



Swiss Off-site Data Centre

Andrii Neronov, Volodymyr Savchenko 

Pablo Fernandez, Victor Holanda Rusu  **CSCS**
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

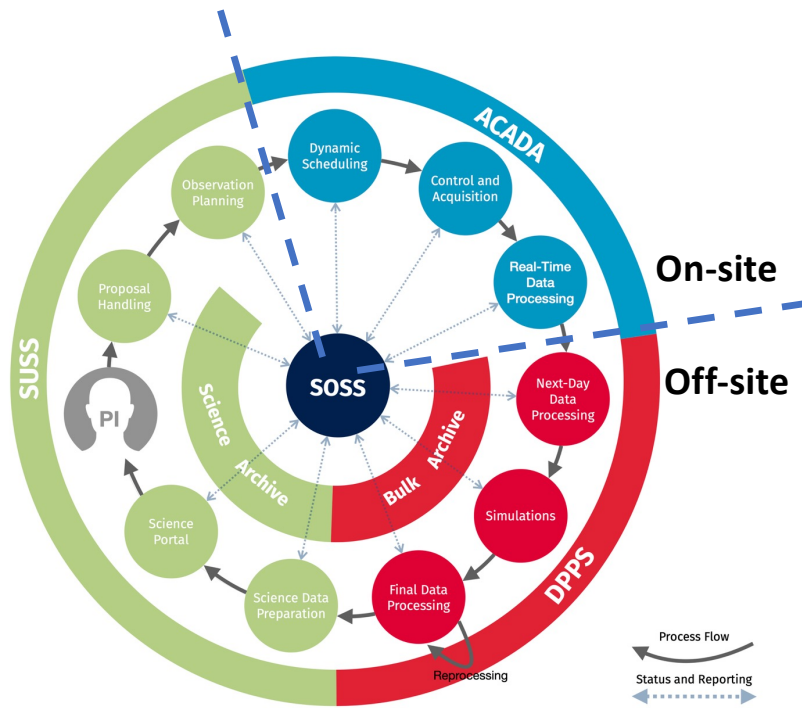
Swiss Off-site Data Centre

Andrii Neronov, Volodymyr Savchenko 

Pablo Fernandez, Victor Holanda Rusu  **CSCS**
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

- Context, kick-off 2022, organisation of work
- Computing environment at CSCS
- DPPS: Bulk archive, Monte-Carlo production
- Science Data Challenge
- Data management for the Large Size Telescope prototype
- Synergies with SKA Regional Center

Context



Data from 2 observatory sites will be processed at Off-site Data Centers

- **CSCS (Switzerland)**
- PIC (Spain)
- DESY-Zeuthen (Germany)
- INAF (Italy)

+ Orchestration and monitoring of services from Science Data Management Center (SDMC) at DESY Zeuthen

- Proposal Handling
- Long-Term and Mid-Term Scheduling
- Automatic Data Product Preparation and Verification
- Science Analysis Tools
- Science Archive
- Science Portal
- Reporting/Diagnosis

- Operations Management System
- Computing Workload Management System
- Bulk Archive and File Transfer Management System
- Data Processing Pipeline System
- Calibration Production Pipeline System
- Data Quality Pipeline System
- Simulation Production Pipeline System
- Common Software Frameworks

Context

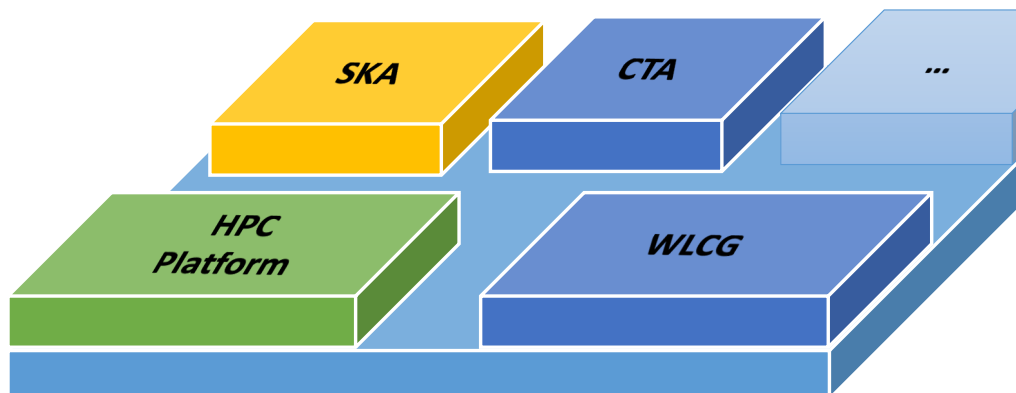


- CSCS Director: *“We are committed to support world-class astronomy in Switzerland. In collaboration with our CTA and SKA partners we will devise solutions for computing and data that best serve the evolving needs of science.”*
- Swiss off-site DC for CTA will run on a new ALPS platform (now in construction), sharing resources with other projects, such as SKA.
- The idea is to exchange expertise and technology between CTA and SKA data management activities in the measure of possible.
- Collaborative effort between CSCS and EPFL.

Context



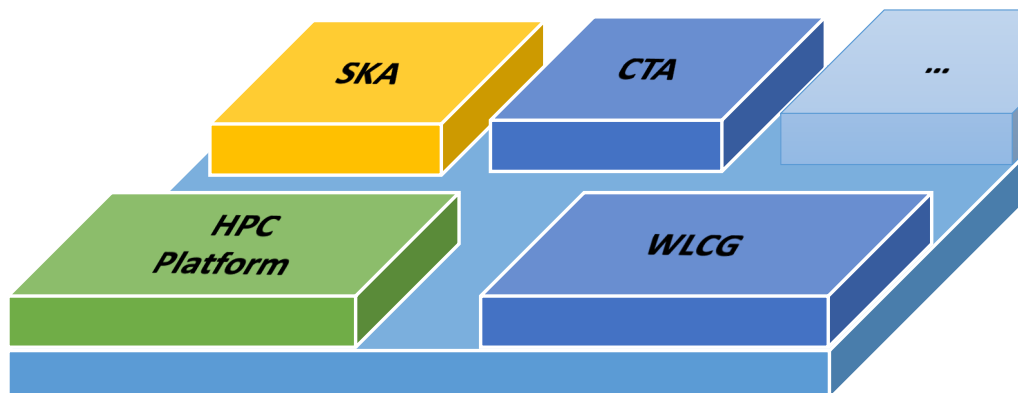
- CSCS Director: *"We are committed to support world-class astronomy in Switzerland. In collaboration with our CTA and SKA partners we will devise solutions for computing and data that best serve the evolving needs of science."*
- Swiss off-site DC for CTA will run on a new ALPS platform (now in construction), sharing resources with other projects, such as SKA.
- The idea is to exchange expertise and technology between CTA and SKA data management activities in the measure of possible.
- Collaborative effort between CSCS and EPFL.



Organisation of work



- CSCS Director: *“We are committed to support world-class astronomy in Switzerland. In collaboration with our CTA and SKA partners we will devise solutions for computing and data that best serve the evolving needs of science.”*
- Swiss off-site DC for CTA will run on a new ALPS platform (now in construction), sharing resources with other projects, such as SKA.
- The idea is to exchange expertise and technology between CTA and SKA data management activities in the measure of possible.
- Collaborative effort between CSCS and EPFL.

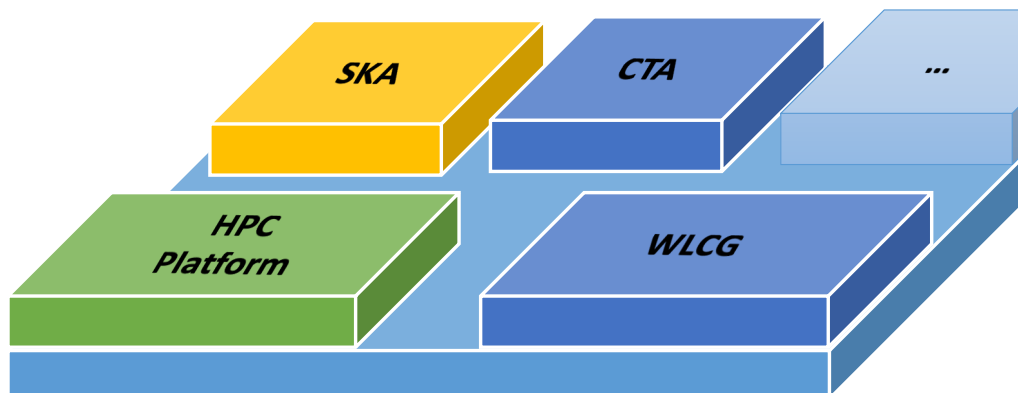


- Proposal to SERI together with SKA, contracts, approval and budget discussions
- Setup of the team
 - Resourcing with existing people, open new positions
 - Admin tasks like open accounts, introduce people, setup collaboration spaces, Sync meetings
- Engineering kick-off meeting March 28, 2022 at CSCS
- Integration into Off-site ICT workpackage:
 - technical info from CSCS to Off-site ICT management
 - visit Off-site ICT to CSCS
 - Face-to-face meeting of four Off-site DCs in Annecy

Organisation of work



- CSCS Director: *“We are committed to support world-class astronomy in Switzerland. In collaboration with our CTA and SKA partners we will devise solutions for computing and data that best serve the evolving needs of science.”*
- Swiss off-site DC for CTA will run on a new ALPS platform (now in construction), sharing resources with other projects, such as SKA.
- The idea is to exchange expertise and technology between CTA and SKA data management activities in the measure of possible.
- Collaborative effort between CSCS and EPFL.

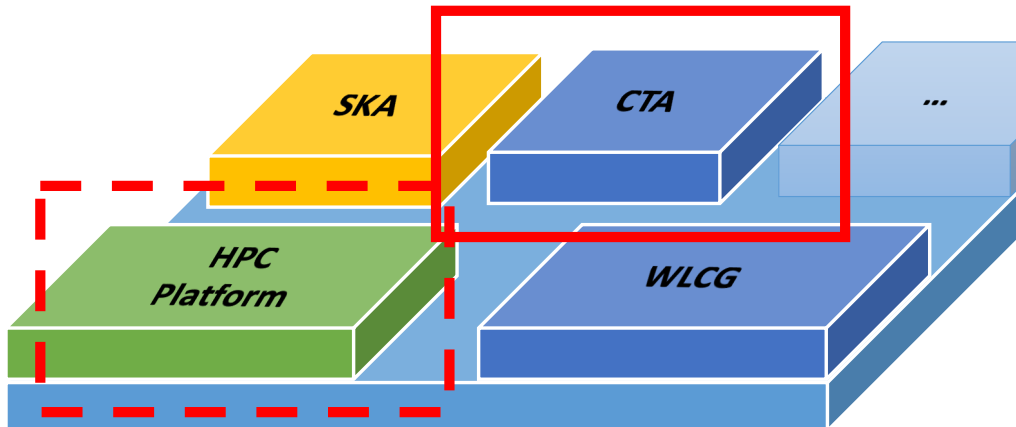


- Proposal to SERI together with SKA, contracts, approval and budget discussions
- Setup of the team
 - Resourcing with existing people, open new positions
 - Admin tasks like open accounts, introduce people, setup collaboration spaces, Sync meetings
- Engineering kick-off meeting March 28, 2022 at CSCS
- Integration into Off-site ICT workpackage:
 - technical info from CSCS to Off-site ICT management
 - visit Off-site ICT to CSCS
 - Face-to-face meeting of four Off-site DCs in Annecy
- Definition of roles inside the WP
 - CSCS: compute & storage on Alps, ARC, dCache, etc.
 - EPFL: link with CTAO (DPPS, SUSS, Science Data Challenge)

Organisation of work

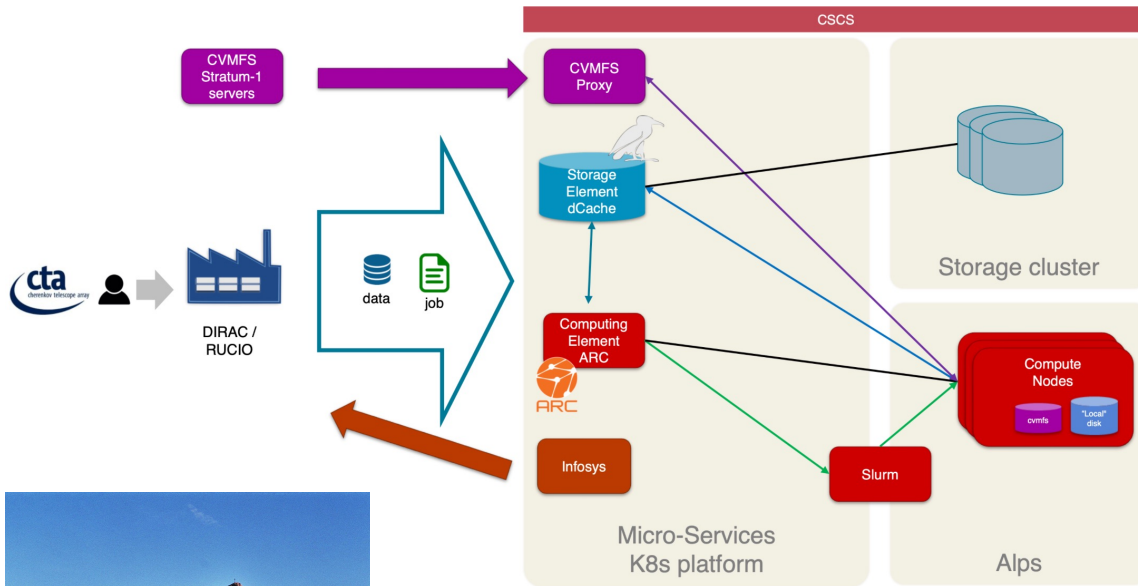


- CSCS Director: *“We are committed to support world-class astronomy in Switzerland. In collaboration with our CTA and SKA partners we will devise solutions for computing and data that best serve the evolving needs of science.”*
- Swiss off-site DC for CTA will run on a new ALPS platform (now in construction), sharing resources with other projects, such as SKA.
- The idea is to exchange expertise and technology between CTA and SKA data management activities in the measure of possible.
- Collaborative effort between CSCS and EPFL.



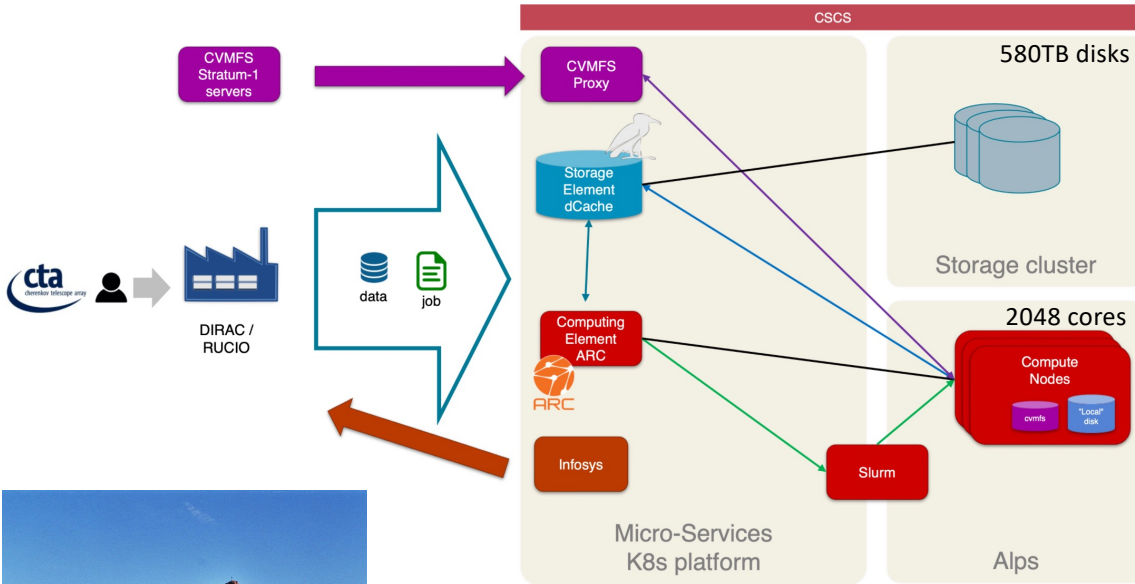
- Proposal to SERI together with SKA, contracts, approval and budget discussions
- Setup of the team
 - Resourcing with existing people, open new positions
 - Admin tasks like open accounts, introduce people, setup collaboration spaces, Sync meetings
- Engineering kick-off meeting March 28, 2022 at CSCS
- Integration into Off-site ICT workpackage:
 - technical info from CSCS to Off-site ICT management
 - visit Off-site ICT to CSCS
 - Face-to-face meeting of four Off-site DCs in Annecy
- Definition of roles inside the WP
 - CSCS: compute & storage on Alps, ARC, dCache, etc.
 - EPFL: link with CTAO (DPPS, SUSS, Science Data Challenge)

Setup of computing environment at CSCS



- Setup of Noir/ALPS virtual cluster
- 16 nodes (2048 cores), 500 TB of storage
 - Middleware: dCache and ARC
 - Other services CVMFS, DIRAC
 - Access through Grid certificates

Monte-Carlo production



Noir/ALPS virtual cluster

- 16 nodes (2048 cores), 500 TB of storage
- Middleware: dCache and ARC
- Other services CVMFS, DIRAC
- Access through Grid certificates

Initially prepared for “Prod6” Monte-Carlo production

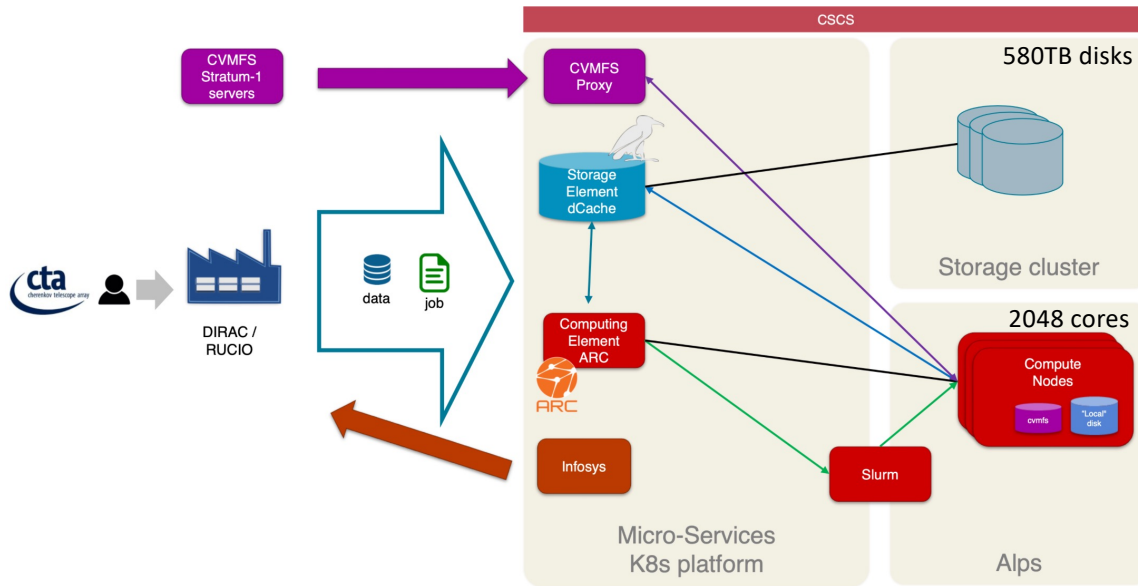
- Request from production team in April, with projected start mid-May
 - Shifted to July, August, September
 - Finally postponed till mid-2023

Alternative Monte-Carlo production proposed (to run on two Off-site DCs: CSCS and PIC), for “divergent pointings” (12 million HS06.hours, 30 TB per DC)

- Starts on Dec. 13, running for 2+ months

Global CTAO requirements		2022	DC proposed pledge	Availability
CTAO	CPU [HS06.hours]	230 000 000		
	CPU peak value [Nb of cores]	12 000		
	Disk [TB]	2 000		
	Tape [TB]	0		

Data management for Large Size Telescope



Noir/ALPS virtual cluster

- 16 nodes (2048 cores), 500 TB of storage
- Middleware: dCache and ARC
- Other services CVMFS, DIRAC
- Access through Grid certificates

Invitation from LST team to join data management for the telescope

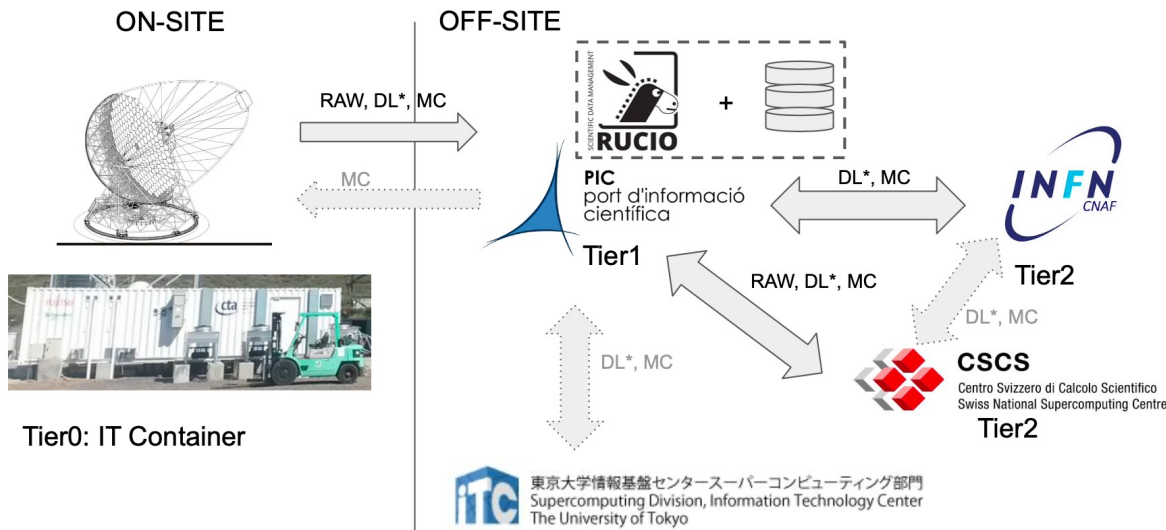
- Agreement within CTA-CH collaboration to enable science use case with real data
- Agreement with SERI that EPFL joins LST collaboration
- Collaboration with Spanish Off-site DC (PIC)

LST1

Year	2021	2022	2023	2024	2025	2026
DLO data flow per year North site (TB)	1390	1765	1477	1044	817	
DLO data flow per year South site (TB)	0	0	208	625	895	
DLO data flow per year (TB)	1 390	1 765	1 684	1 669	1 712	
Cumulative DLO data (TB)	1 390	3 156	4 840	6 509	8 221	

OFF-SITE IC-INFRA
Data Centre Model Implementation
Cost Book explanations
V1.0

Data management for Large Size Telescope



Noir/ALPS virtual cluster

- 16 nodes (2048 cores), 500 TB of storage
- Middleware: dCache and ARC
- Other services CVMFS, DIRAC
- Access through Grid certificates

Invitation from LST team to join data management for the telescope

- Agreement within CTA-CH collaboration to enable science use case with real data
- Agreement with SERI that EPFL joins LST collaboration
- Collaboration with Spanish Off-site DC (PIC)

Currently on-going / planned activities:

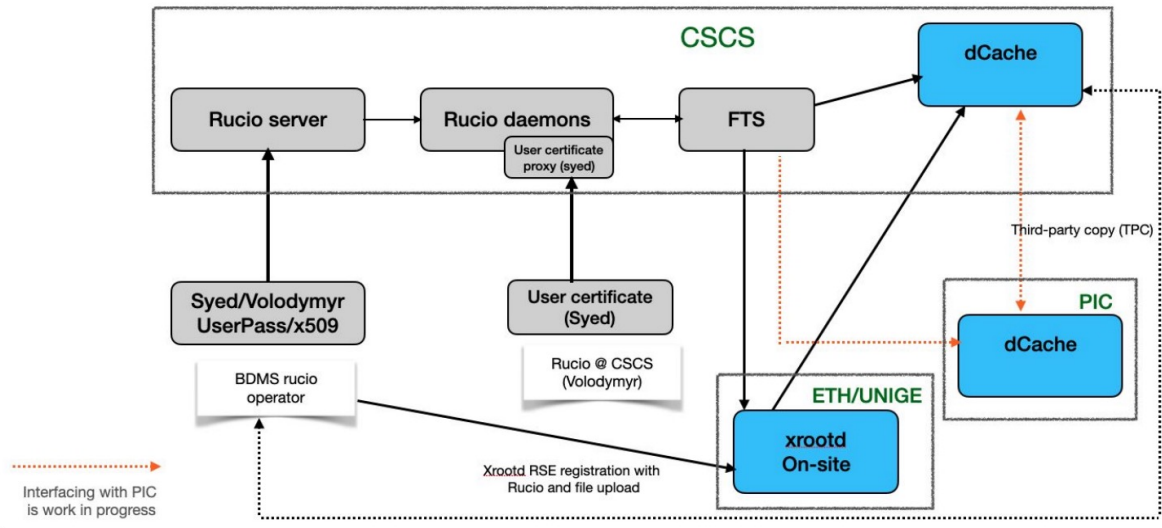
- Store a copy of raw data (currently 1.5 PB, +1 PB/yr)
- Post-processing of MC data (using Noir cluster)
- Regular re-processing of all data (using Noir cluster)
- Develop services to Swiss members of LST
- Synergies with CTA development: RUCIO (Bulk Archive), DIRAC (workflow management system)

LST1

Year	2021	2022	2023	2024	2025	2026
DLO data flow per year North site (TB)	1390	1765	1477	1044	817	
DLO data flow per year South site (TB)	0	0	208	625	895	
DLO data flow per year (TB)	1 390	1 765	1 684	1 669	1 712	
Cumulative DLO data (TB)	1 390	3 156	4 840	6 509	8 221	

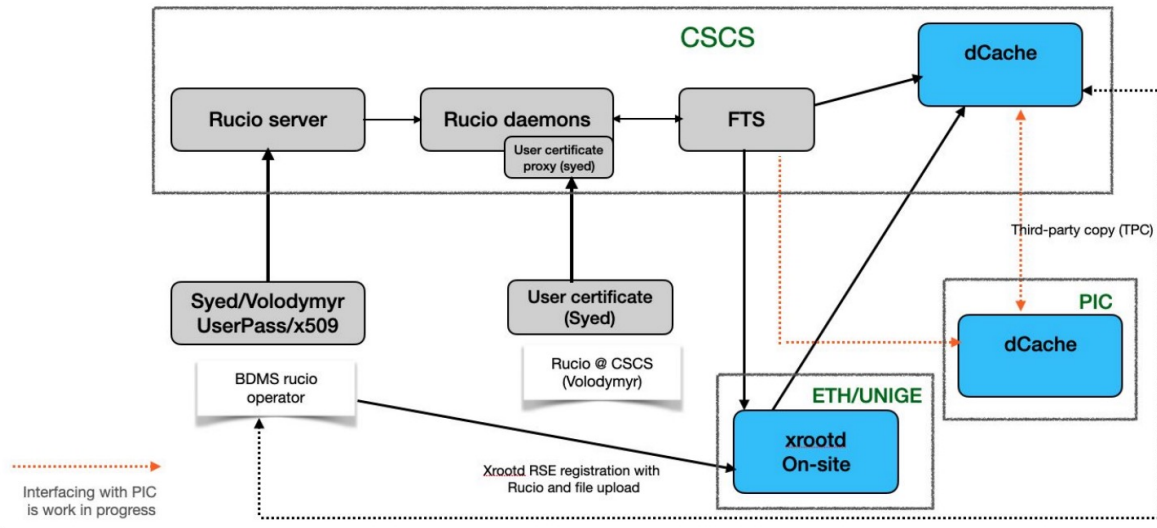
OFF-SITE IC-INFRA
Data Centre Model Implementation
Cost Book explanations
 V1.0

Bulk Data Management System, Workflow Management System technologies



- Assistance to BDMS prototyping in collaboration with DPPS team at UNIGE and ETHZ
- Deployment of RUCIO at CSCS (on Kubernetes on CSCS OpenStack cluster, not related to Noir)
- Access through Grid certificates found to be cumbersome.
- Collaboration with SKA team at EPFL (also considering to use RUCIO for their Regional Centers) on replacement of Grid certificates with tokens
- Collaboration with Spanish Off-site DC on RUCIO-managed data transfer

Bulk Data Management System, Workflow Management System technologies



- Assistance to BDMS prototyping in collaboration with DPPS team at UNIGE and ETHZ
- Deployment of RUCIO at CSCS (on Kubernetes on CSCS OpenStack cluster, not related to Noir)
- Access through Grid certificates found to be cumbersome.
- Collaboration with SKA team at EPFL (also considering to use RUCIO for their Regional Centers) on replacement of Grid certificates with tokens
- Collaboration with Spanish Off-site DC on RUCIO-managed data transfer
- DPPS team has participated in a Kubernetes hackathon at CSCS, developing "kubernetesisation" of DIRAC, a candidate Workflow Management System.

Events | Upcoming Events

28.11.2022-02.12.2022

EUROCC PROJECT - KUBERNETES HACKATHON

📍 Hotel Dante

💰 Free

The Swiss National Supercomputing Centre (CSCS), in collaboration with the **EuroCC Project**, is pleased to announce a **Kubernetes Hackathon**.

Science Data Challenge



- CTA collaboration plans a “Science Data Challenge” (SDC) in 2023.
- Off-site DCs have been invited to contribute to generation of mock CTA data for the SDC

Science Data Challenge



ESCAPE

ESAP MOA Query

ESAP MOA Query

SELECT TOP 100 FROM vobastore

Service Metadata

Interactive Analysis Workflows

Search for Workflows

Advanced Search

CTA Geminny Example

Description: CTA Geminny Example

Author: [Name]

Runtime Platform: Python

Keywords: jupyter-notebook

Results from http://voparis-tap-astro.obspm.fr/_system_/tap/run/tap

Record	dataproduct_type	dataproduct_subtype	data_level	obs_collection	obs_id	obs_sra	obs_publication_id	obs_creator_id	access_url	access_remote	access_remote	target_name	target_id
1	spectrum	1	raw_spectra	raw_spectra	1	1	1	1	1	1	1	1	1
2	spectrum	2	raw_spectra	raw_spectra	2	2	2	2	2	2	2	2	2
3	spectrum	3	raw_spectra	raw_spectra	3	3	3	3	3	3	3	3	3

Interactive Analysis Compute Facilities

Search for Facilities

CSCS CTA BinderHub

Description: CSCS CTA BinderHub

Link: <https://binderhub-cs.cscs.ch/>

Python 3 (system)

Plot 1: Line graph showing flux vs frequency for different parameters.

Plot 2: Heatmap showing correlation between variables.

Plot 3: Line graph showing flux vs frequency for different parameters.

Plot 4: Line graph showing flux vs frequency for different parameters.

Plot 5: Line graph showing flux vs frequency for different parameters.

Plot 6: Line graph showing flux vs frequency for different parameters.

- CTA collaboration plans a “Science Data Challenge” (SDC) in 2023.
- Off-site DCs have been invited to contribute to generation of mock CTA data for the SDC
- Data challenge organisers are considering to use an online data analysis environment based on ESFRI Science Analysis Platform (ESAP) for participants of SDC.
- Swiss Off-site DC team has made first tests of ESAP deployment at CSCS and contributed to ESAP development (Kubernetes deployment)

Science Data Challenge



ESAP MOA Query

```
SELECT TOP 100 FROM voparisastro
```

Service Metadata

- http://voparis-tap-astro.obspm.fr/_system_/tap/run/tap

Selected Services

- http://voparis-tap-astro.obspm.fr/_system_/tap/run/tap

Results from http://voparis-tap-astro.obspm.fr/_system_/tap/run/tap

id	name	description	url
1	spectrum
2

Interactive Analysis Workflows

Search for Workflows

CTA Geminopy Example

Interactive Analysis Compute Facilities

Search for Facilities

CSCS CTA BinderHub

- CTA collaboration plans a “Science Data Challenge” (SDC) in 2023.
- Off-site DCs have been invited to contribute to generation of mock CTA data for the SDC
- Data challenge organisers are considering to use an online data analysis environment based on ESFRI Science Analysis Platform (ESAP) for participants of SDC.
- Swiss Off-site DC team has made first tests of ESAP deployment at CSCS and contributed to ESAP development (Kubernetes deployment)
- Work in progress on Authentication and Authorisation and on the choice of multi-core computing technology to be provided to SDC participants.

Science Data Challenge



The image displays three screenshots of the ESAP (ESFRI Science Analysis Platform) web interface. The top-left screenshot shows the 'Interactive Analysis Workflows' page with a search bar and a list of workflows, including 'CTA Geminopy Example'. The bottom-left screenshot shows the 'Interactive Analysis Compute Facilities' page with a search bar and a list of facilities, including 'CSCS CTA BinderHub'. The right-side screenshot shows a JupyterLab environment with several plots, including line graphs and a heatmap, representing data analysis results.

- CTA collaboration plans a “Science Data Challenge” (SDC) in 2023.
- Off-site DCs have been invited to contribute to generation of mock CTA data for the SDC
- Data challenge organisers are considering to use an online data analysis environment based on ESFRI Science Analysis Platform (ESAP) for participants of SDC.
- Swiss Off-site DC team has made first tests of ESAP deployment at CSCS and contributed to ESAP development (Kubernetes deployment)
- Work in progress on Authentication and Authorisation and on the choice of multi-core computing technology to be provided to SDC participants.
- A proposal EPFL+CSCS is submitted to Swiss Science Data Centre (SDSC) for a data science support of SDC.

