PAUL SCHERRER INSTITUT

PSI

CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Developing Real-time Services for Large Volume Experiment-Data Analysis utilizing Supercomputing and Cloud technologies at CSCS (SELVEDAS)

Dr. Mei-Chih Chang
mei-chih.chang@psi.ch
PSI Science IT
22.10.2020

PSI challenges

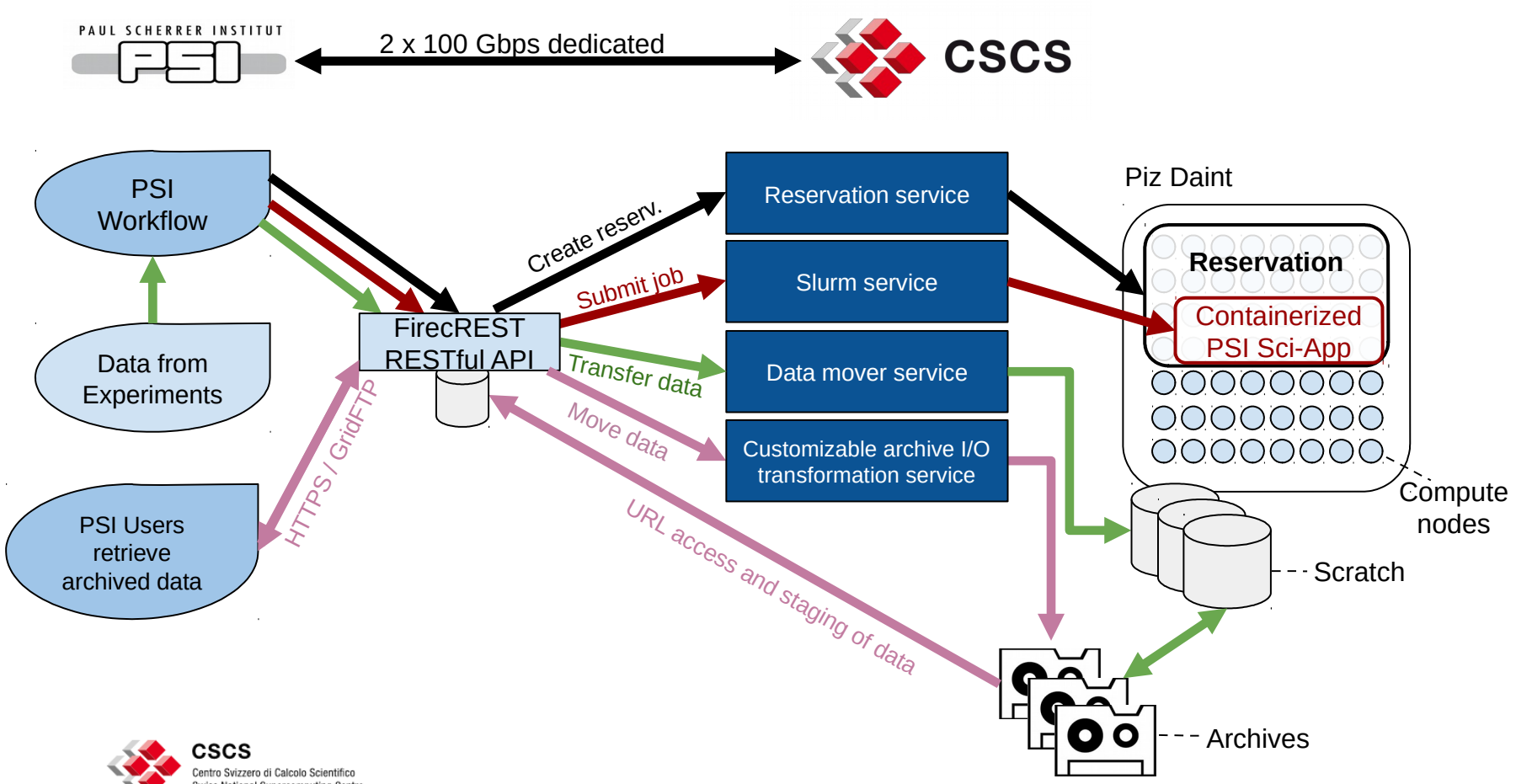
- Massive growth of data
  - 4-5 PB of data in 2018
  - 10-20 PB expected in 2019-2023
  - ~100 PB in 2024
- Require large IT infrastructure to process the data
  - Large number of compute nodes
  - Use of accelerator

PAUL SCHERRER INSTITUT

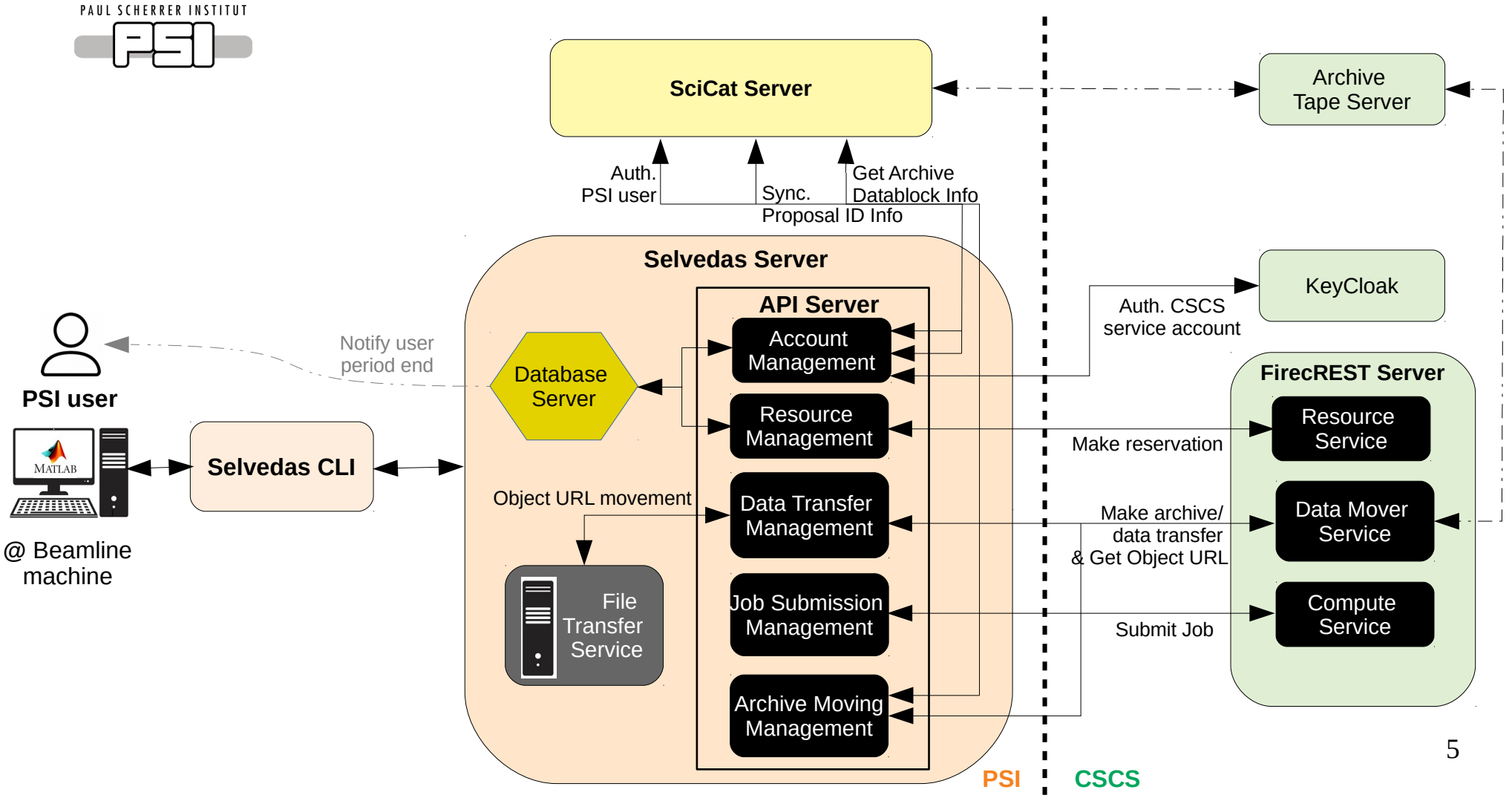
SELVEDAS Project

- Goal
  - Develop scalable and extensible data services (management, processing and analysis) on top of HPC and Cloud at CSCS for computation, storage and networking to support PSI scientists' remote experiments.

Architectural Design – Hybrid Cloud



Architectural Design – Data Catalog Extension



Architectural Design - Cross-site Authentication

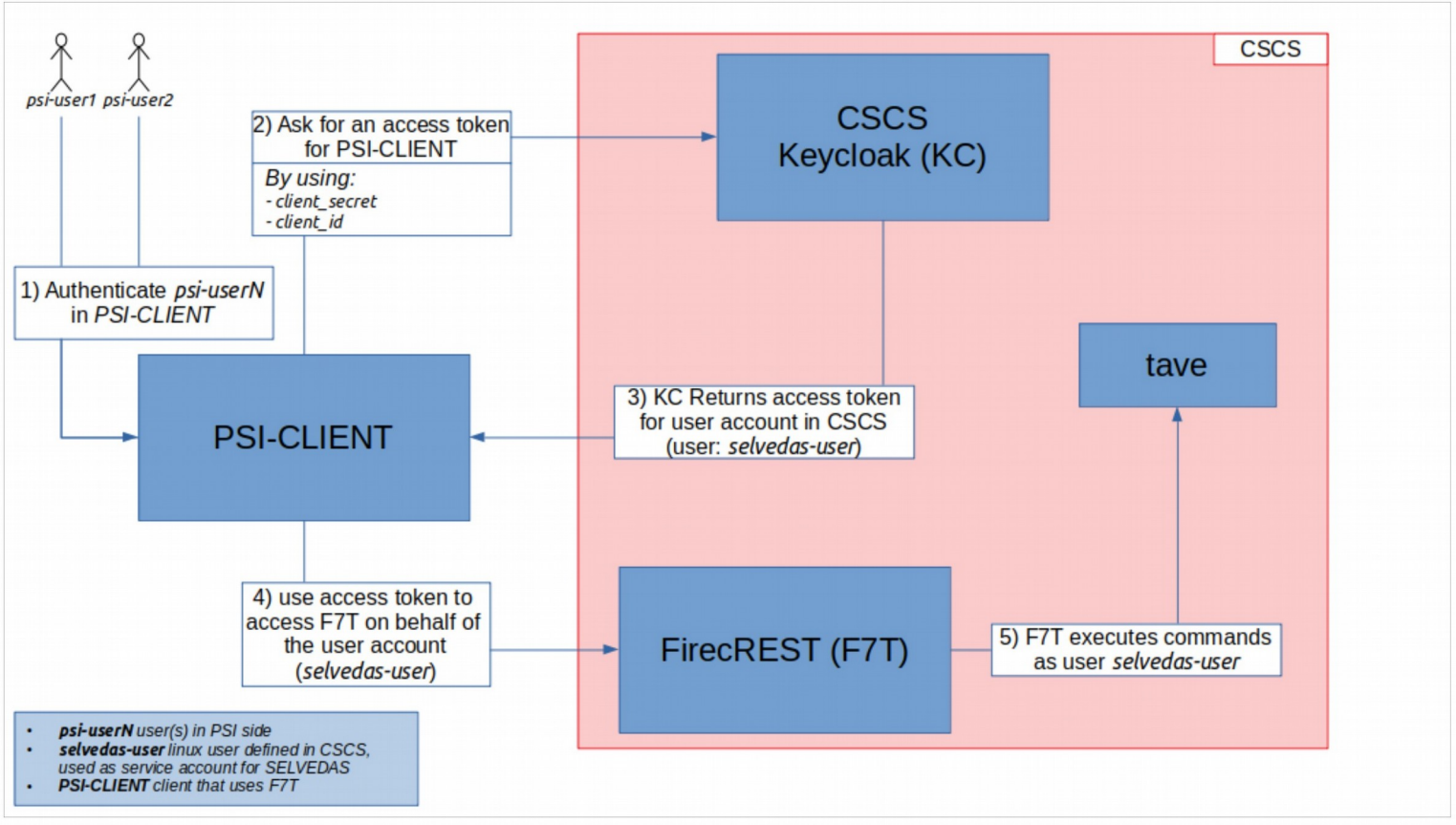


Figure 3. Cross-site authentication diagram

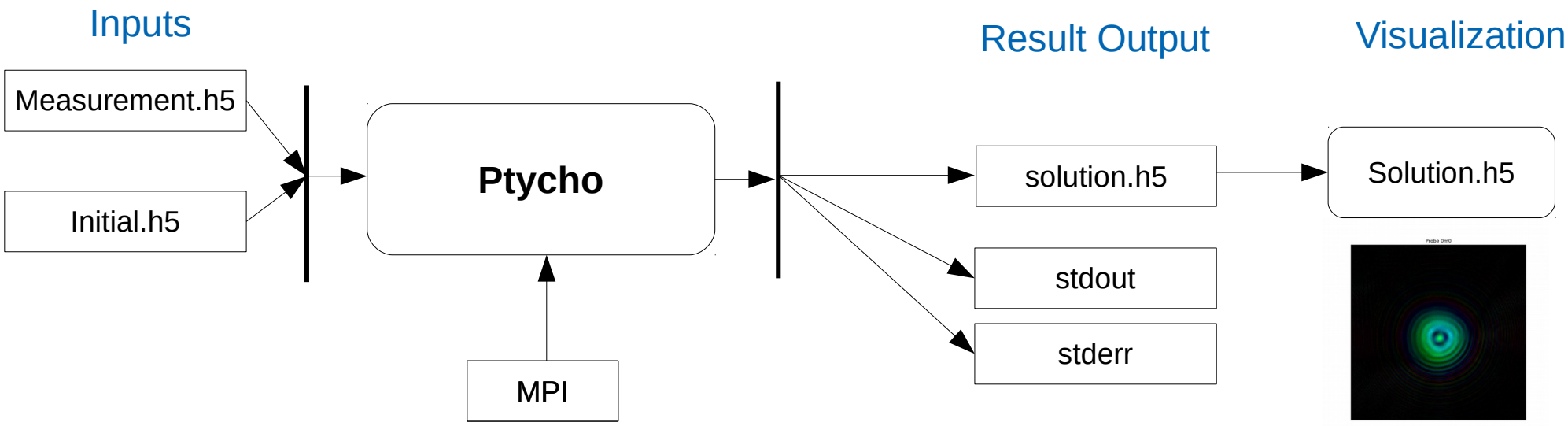
Use cases

UC1:
Analysis of archived data

UC2:
Fast feedback experiment

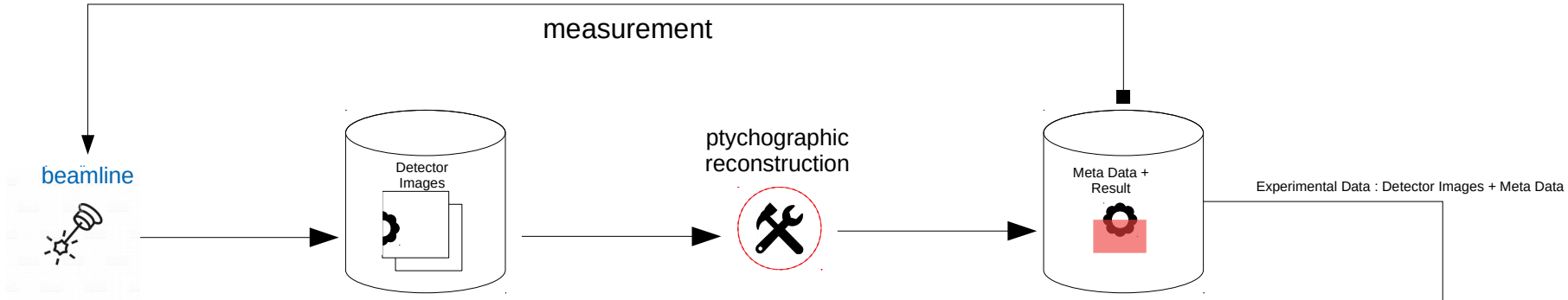
UC3:
Retrieve archived data through a portal

Use case 2 – ex. beamline cSAXS



Ptycho: implemented ptychorgrahic reconstruction algorithm includes “Difference Map” & “Maximum Likehood” with GPU supported (MPI)

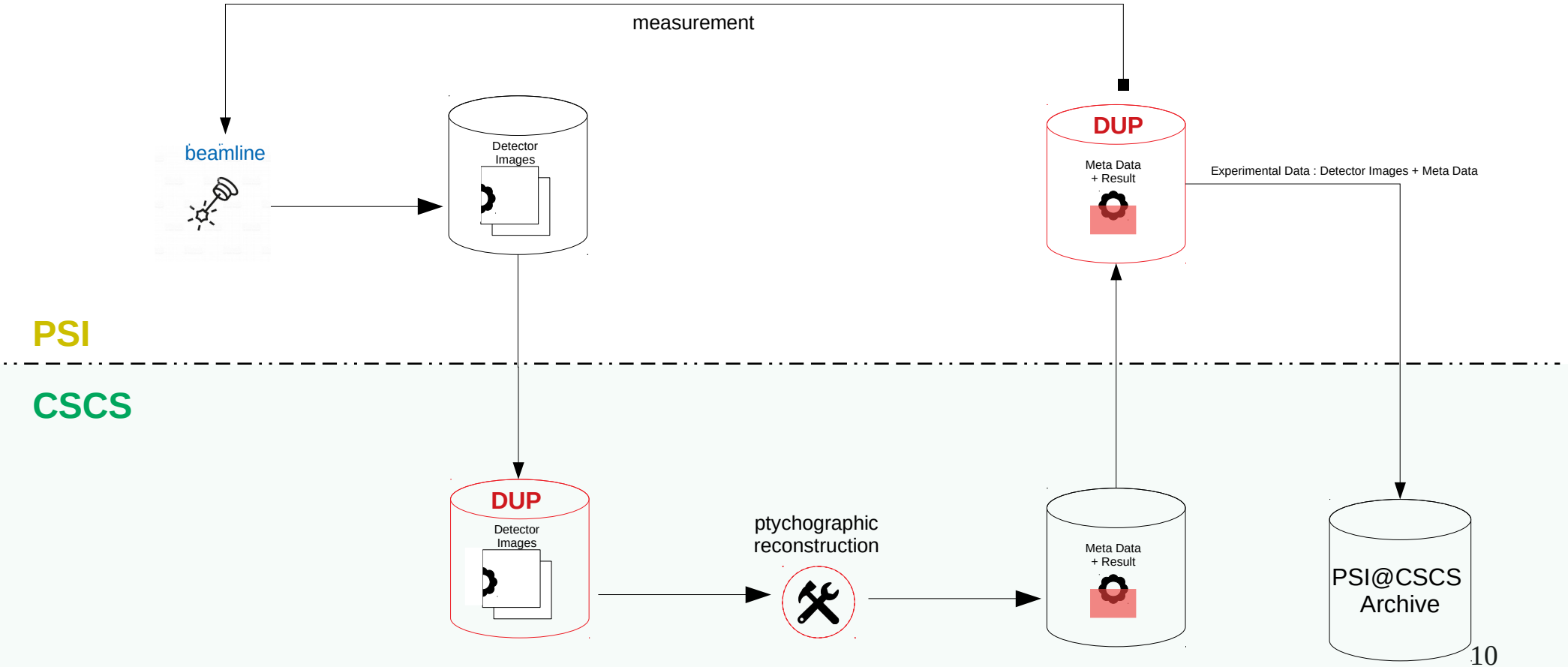
Current cSAXS Workflow



PSI

CSCS

New cSAXS Workflow



cSAXS workflow batch file

```
#!/bin/bash

# Step 1. Authentication token
time TOKEN=$(python selvdscli.py auth -n name -p password -g group)
#echo "Auth Token:"
#echo $TOKEN

# Step 2. Upload measurements_file
FILE1="measurement_f2_S01308.h5"
SRC="/home/l_chang_m/"
DST="/xxxx/mchang"
time RES1=$(python selvdscli.py transfer -i -f $FILE1 -s $SRC -d $DST -t $TOKEN)
#echo $RES1

# Step 3. Upload initial_conditions_file
FILE2="initial_conditions_S01308.h5"
time RES2=$(python selvdscli.py transfer -i -f $FILE2 -s $SRC -d $DST -t $TOKEN)
#echo $RES2

# Step 4. Upload program with tar file
FILE3="ptycho-gpu-mpi-avx-docker.tar"
time RES3=$(python selvdscli.py transfer -i -f $FILE3 -s $SRC -d $DST -t $TOKEN)
#echo $RES3

# Step 5. job submission
FILE4="run-hybrid-gpu6.sh"
time RES4=$(python selvdscli.py job -f $FILE4 -s $SRC -t $TOKEN -D)
#echo $RES4

# Step 6. download result file
FILE5="solution-1.h5"
SRC_d="/xxxx/mchang"
DST_d="/home/l_chang_m/"
time RES5=$(python selvdscli.py transfer -o -f $FILE5 -s $SRC_d -d $DST_d -t $TOKEN)
#echo $RES5
```

Job submission batch file

```
#!/bin/bash -l
sarus load /xxxx/mchang/ptycho-omp-mpi-avx-podman.tar ptycho-omp-mpi-avx:0.1

#SBATCH --job-name=csaxs-test-job
#SBATCH --output=csaxs-test-job-%j.out
#SBATCH --error=csaxs-test-job-%j.err
#SBATCH –cpus-per-task=16

module load sarus

DEBUG_FLAGS=$((64+256))

echo "Start: $(date)"
srun sarus run --mpi --mount=type=bind,source=$SCRATCH/mchang/,destination=$SCRATCH/mchang/
load/library/ptycho-omp-mpi-avx:0.1 /usr/bin/ptycho –debug_flags=$DEBUG_FLAGS
$SCRATCH/mchang/{measurement_f2_S01308.h5, initial_conditions_S01308.h5, solution-1.h5}
echo "Finish: $(date)"
```

Performance Report

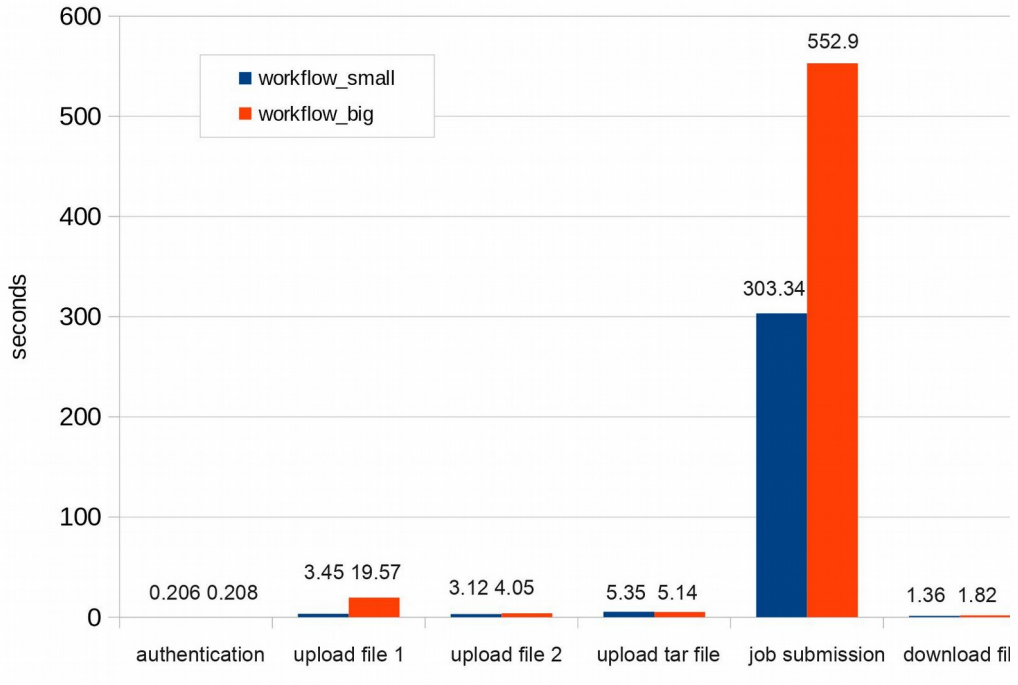
workflow_small

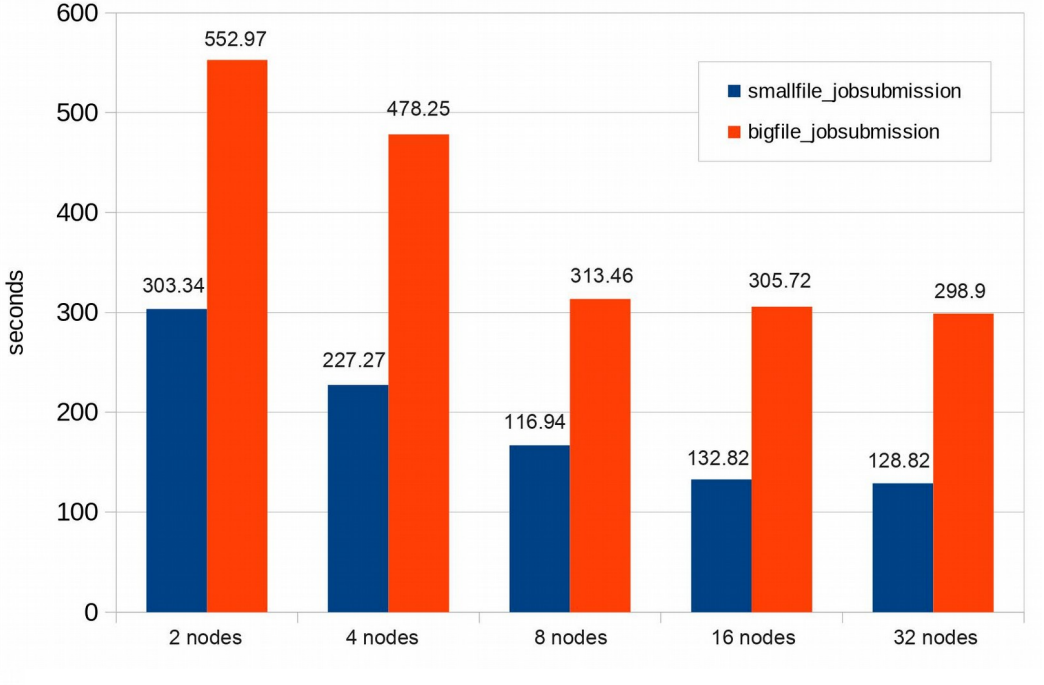
Small input files	Size
measurement_f2_S01308.h5	207MB
initial_conditions_S01308.h5	1.53MB
ptycho-gpu-mpi-avx-docker.tar	294MB
solution-1.h5	1.89MB

workflow_big

Big input files	Size
S00035_data_1000x1000.h5	2.1G
S00035_initial_conditions_1000x1000.h5	139 M
ptycho-gpu-mpi-avx-docker.tar	294MB
Solution-0.h5	64MB

Performance Report





On demand Service – advanced resource reservation

- Create a reservation service to reserve computation nodes

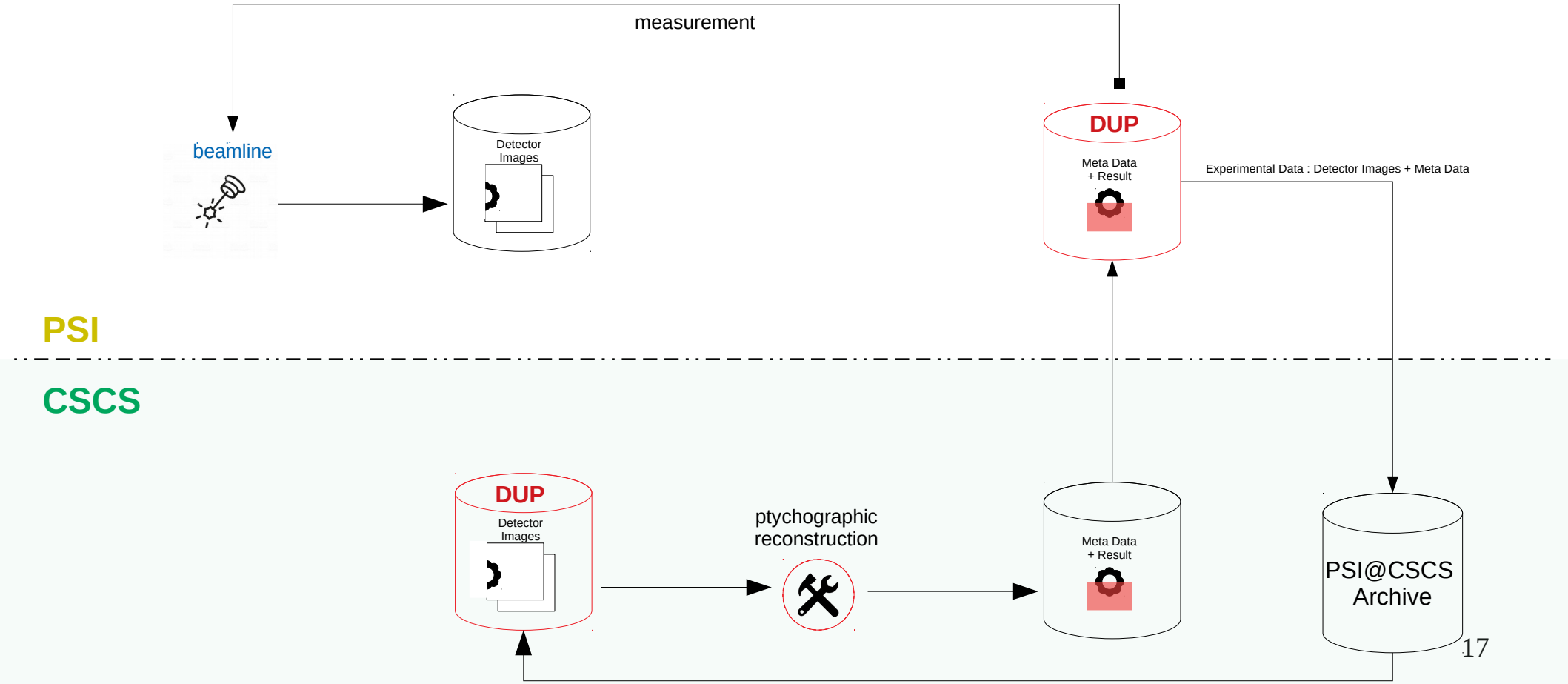
```
# Step1. Authentication token
time TOKEN=$(python selvdscli.py auth -n name -p password -g groupid)
#echo "Auth Token:"
#echo $TOKEN

# Step2. create a reservation
time RES1=$(python selvdscli.py reservation create -p 20.500.11935/20110285 -n 4 -y knl -d test/ -t $TOKEN)
#echo $RES1

# (Step3. update a reservation)
time RES2=$(python selvdscli.py reservation update -p 20.500.11935/20110285 -n 4 -y knl -s 2020-09-29T19:22:00
-e 2020-09-30T19:22:00 -d test/ -t $TOKEN)
```

UC1 & UC3

UC1:
Analysis of archived data

UC3:
Retrieve archived data through a portal

New cSAXS Workflow with Archive data



cSAXS archive move batch file

```
#!/bin/bash

# Step 1. Authentication token
time TOKEN=$(python selvdscli.py auth -n name -p password -g group)
#echo "Auth Token:"
#echo $TOKEN

# Step 2 move a dataset
time RES1=$(python selvedas_cli.py move archive -i 20.500.11935/7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97 -d test/data -t $TOKEN)
#echo $RES1

# Step 3 wait for the result
time RES2=$(python selvedas_cli.py move status -i 20.500.11935/7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97 -a -t $TOKEN)
#echo $RES2
```

cSAXS get archive URL batch file

```
#!/bin/bash

# Step 1. Authentication token
time TOKEN=$(python selvdscli.py auth -n name -p password -g group)
#echo "Auth Token:"
#echo $TOKEN

# Step 2 get datablock URL
time RES1=$(python selvedas_cli.py move cloud -i 20.500.11935/7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97 -d test/data -t $TOKEN)
#echo $RES1

# Step 3 wait for the result
time RES2=$(python selvedas_cli.py move status -i 20.
```

cloud url at /scratch/snx2000/psisat/slstomcat/test/data/*.tar

email

noreply-selvedas@psi.ch

Wed 14/10/2020 07:42

To: Chang Mei-Chih (PSI) <mei-chih.chang@psi.ch>;

7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97_0_2020-01-29-16-12-30.tar:
https://object.cscs.ch/v1/AUTH_3433719da17e479181ca702e5aa53c28/psisat/36b65ab7fe8fe2be1c9ac0c57a748466/7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97_0_2020-01-29-16-12-30.tar?temp_url_sig=dde06beb3b0c9f89848e8a912e2db3c06466b6db&temp_url_expires=1605246141
7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97_1_2020-01-29-16-12-31.tar:
https://object.cscs.ch/v1/AUTH_3433719da17e479181ca702e5aa53c28/psisat/5bbd430cab8afde7e0b8ea57df3e7948/7f8adb54-6abe-4c02-a1ad-c9b4a6d60f97_1_2020-01-29-16-12-31.tar?temp_url_sig=c14fadac0f1ab30562e066b02106f7e70b91a231&temp_url_expires=1605246147

demo

- Three parts
  - 1st : Run workflow_small_new.sh from my local laptop by VPN at PSI.
  - 2nd : Resource reservation with 8 compute nodes and run 8 job submissions at the same time to check all results.
  - 3rd: Archive move service

SELVEDAS TEAM

PAUL SCHERRER INSTITUT

- Alun Ashton (Project co-lead)
- Mei-Chih Chang
- Hans-Christian Stadler
- Leonardo Sala
- Stephan Egli
- Peter Huesser
- Karel Stadler

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

- Cerlane Leong (Project lead)
- Juan Pablo Dorsch
- Tomas Aliaga
- Mario Valle
- Sadaf Alam
- Andreas Jocksch
- Maxime Martinasso
- Vasileios Karakasis
- Mark Klein
- Guy-Mael Horclois Le Pironnec

PAUL SCHERRER INSTITUT

PSI

Thank you!

Q & A