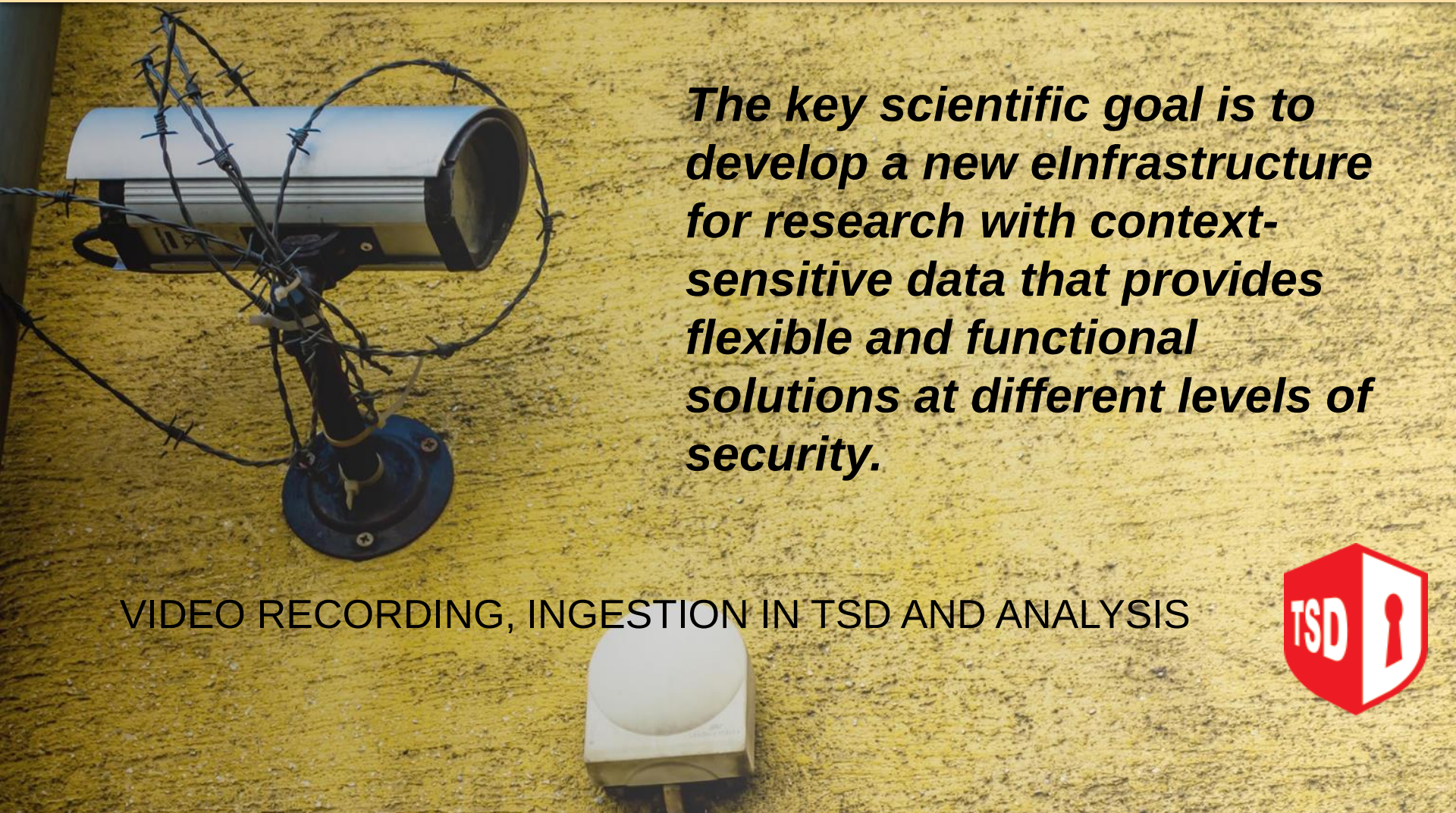
av Gunhild M. Haugnes – 11. november, 2015



https://titan.uio.no/node/983

TSD

SENSITIVE VIDEO (eVIR)



The key scientific goal is to develop a new eInfrastructure for research with context-sensitive data that provides flexible and functional solutions at different levels of security.

VIDEO RECORDING, INGESTION IN TSD AND ANALYSIS

ALLERGIES AND PUBLIC DISEASES

PreventADALL – Preventing Atopic Dermatitis and Allergies *(Prof.* Karin Lødrup Carlsen, Oslo University Hospital)

Identifying environmental and lifestyle factors through pregnancy, childbirth and childhood that affect the development of allergic and other public diseases, also later in life.

> 60000 NETTSKJEMA ANSWERS

MOBA

MoBA – Norwegian Mother and Child Cohort Study –
(FHI, Folkehelseinstitutt)

The largest cohort study in Europe to correlate mother and child health status. Since 1998 data are collected from 17th week of pregnancy.

DATA COLLECTED THROUGH NETTSKJEMA

From paper to bits and bytes

Anno 2000

Anno 2018

COMING SOON … SENSORS

«How do I feel ?» Technology-supported health care for people with severely limited communication skills
(SINTEF , OUS)

DATA COLLECTED DIRECTLY FROM SENSOR, INGESTED AND MONITORED IN TSD. SEVERE CONDITIONS NOTIFIED TO THE HEALTH CARE PERSONELL

Enabled by TSD today

- Large scale human genomics
- Large scale image studies (fMRI etc)
- Regular and complex computational analysis
- Controlled sharing of data and collaboration across borders
- Electronic questionnaires (in a secure way)
- Digital consent (soon dynamic consent)
- Remote access to sensitive data
- Research on sensitive recordings (video and sound)
- Data collection from apps and IoT
- Dictaphone via your smartphone

TSD as a research platform

- Wide user agreements with HSØ, NIPH, NTNU, UiT, OsloMet
- TSD is used by a wide range of research institutions
- Ongoding collaborations with similar infrastructures (UiB Safe, HUNT and EUTRO (UiT))
- Collaborations with Dir for eHealth – HAP
- Collaborations with Statistics Norway
- National provider through Sigma2, funded by NFR
- Partners in NeIC Tryggve2 (http://wiki.neic.no/Tryggve)
- Service providers in EOSC-HUB (etterfølger til EUDAT), leading WP 6.6
- Collaborations with Elixir Norway
- Collaborations with Sykehuspartner
- Collaborations with Privacy Officer at OUS
- Main platform for the BigMed-project
- Established start-up collaboration with Inven2 and IBM

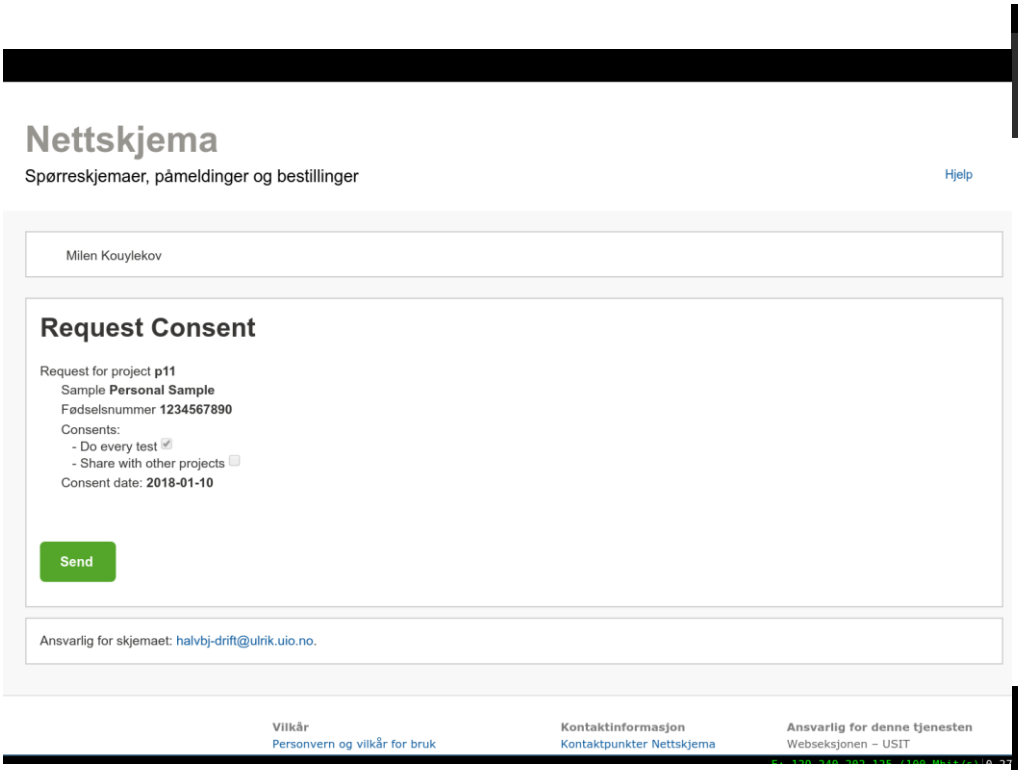
TSD in the future

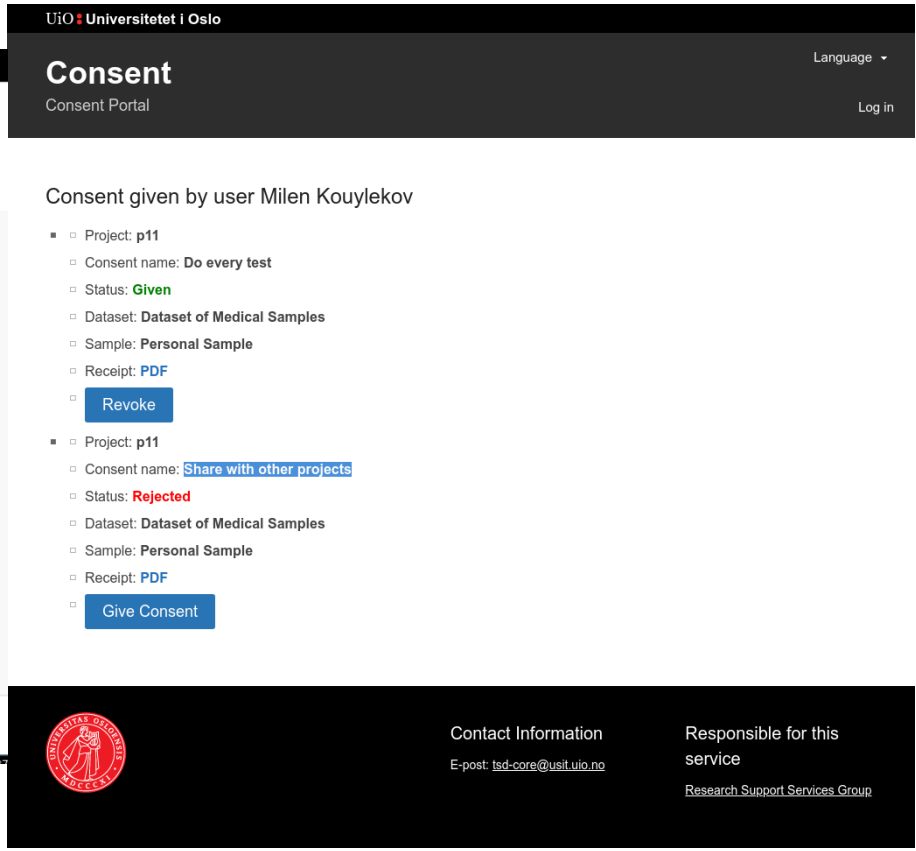
- Technical development and growth
  - Digital dynamic consent system
  - Full audit of databases
  - Deliver data back to the research subject
  - ML/AI capacity

- National platform for sensitive data – via Uninett Sigma2
- Central role in the Helseanalyseplattformen – downstream of data hand-out
- We want to be in on the development of a national clinical solution
- for HPC and mass-storage in the clinic
- IT-technology “research”
- International collaboration
- National role as innovation platform
  - Requested as partner in 7 NFR einfra proposals H2018

UiO : University of Oslo

Digital dynamic consent



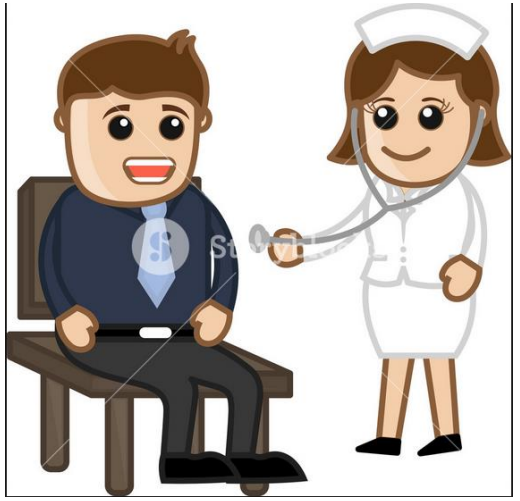


In production by 1/1-2019, looking to integrate more IDPs

What´s in it for me?

- Research objects wants feedback

- The consent solution and our «myTSD» webpage will solve this. First use-case is in the pipeline with the Dept of Psychology at UiO.



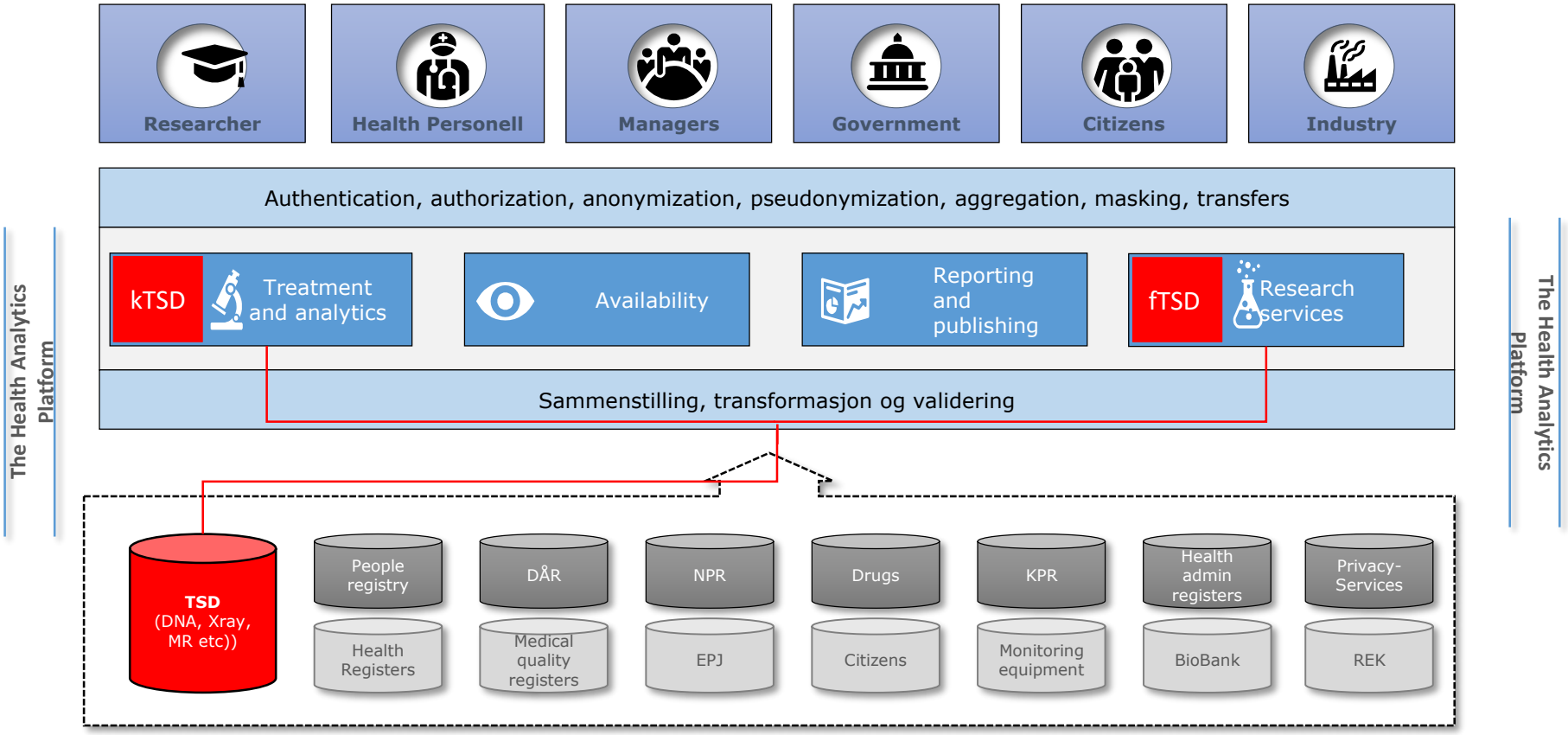
Picture courtesy of Storyblocks

TSD and “Helseanalyseplattformen” / HelseNorge.no

- TSD as a partner:
  - TSD has established competence, trust and insight
  - User driven development
  - Dynamic, agile, and low time to market
  - Low price, suitable for researchers
  - Established services :
    - Compute capacity – TSD is the “Analysis platform”
    - Data collection nnsamling
    - Apps
    - REST-APIs for data transfer
    - Dynamic consent (1.Jan 2019)

- Use it in HelseNorge.no!
- Reuse for a future clinical platform!

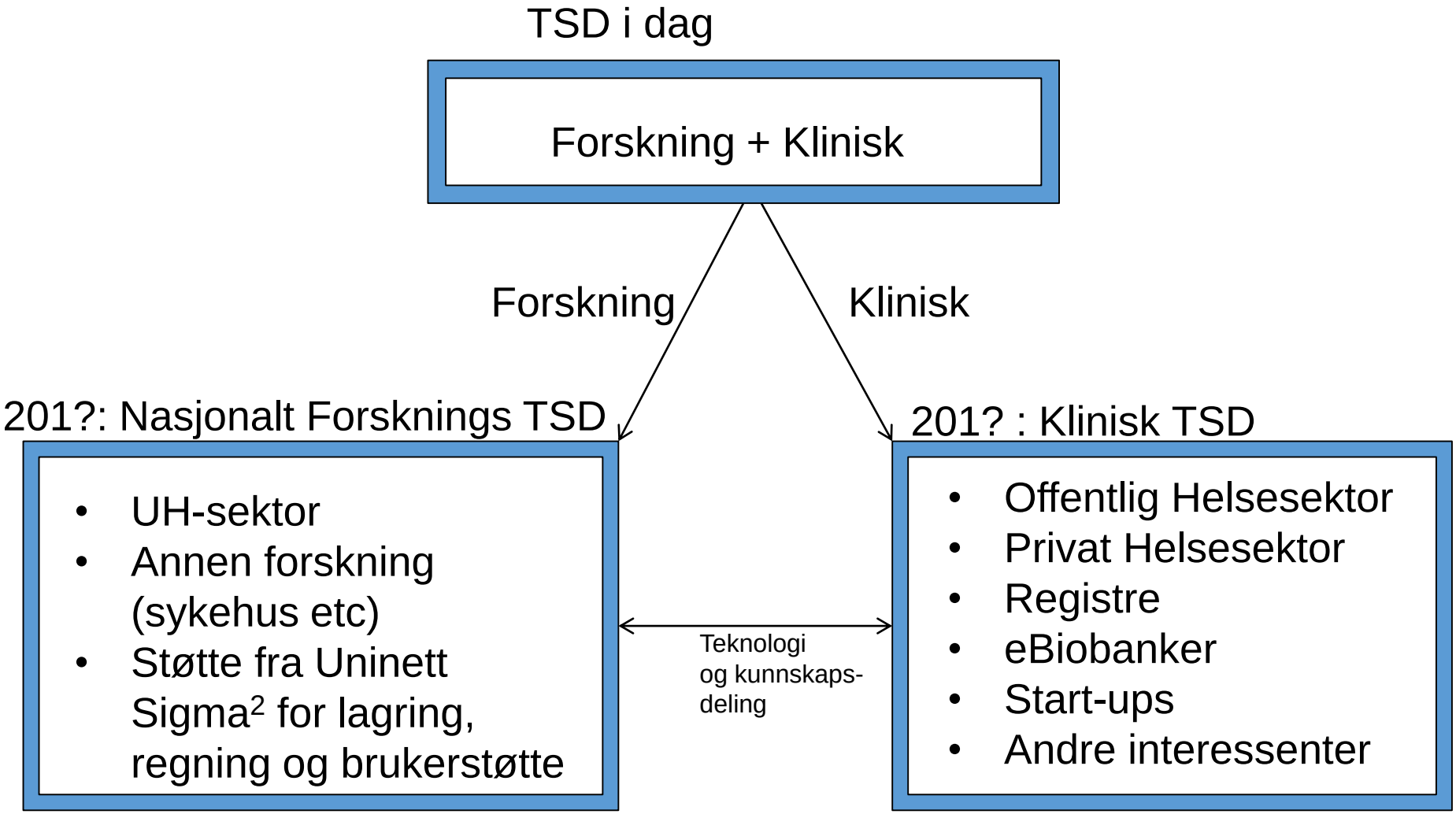
Helseanalyseplattformen



Slide from Dir of eHealth, modified by TSD

UiO : University of Oslo

TSD organisering



Thank you for the attention

TSD

tsd-contact@usit.uio.no

USIT

gardot@usit.uio.no