T5.2 - Analysis and Post-processing across borders (M1-36)

Lead: SIGMA2. Participants: CSC, UT/ETAIS, SNIC, UICE, SDU, FMI, UIO-USIT/SIGMA2, UGOT/GGBC, UIO-GEO/SIGMA2, UIO-INF/SIGMA2, UH, CGG/UCPH

T5.2.1: Cross-border data processing workflows (M1-36)

T5.2.2: Code Repositories, Containerization and "virtual laboratories" (M1-36)

T5.2.3: Platform as a Service for Scientific Cloud Computing and Cloud Native Execution mechanisms (M1-36)
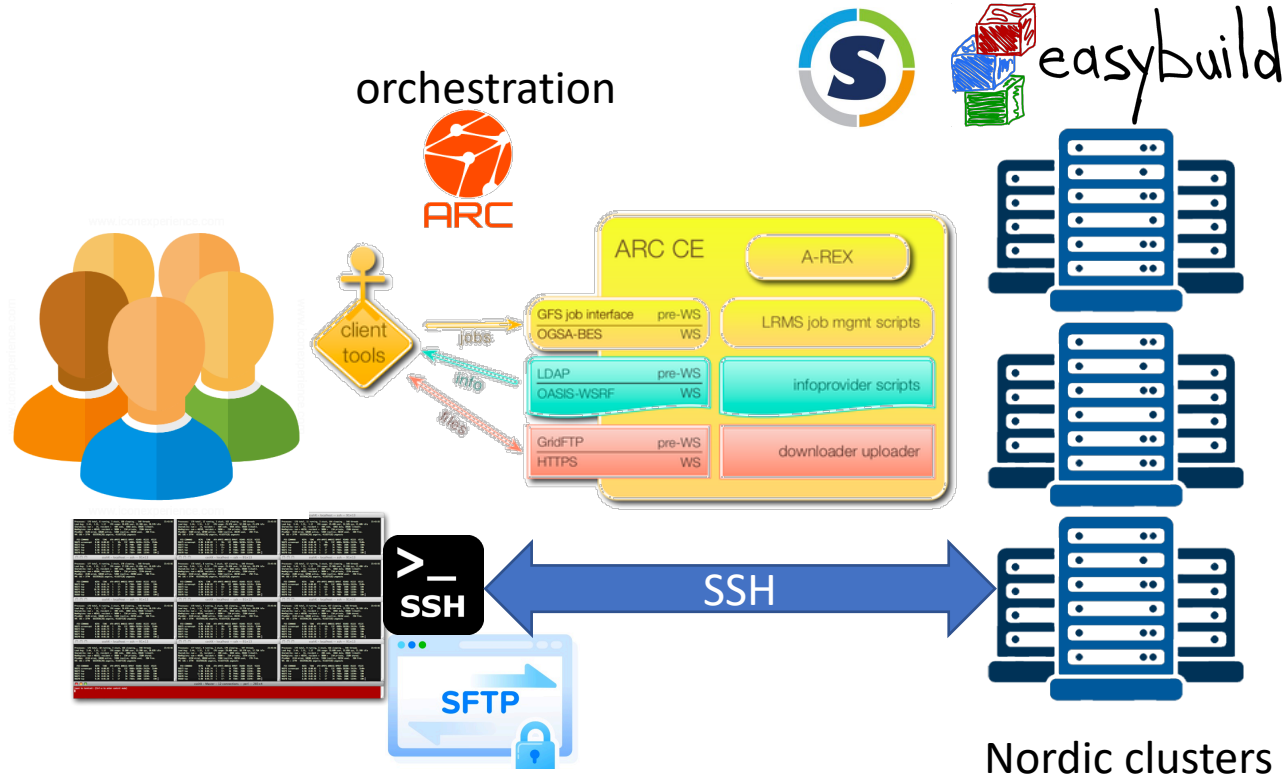
Use cases – NLP (Natural Language Processing)

User needs:

- Easily use NLP software (don't need to install themselves).

- Submit jobs to different Nordic clusters

- Have the same software environment on different Nordic clusters

- Easily package the current environment setup, in order to share and make possible to easily replicate the research results (see Research group E below).

Use cases – NLP (Natural Language Processing)

Models

orchestration

easybuild

ARC CE

A-REX

client tools

jobs

GFS job interface   pre-WS
OGSA-BES   WS

LRMS job mgmt scripts

LDAP   pre-WS
OASIS-WSRF   WS

info

infoprovider scripts

GridFTP   pre-WS
HTTPS   WS

files

downloader uploader

SSH

SFTP

Nordic clusters

Use cases – NLP (Natural Language Processing)

Status:

- The framework needed to start building NLPL modules provided
  - https://source.coderefinery.org/nlpl/easybuild

- Scripts to locally test in shared resources created

- Tested different configurations for basic packages

- Created patches for MPI Machine learning applications (e.g. Tensorflow)

Sabry Razick
USIT/UiO

Use cases – Climate change

orchestration

HTTPS

PULSAR

docker

Nordic clusters

Production: climate.usegalaxy.eu (jobs to cPouta)
In progress: climate.usegalaxy.no
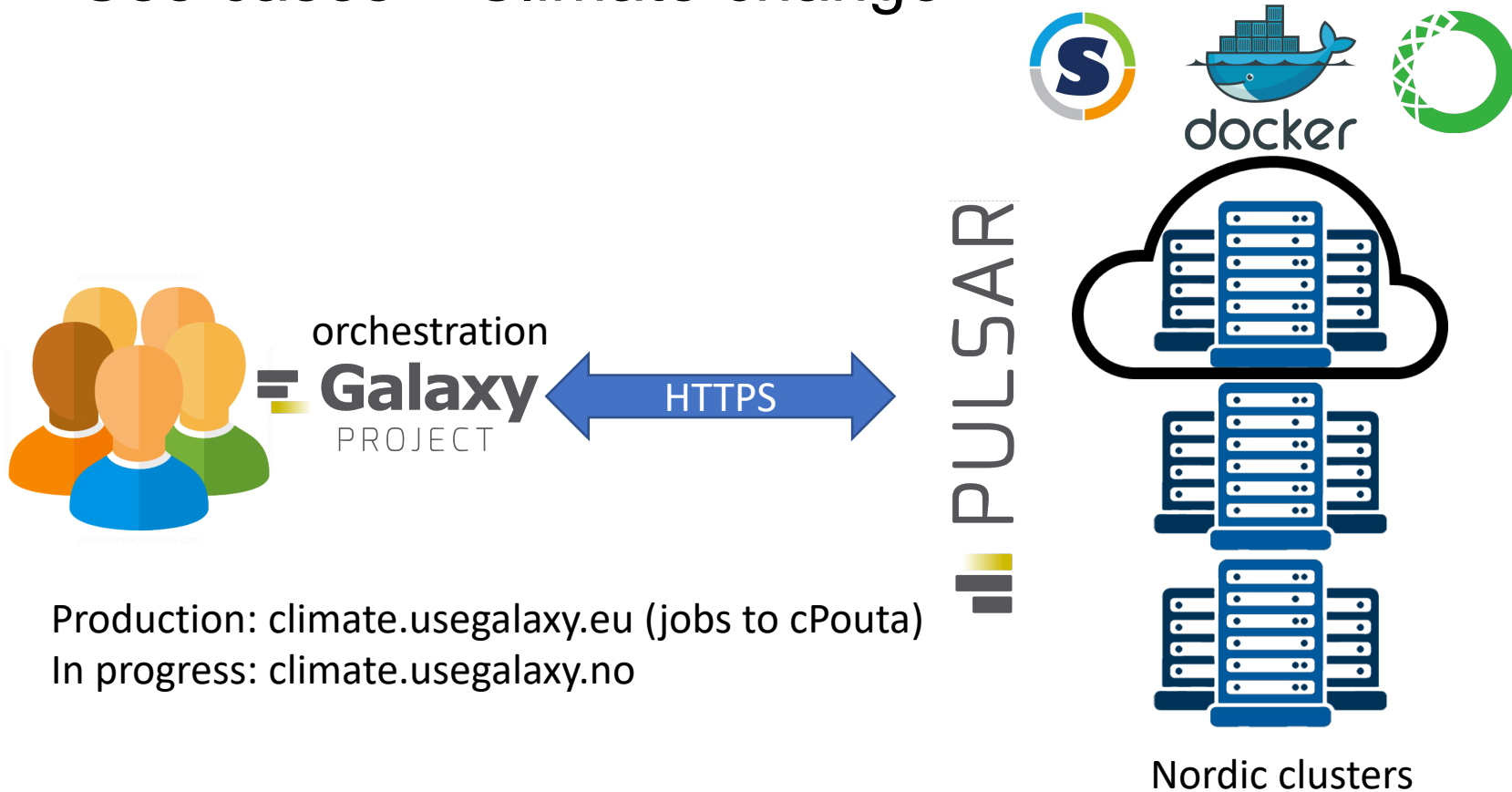
Use cases – Climate change

Issue (Access control and Accounting):

- How would Galaxy users authenticate against the backend HPC, and how Galaxy would perform accounting for CPU/GPU hours?

- Options:
    - A: Map Galaxy users to HPC users (ongoing dialogue with Puhuri project)
    - B: Local Accounting and access control in Galaxy and one project user at HPC

Use cases – DeepDive

orchestration

switches

SSH

SFTP

SSH

SFTP

SSH

SFTP

Nordic clusters
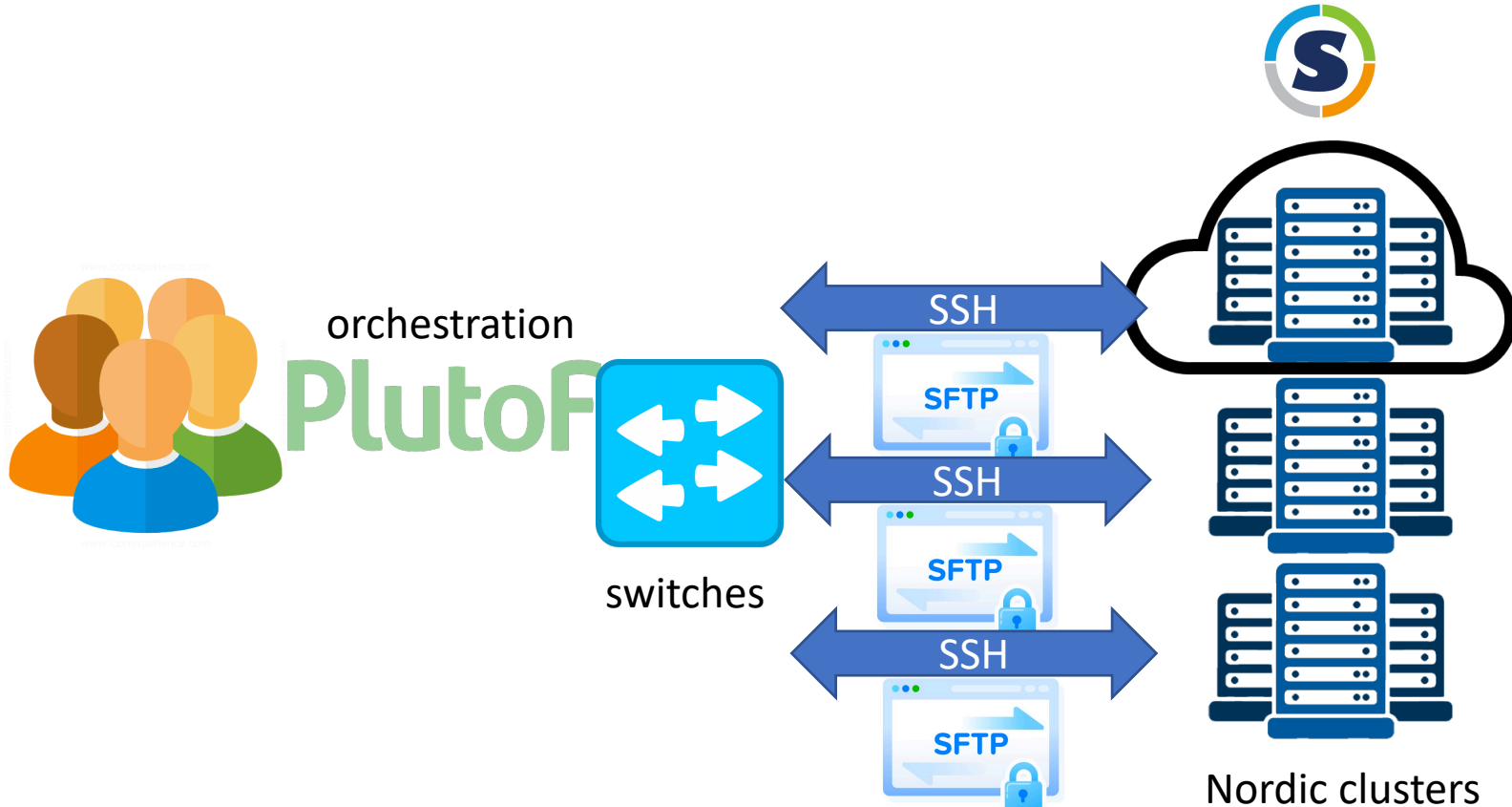
T5.4 - Sensitive Data (M1-36)

Lead: UIO-USIT/SIGMA2 (2PM). Participants: UIO-USIT/SIGMA2 (11PM), CSC (7PM), SNIC (11PM), Computerome (11PM), RH (5PM), UCPH (5PM)

- T5.4.1: Cross borders orchestration of sensitive data clouds

- T5.4.2: AAI and operations procedure for a Nordic eHealth Cloud

Use cases - Precision medicine

The main challenge will be to combine data from three countries. In Denmark, data has been imported to the secure cloud. In Norway, somehow difficult to get the needed permissions to use data, still in dialogue. Typically, it takes 17 months from protocol to access.

Data sources: deCODE genetics (in addition to other national data archives). Data transfer:
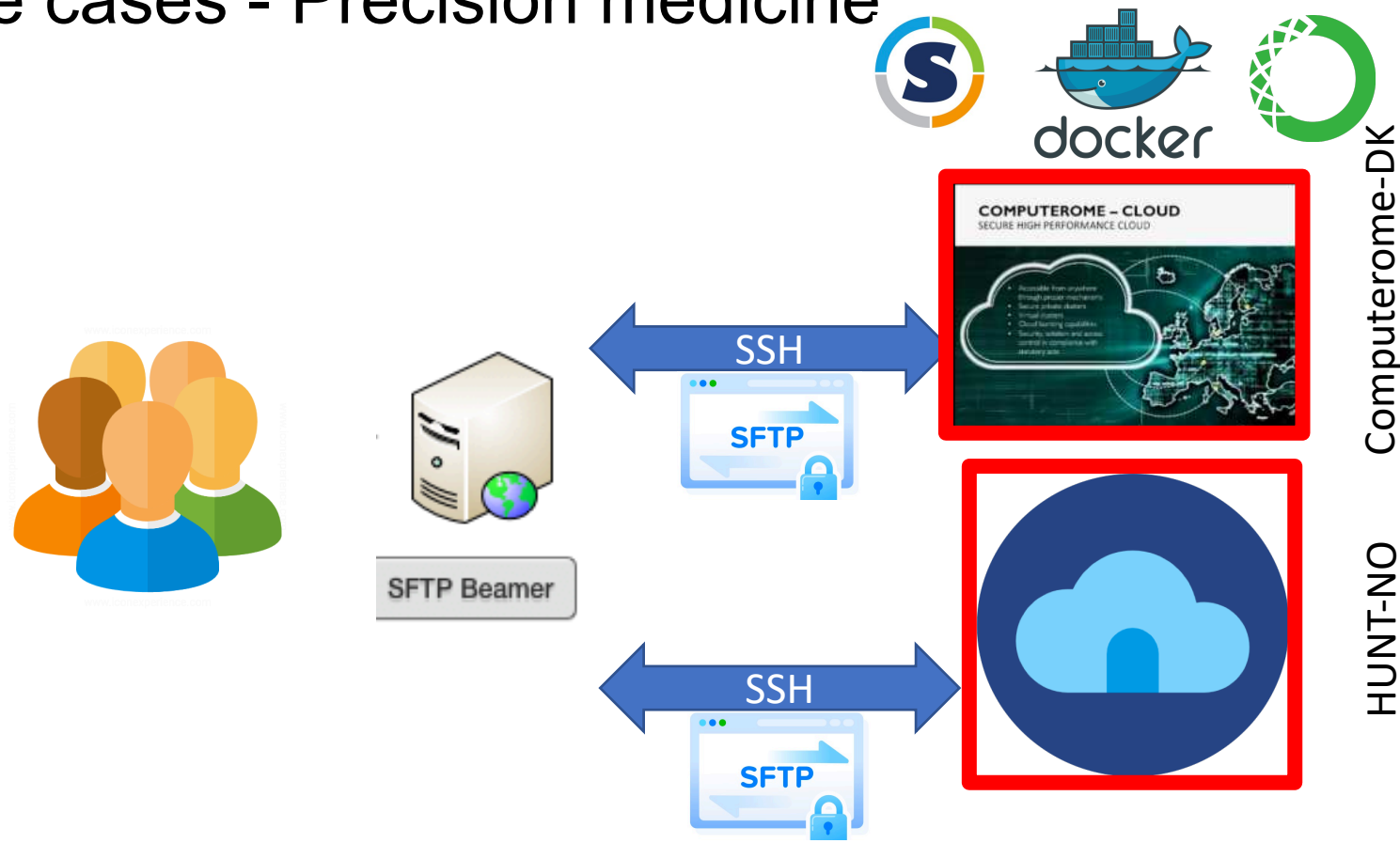- encrypted HDD
- Special "commercial" tool for filtered data importing
- SFTP beamer could be used for data transfer

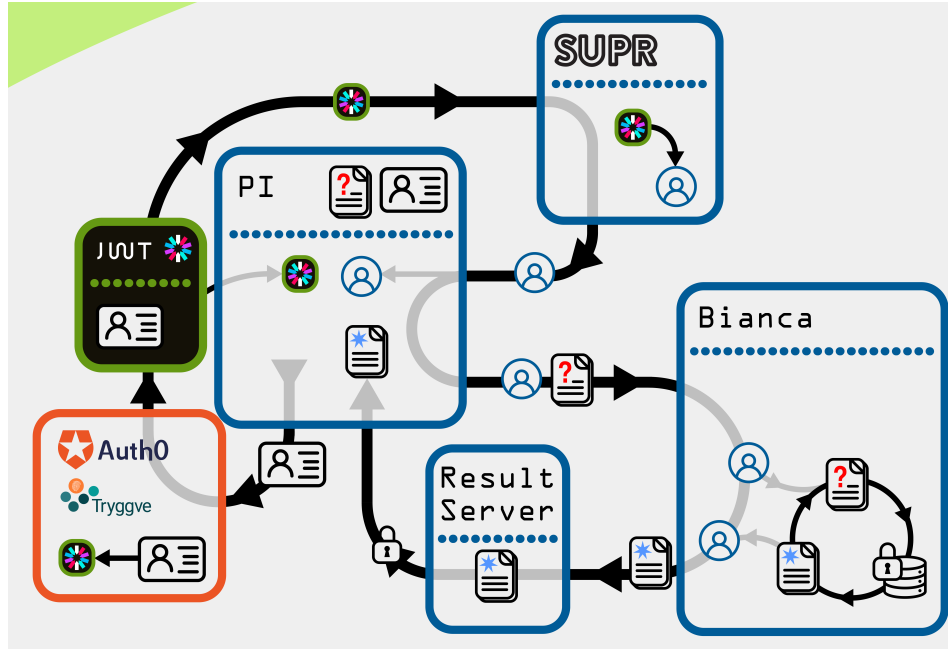**Locations:** HUNT cloud (IaaS), Computerome (secure cloud - IaaS)

Cross-border data processing (data from site x with data from site y): options:
- Moving data from x → y or y → x
- Stream data from x → y or y → x

Use cases - Precision medicine

Federated token based access – in progress

Thank you