Services for sensitive data
and their integration with European e-infrastructures

Antti Pursula, CSC

Maria Francesca Iozzi, USIT UiO

CSC

UiO : University of Oslo

DI4R Conference, Brussels

1.12.2017

EUDAT

neic
NORDIC E-INFRASTRUCTURE COLLABORATION

Use of human data in medical research

PERSONALIZED MEDICINE

NEW CURES FOR DISEASE

IMPROVED QUALITY OF LIFE

Vast amounts of human data is being collected by different actors and for different purposes, e.g.:

- Biobanks
- Large genome studies
- Register organizations
- Hospitals
- Personal health apps, etc.

Society benefits when these data collections are made available for research!

Protecting the privacy of individuals

CHALLENGE IS TO PREVENT NON-AUTHORIZED USE OF DATA WHILE UTILIZING THE DATA FOR THE BENEFIT OF SOCIETY AND INDIVIDUALS

Research use of health data requires:

- Compliant data protection and IT security practices
- Secure data environments
- Trusted data access
- Ethical permissions
- Effective legislation

Health and well-being data on humans is sensitive, personal data that needs to be protected, even after de-identification

Challenge of providing services for sensitive data

- It's not only medicine but all research that concerns data about humans: social sciences, language research, digital humanities etc.

- Privacy of individuals vs. benefit of society

- General Data Protection Regulation (GDPR) to be implemented by May 2018

- Service providers need to balance between security and usability; and be compliant with laws and regulations

How is this done

- Service intended to manage or process sensitive data will need to be carefully designed to satisfy requirements. This typically leads to solutions such as:
    - Separated from the internet
    - Multifactor authentication
    - Private network connections
    - Encryption in transfer and storage
    - Well-defined authorization mechanisms
    - Strict separation of customer data from each other
    - Certifications, compliancy
    - …

How is this done…

- On the other hand: several data sources? European wide access? Tailored software pipelines? Integration to other infrastructure components?
  - Additional level of complexity

- Also, where the data is analyzed or stored; and who has the control over the data are different things.
  - Taking into use external secure service does not imply transferring the control over the data to the provider

Secure services and selected use cases

Operational services for sensitive data – CSC ePouta secure cloud

- ePouta secure cloud is a closed environment that meets elevated information security level regulations as defined by the governmental authorities in Finland.

- Suitable for all fields of science

- The cloud service combines virtual computational resources with the customers' own resources using a dedicated network connection (OPN or MPLS)

- Not only a platform for processing of data but also integrated secure layered storage

- Developed to large extent within ELIXIR Finland (located at CSC)

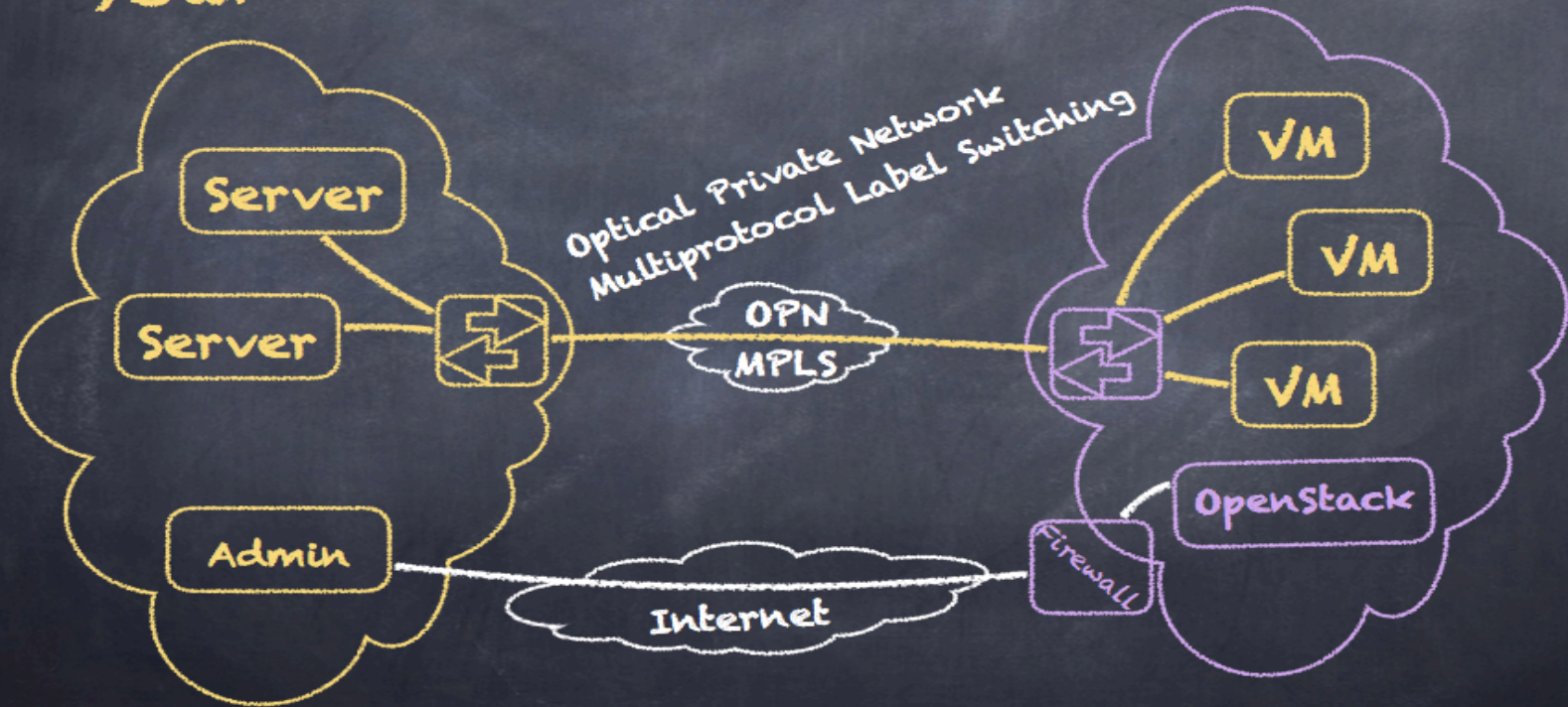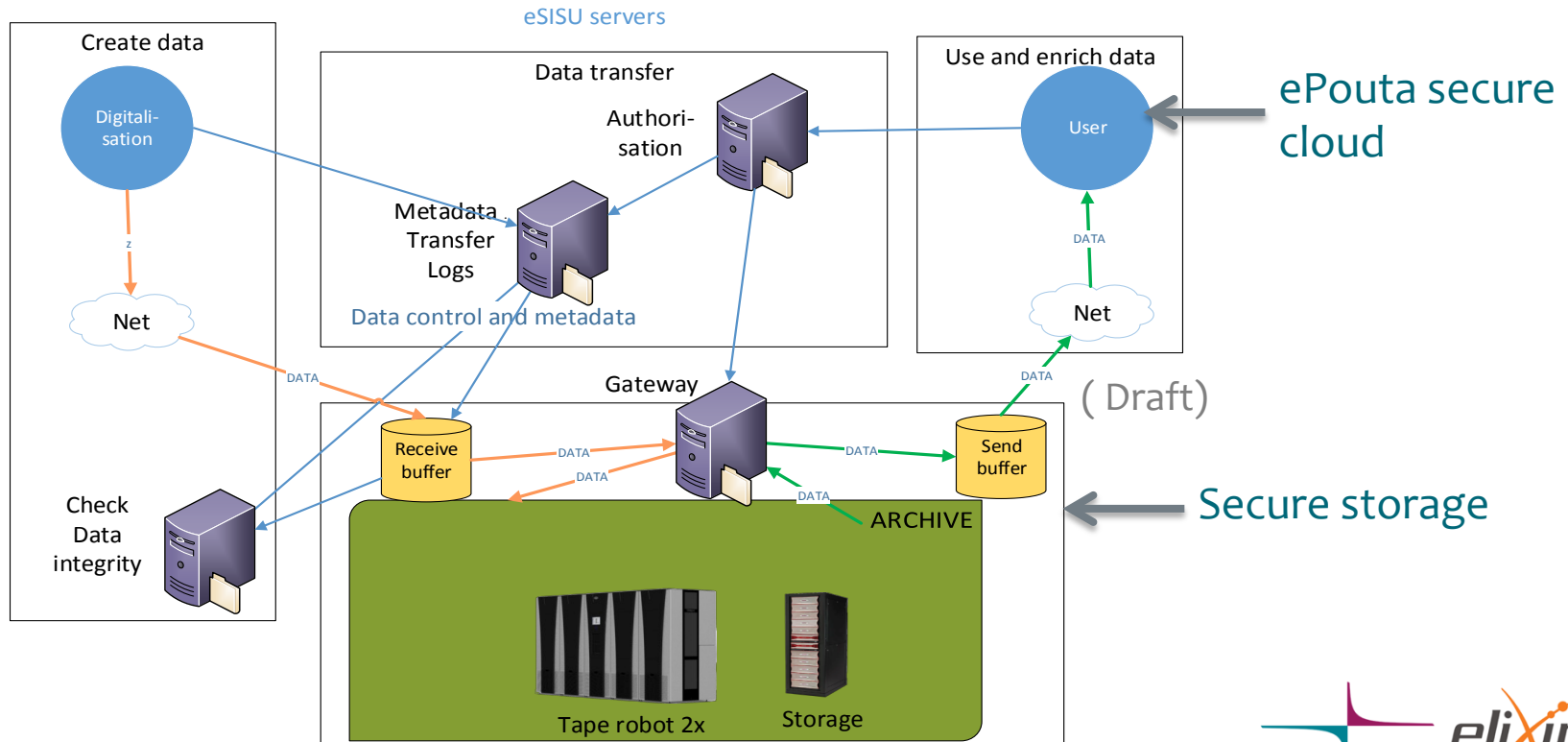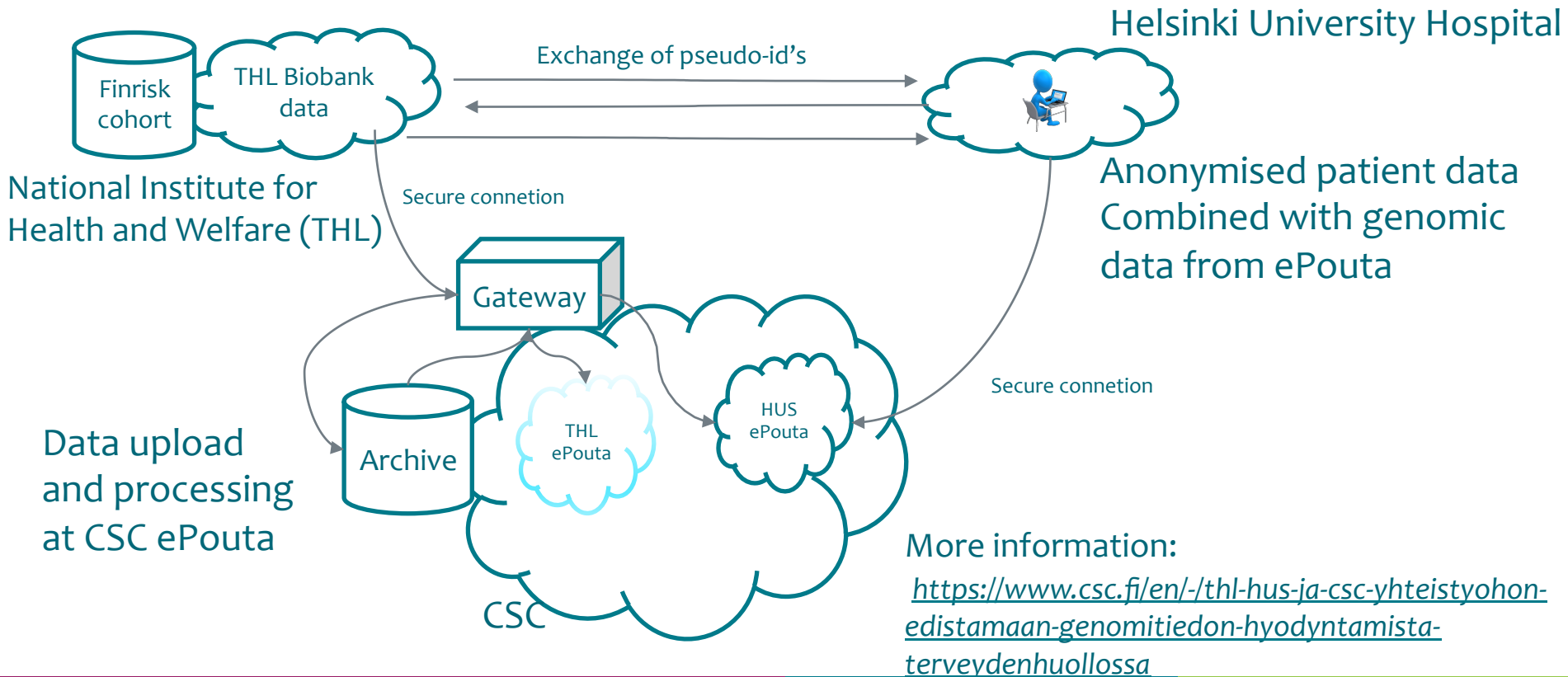- Service provider is CSC – also a EUDAT service provider

C S C

Data management for Sequence Initiative Suomi (SISu)

See also: *European Journal of Human Genetics* (2017), doi:10.1038/ejhg.2016.205

Finnish study to use genome data in health care

Exchange of pseudo-id's

Helsinki University Hospital

National Institute for Health and Welfare (THL)

Secure connetion

Anonymised patient data Combined with genomic data from ePouta

Gateway

Data upload and processing at CSC ePouta

Archive

THL ePouta

HUS ePouta

Secure connetion

CSC

More information:
https://www.csc.fi/en/-/thl-hus-ja-csc-yhteistyohon-edistamaan-genomitiedon-hyodyntamista-terveydenhuollossa

Operational services for sensitive data – TSD secure service

- TSD is designed to host and process sensitive health data in compliance with the Norwegian law; now used for sensitive data from all fields

- Use case driven development

- Service launched in May 2014.

- TSD system development focused to keep the balance on the the edge between security and usability always in compliance with the law.

- Includes data Collection from Web-based questionnaires, which can be used on portable devices

> 1800
USERS

Circa 400
PROJECTS

> 1,7 PiB DATA

TSD Today
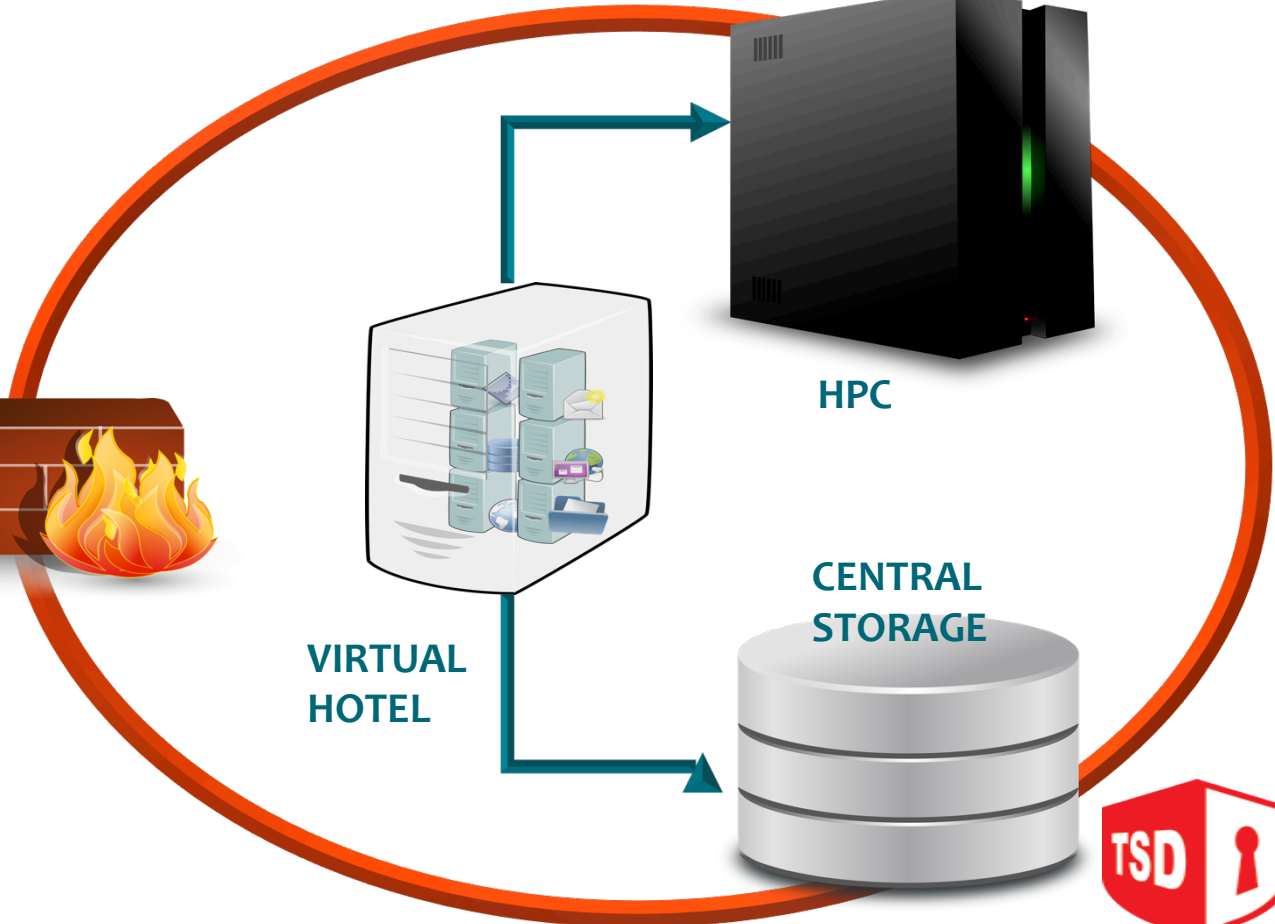
TSD

TSD

GATEWAYS
2-FACTOR
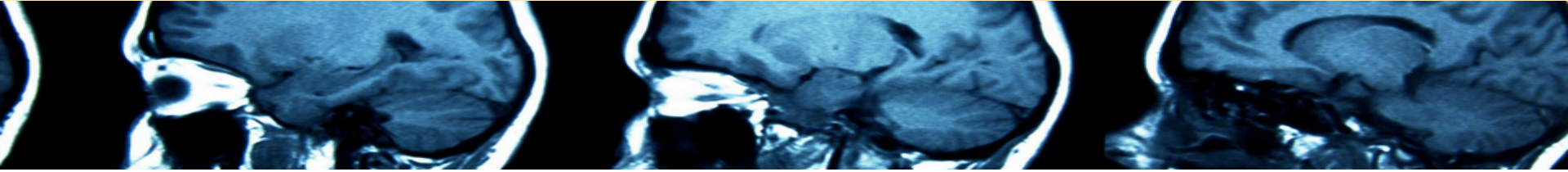AUTH

HPC

VIRTUAL
HOTEL

CENTRAL
STORAGE

TSD
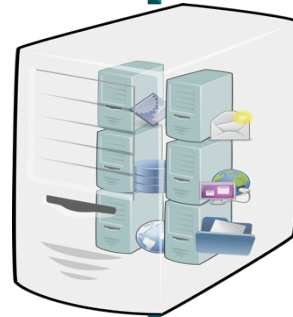
Secure Import/Export

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)

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 ELECTRONICALLY AND
AUTOMATICALLY SENT TO TSD

17

Integration of the secure services with European e-infrastructures

- Integration to European e-infrastructures is a further challenge

- Novel research can be done by allowing user to access services and data across Europe, which calls for integrating data and services together

- The following shows current activity towards integration of TSD and ePouta to international infrastructures

Integration of ePouta and TSD

- A case study to enhance the interoperability between existing services for sensitive data
- Part of Nordic NeIC Tryggve project
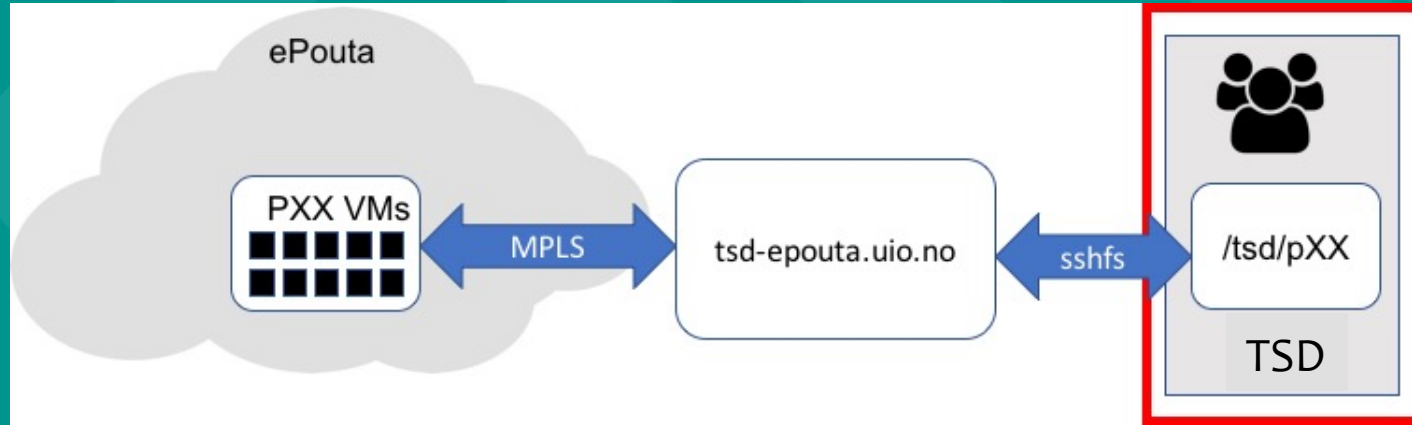
- Use case scenario: The user need more computing capability.
- The TSD closed network is expanded into the ePouta secure cloud infrastructure.

Collaboration with the European Genome-Phenome Archive EGA

Work in progress towards integrating the two systems into European Sensitive data archiving infrastructure
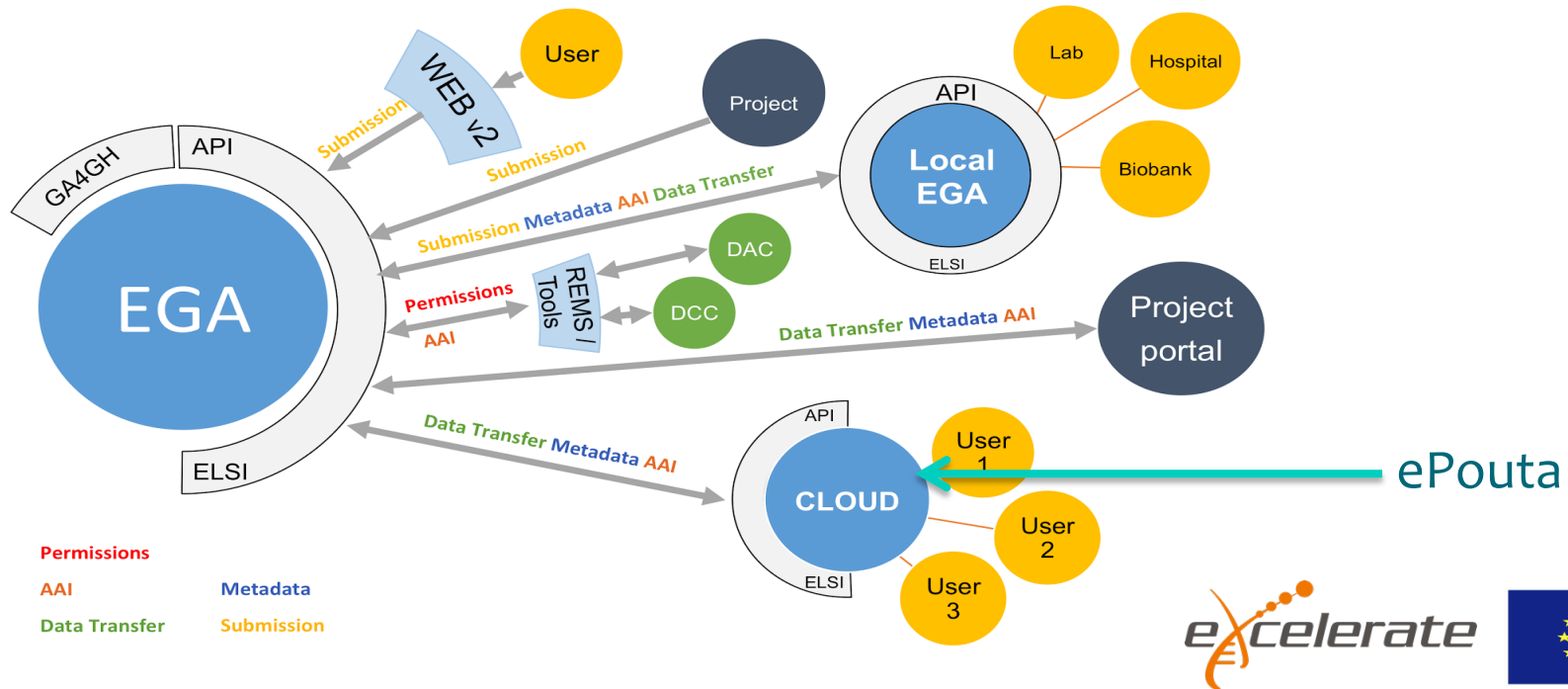
Integration with the EUDAT B2services

- *Use case scenario: users wants to make the anonymized metadata available on public repository*
- *Solution: users upload data together with metadata through a B2SHARE instance running inside TSD*
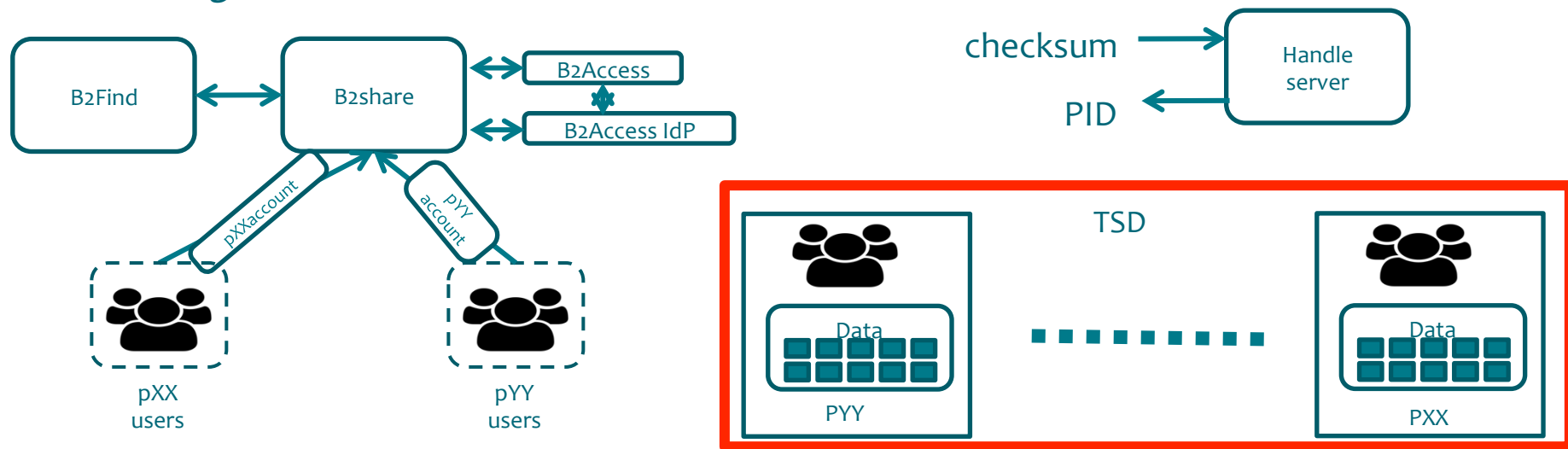
In preparation for EOSC

- EUDAT Sensitive data working group is collecting the requirements from several communities handling sensitive data

- Both CSC (ePouta) and SIGMA2 (TSD) are partners of the EOSC-hub project

- Both service providers are preparing the legal and technical framework on offering the secure services on a European scale with EUDAT CDI and through EOSC service portfolio

- In particular, the EOSC-hub project includes tasks for facilitating the use of these service for European research

Summary

- Using sensitive data for research is a challenge to the e-infrastructure service providers – but it can be solved

- Operational national and regional services satisfying security requirements exist already, and extending these into truly European services is being pursued

- CSC (ePouta) and USIT (TSD) are working towards EUDAT and EOSC integration for wider accessibility

Thank you

Contact: antti.pursula@csc.fi