CTAO
Cherenkov Telescope Array Observatory

Cherenkov Telescope Array Observatory (CTAO) report

7th RUCIO Community Workshop, 30th September to 4th October, 2024

San Diego Supercomputer Center, USA

Syed Hasan* (ETH Zurich), shasan@phys.ethz.ch for the DPPS BDMS team & collaborators:
H. Li, E. Lyard, R. Walter (UniGe), S. Gallozi, G. Zacharis (INAF-OAR), F. Gillardo (LAPP), M. Linhoff (TU Dortmund), Volodymr Savchenko (EPFL/CSCS), N. Benekos (CTAO), K. Kosack (CTAO, CEA Paris)

CTAO

1 Cherenkov Telescope Array Observatory (CTAO)

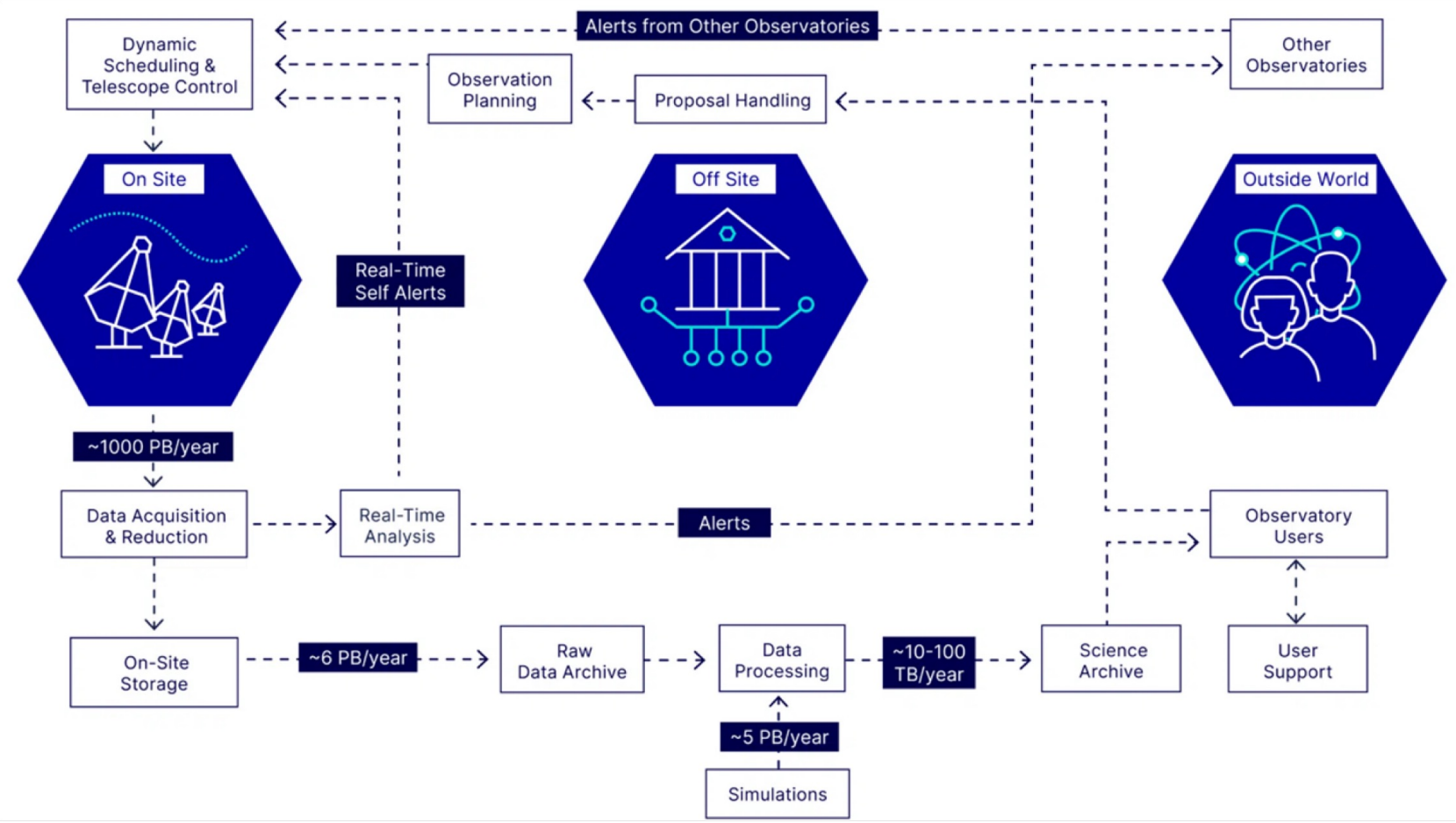
2 Data Processing and Preservation System (DPPS)

3 Role of RUCIO in DPPS – Bulk Data Management System (BDMS)

4 BDMS Prototyping in Docker and Kubernetes (K8s)

5 Conclusion

Cherenkov Telescope Array Observatory (CTAO) - Data Flow

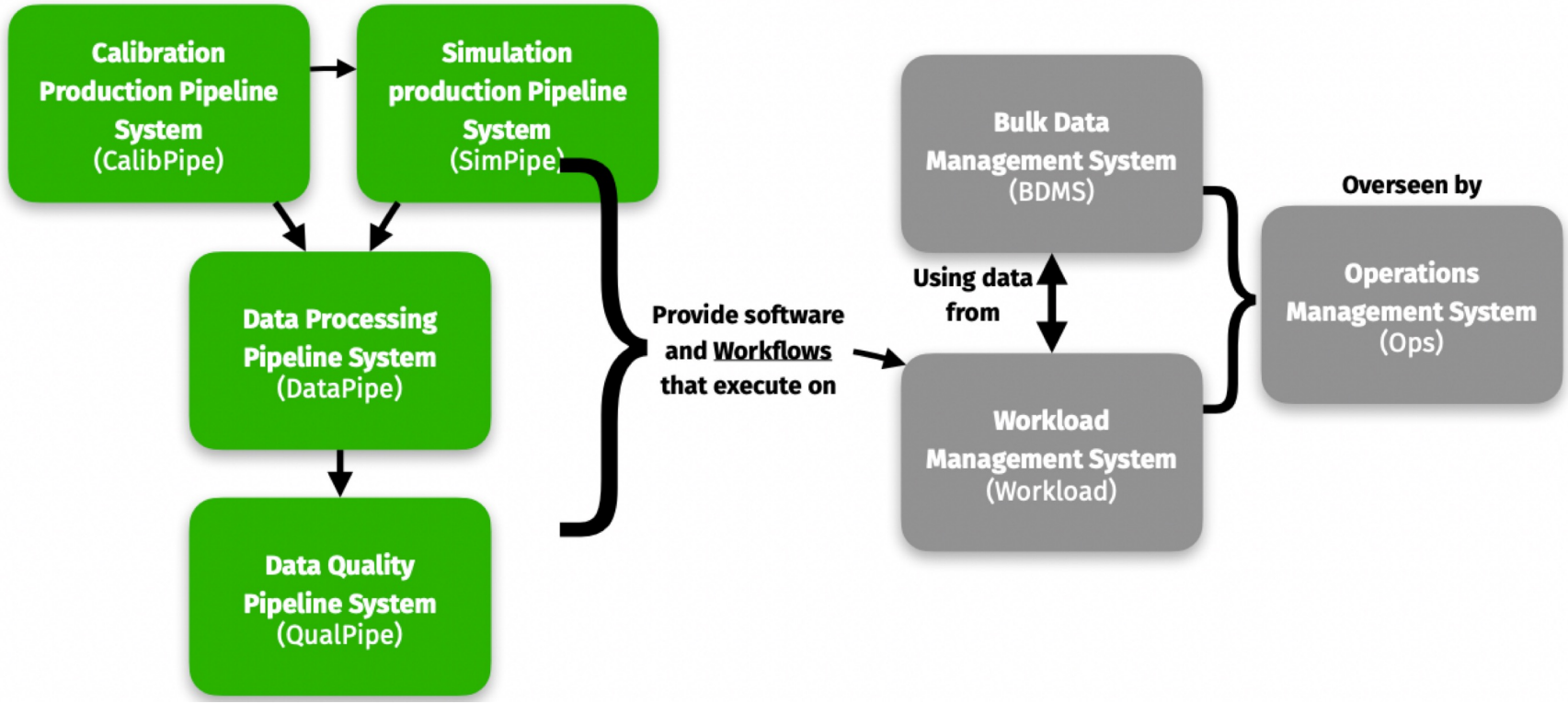


- Control 60+ telescopes
- BIG DATA project, generate hundreds of petabytes (PB) of data in a year (at least 6 PB after compression)
- Two sites and four off-site data centres

Source: Igor Oya (ACADA), The Online data taking system of the Cherenkov Telescope Array Observatory (RICAP 2024)
https://agenda.infn.it/event/35353/contributions/234472/

Data Processing and Preservation System (DPPS) in CTAO

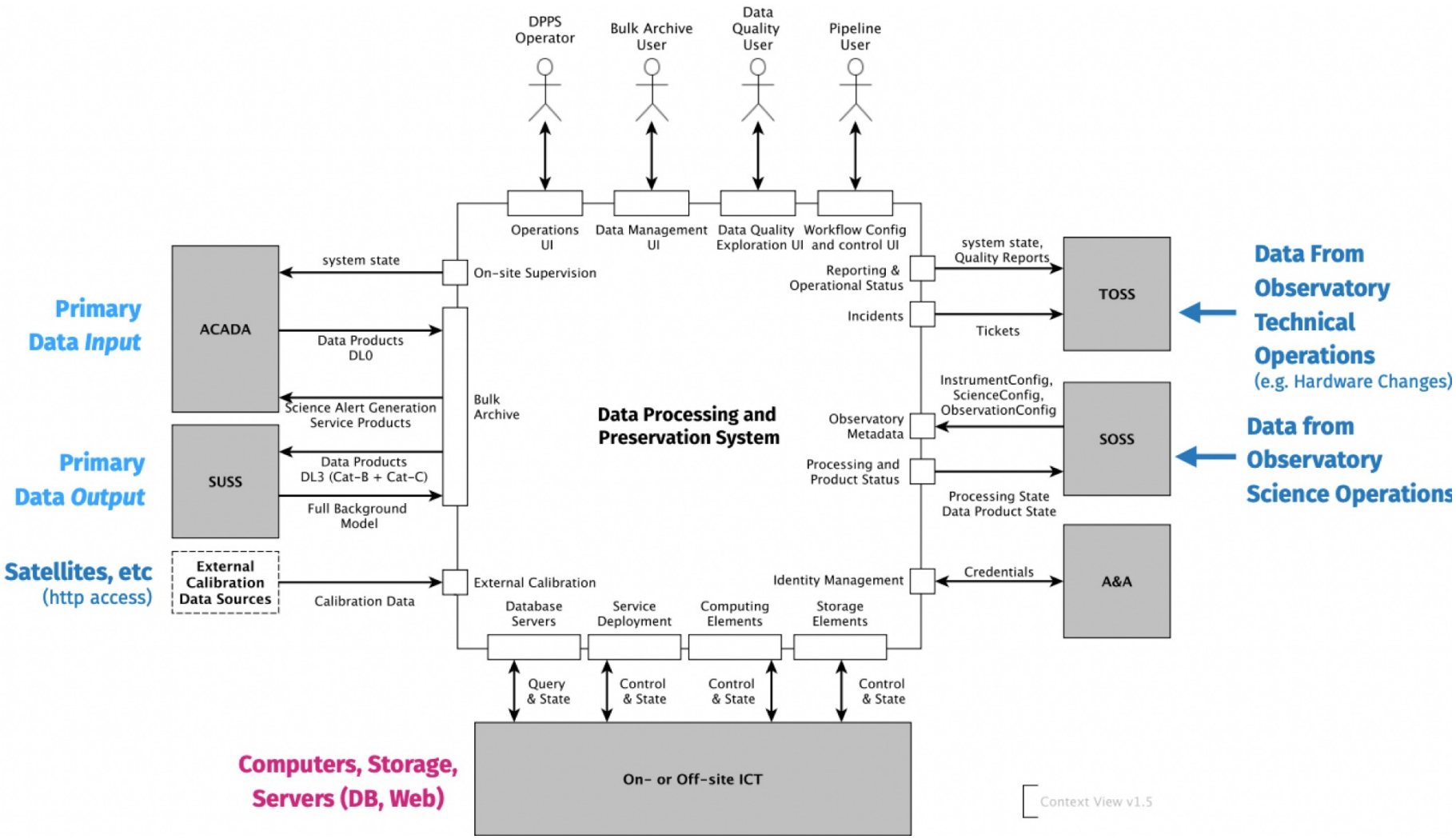
Functional Decomposition to Subsystems



Source: DPPS Introduction, DPPS workshop at DESY (Zeuthen), 10-12th October, 2022

DPPS - Context

- DPPS is a software system responsible for the long-term preservation, (re-) processing, and quality monitoring of low-level data products acquired from real observations or simulated internally
- DPPS also ensures that all required data products are preserved (at least two copies must exist at distinct off-site data centers), traceable and reproducible



Source: DPPS Introduction, DPPS workshop at DESY (Zeuthen), 10-12th October, 2022

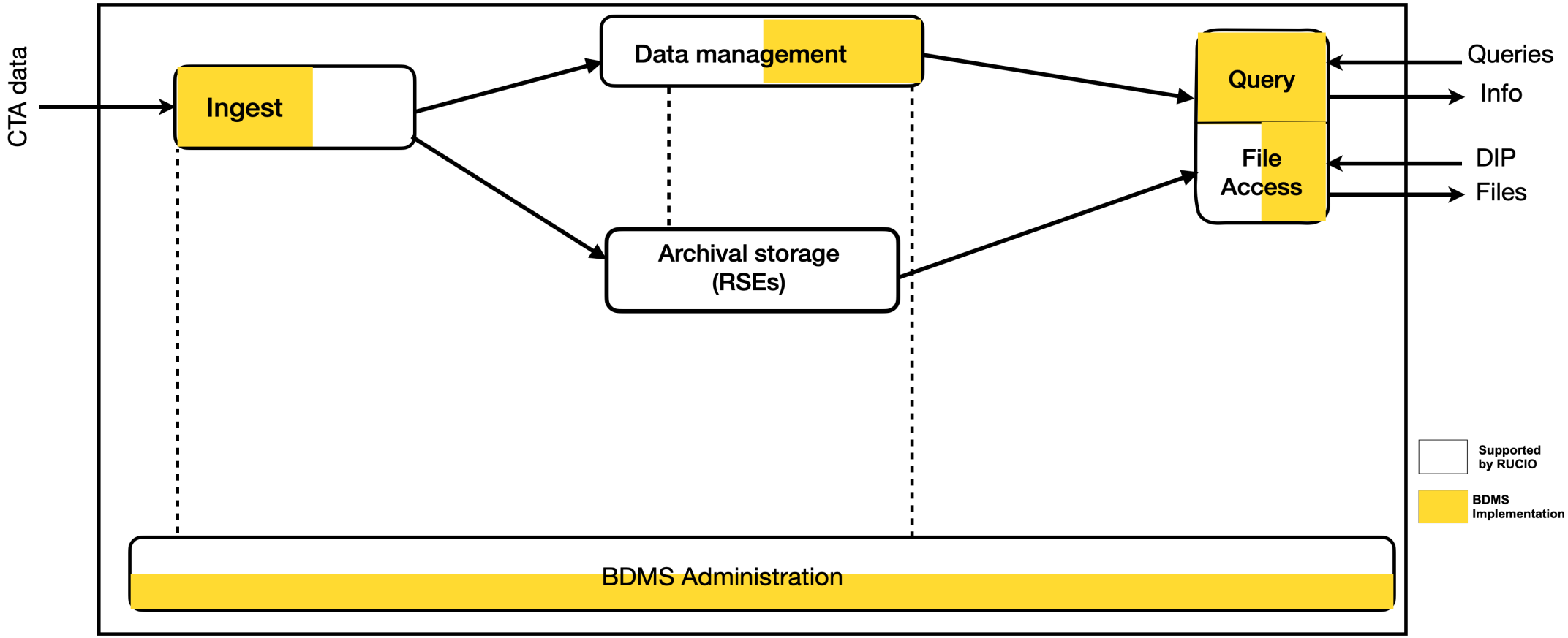
DPPS – Contributors

- Software teams in
  - Switzerland
    - ETH Zurich
    - University of Geneva
  - Italy
    - INAF OAR

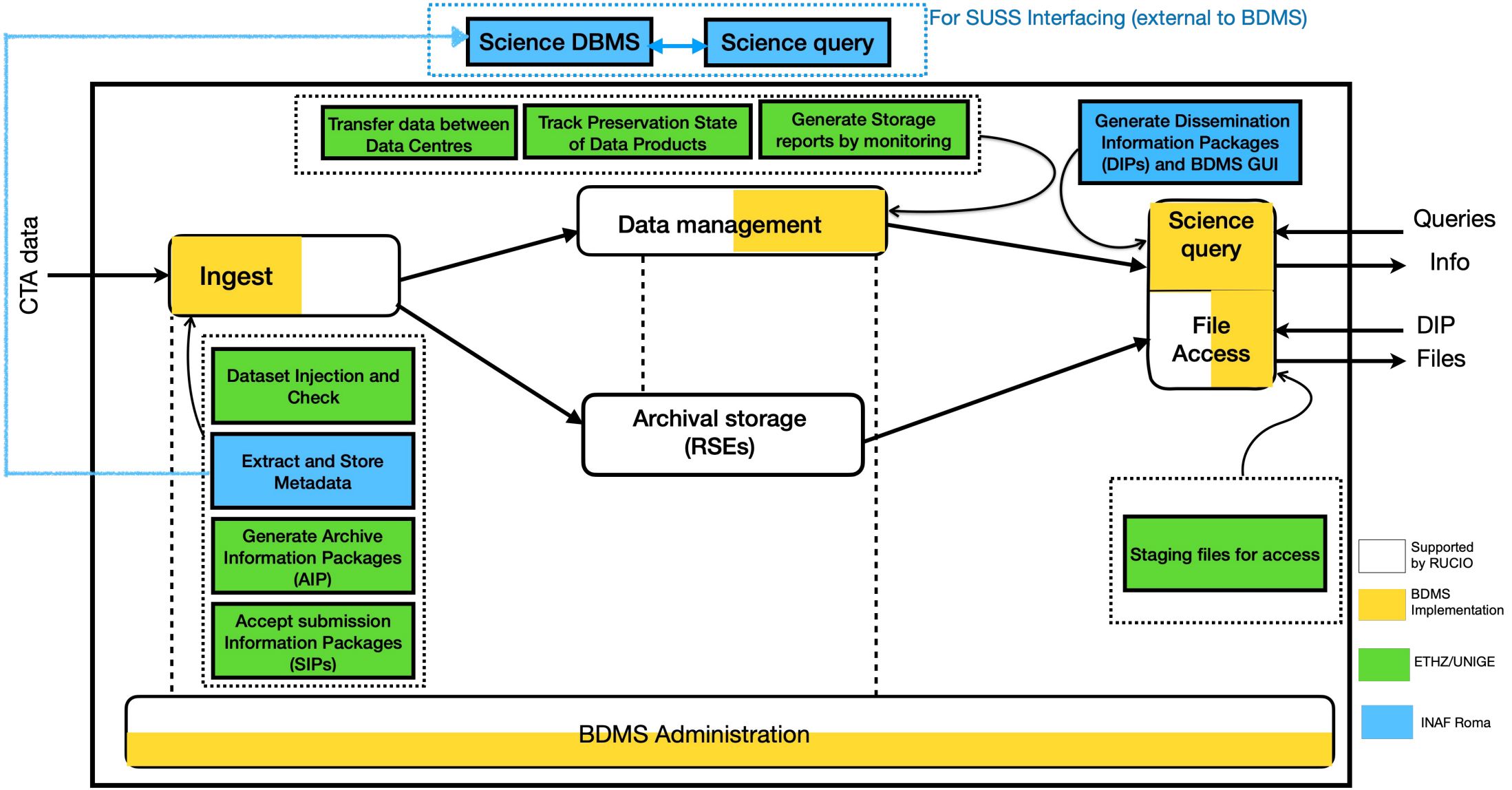
- Data Center teams in
  - Switzerland
    - ETH Zurich/CSCS
    - EPFL
  - Spain
    - PIC
  - Italy
    - Frascati
  - Germany
    - DESY

Open Archival Information System (OAIS) – Logical Flow

- OAIS standards design from high energy astronomy archive experience
- INTEGRAL archive to be the first one to be implemented in the framework of OAIS
- For CTAO, Rucio provides a declarative engine for *Bulk archive* with data being stored at multiple off-site datacenters



Role of RUCIO in DPPS – BDMS Functional decomposition



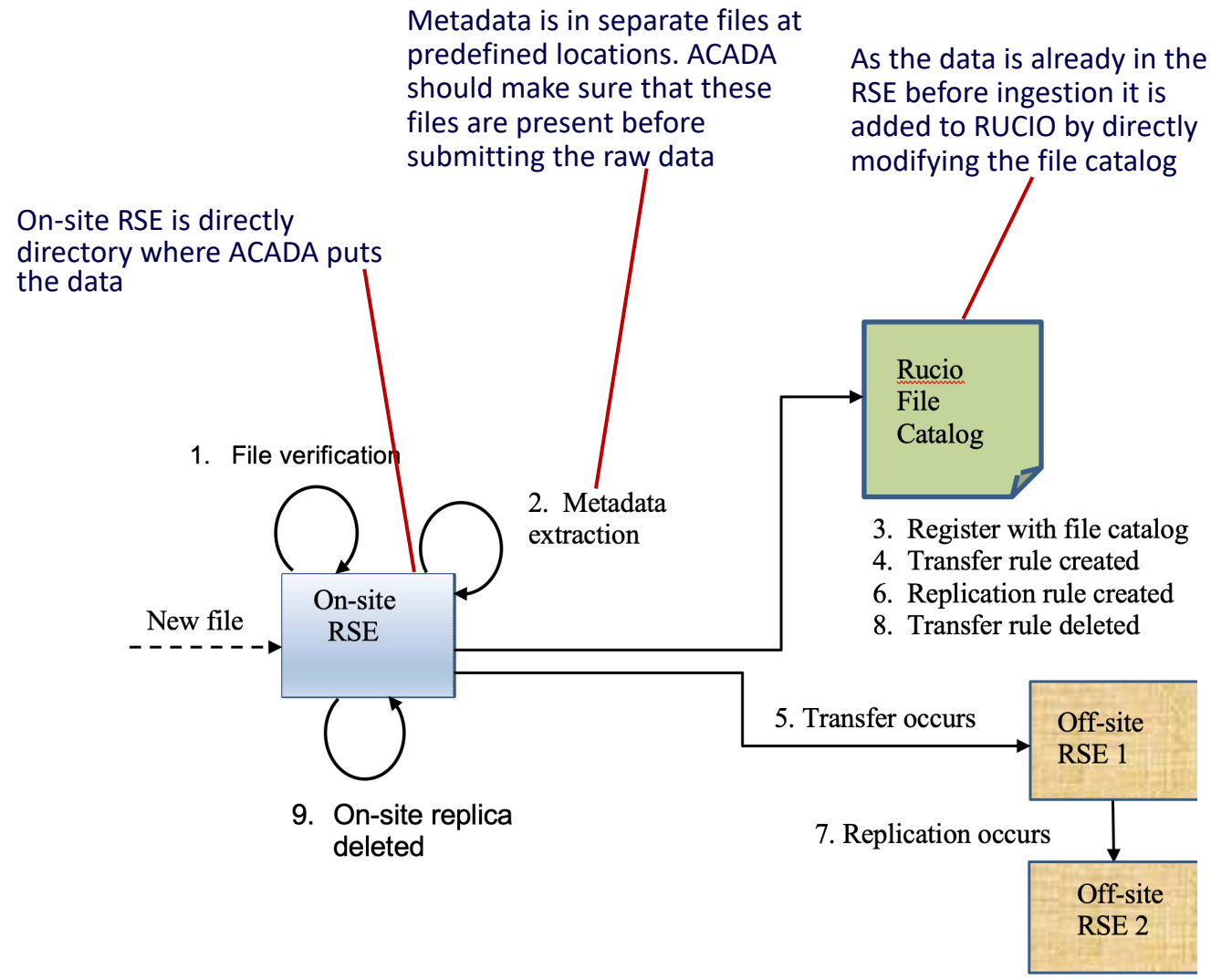
Overview on use-cases for DPPS BDMS

- **BDMS Ingest**
  - Add new data for either long-term archive or as temporary product
  - Data will come either from ACADA (DL0) or WMS (DL > 0 and MC data)
  - Produces archived products and temporary stored products

- **BDMS Data management**
  - ACADA Case – on-site to off-site data transfers: The data at the CTAO-North site in La Palma will first be transferred to an on-site Rucio storage element, followed by an off-site transfer to a data center (DC) and replication to at least one additional DC, thereby creating two replicas. Finally, the original copy on-site will be deleted
  - WMS Case: Currently we use Rucio file catalog plugin for DIRAC; Proposed a plugin for WMS software (DIRAC) for ingesting files to RUCIO

BDMS Ingest and Data Management: ACADA Case

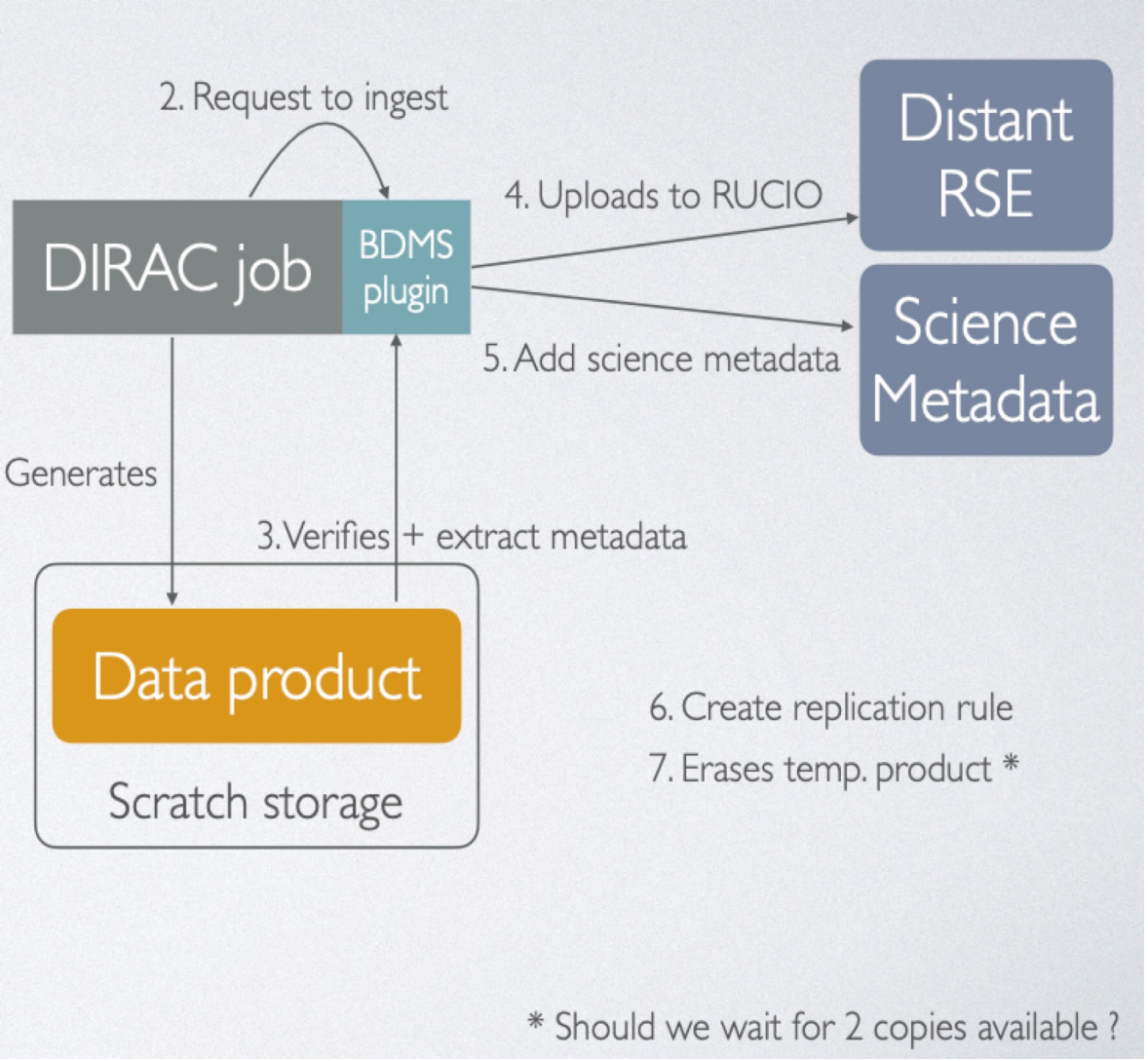
- ACADA plans to deliver DL0 (raw data) to DPPS on-site, via an interface directory and trigger files
- BDMS will need to have on-site scripts that will run the verification and metadata extraction



BDMS Ingest and Data Management: WMS Case

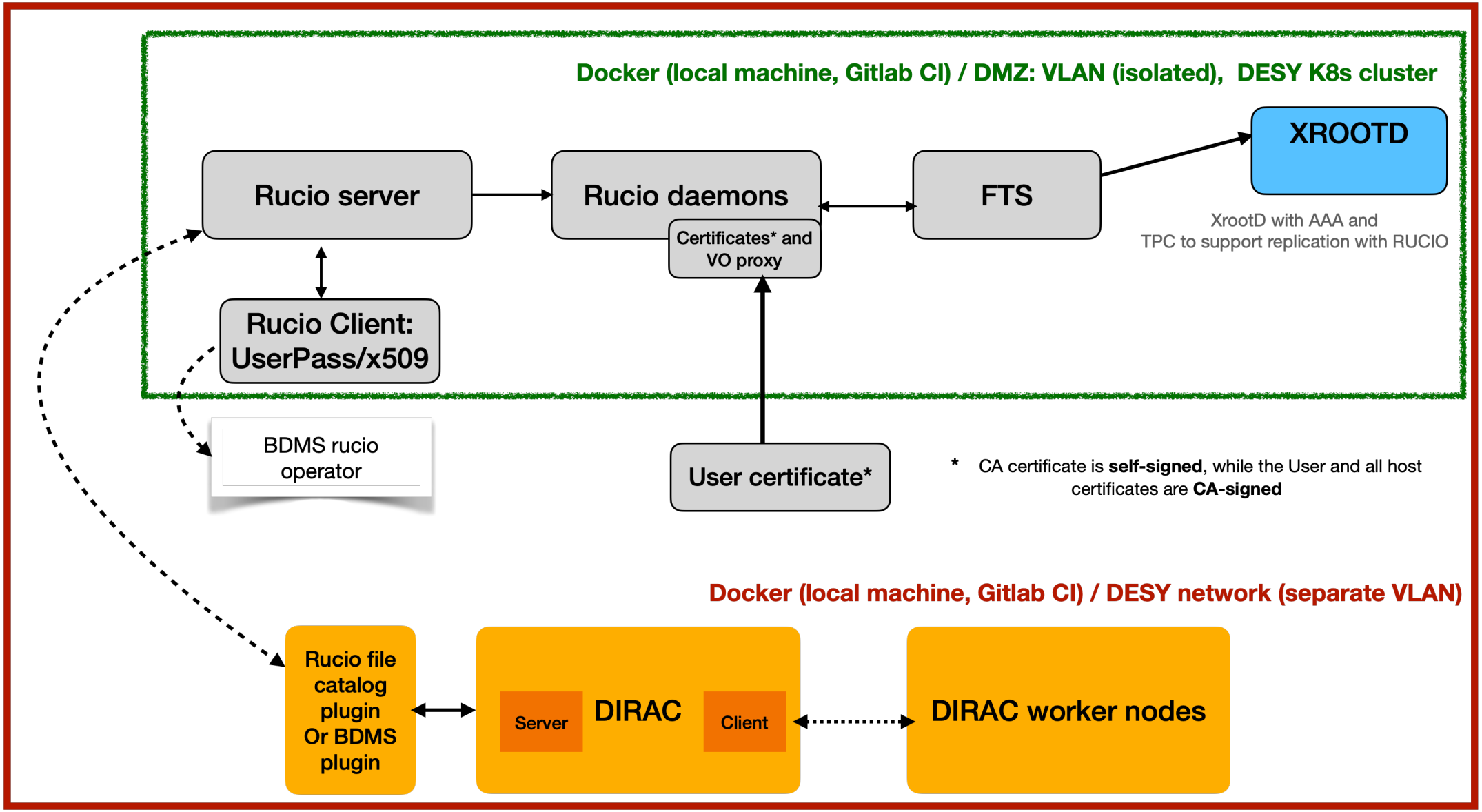
- Not as straightforward due to distributed architecture

- Could go the RUCIO-plugin way, i.e., creating our own plugin for DIRAC
  - BDMS plugin would perform verification and metadata extraction before uploading to RUCIO
  - For temporary datasets only metadata extraction is performed
  - Temporary datasets to be stored in a separate RUCIO scope from the archive data. DIRAC would have read/write/erase access to that scope only



BDMS Prototyping in Docker and Kubernetes (K8s)

BDMS prototypes: Docker / DESY K8s Test Cluster



BDMS Prototyping experience – Docker vs. Kubernetes

Attribute	Docker Compose	Kubernetes
Motivation	Local and fast prototyping, ease, testing	Ideal for production environments (CTAO off-site data centers have kubernetes infrastructure, e.g., PIC/CSCS)
RUCIO	All Rucio components have docker containers	Deployments create pods, services, ingress, replicasets; DB and RSEs pods have storage; RSEs and client have persistence
Management	Gitlab CI, Ansible, Rancher desktop	Rancher (K8 mgmt with high-availability), Fleet
Container orchestration	Docker Compose YAML file	YAML files, Helm Charts
Infrastructure	Single-node	Multi-node, distributed
Deployment tools	Docker Compose CLI	Helm, kubectl (local client to K8s cluster), vault, shuttle
Networking	Built-in networking	Ingress controller (nginx), K8s service (cluster type IP vs. LoadBalancer IP), Firewall configuration
Storage management	Docker volumes (local storage)	Persistent volumes (PVC), StatefulSets
Configuration management	Docker Compose YAML file, mounting certificates	configmaps, values.yaml (helm charts), K8s secrets, cert manager
Monitoring, Failure recovery	Prometheus/Grafana, Docker logs, manual restarts of containers	Prometheus/Grafana, kubectl logs, automated pod restarts, self-healing, deployment robust to cluster upgrades
Gitlab CI	VM with gitlab runner (shell), sonarqube to test coverage on Kubernetes runner	VM with gitlab runner (shell) needs kubectl, kubeconfig (to authenticate) and routing set-up (to access) the K8s Test cluster

BDMS prototype at DESY Kubernetes Cluster

```
NAME                                      READY   STATUS      RESTARTS       AGE
client                                    1/1     Running     0              4d5h
daemons-abacus-account-58575d8d9b-vx5fx   1/1     Running     6 (4d6h ago)   4d7h
daemons-abacus-rse-d5ddf9d7d-pbwxh        1/1     Running     6 (4d6h ago)   4d7h
daemons-conveyor-finisher-64f6ddb659-b9zhg 1/1    Running     0              4d6h
daemons-conveyor-poller-545f688b9b-z6tfv  1/1     Running     0              4d6h
daemons-conveyor-submitter-77449ddb57-b6c2g 1/1   Running     0              4d6h
daemons-judge-cleaner-5fc897c4cd-knbcz    1/1     Running     0              4d6h
daemons-judge-evaluator-f64fb6b5b-s87p5   1/1     Running     0              4d6h
daemons-judge-injector-5469745756-z6w57   1/1     Running     6 (4d6h ago)   4d7h
daemons-judge-repairer-cf86976cb-6vxtb    1/1     Running     0              4d6h
daemons-reaper-68f6d4f779-j8npl           1/1     Running     6 (4d6h ago)   4d7h
daemons-renew-fts-proxy-28785612-vx7rd    0/1     Completed   0              14h
daemons-renew-fts-proxy-28785972-7w7v2    0/1     Completed   0              8h
daemons-renew-fts-proxy-28786332-xk4r6    0/1     Completed   0              160m
daemons-undertaker-6f768f987f-fj2zt       1/1     Running     5 (4d6h ago)   4d7h
fts-mysql-58bd95ff57-fpk5f                1/1     Running     0              4d7h
fts-server-87dc66d59-9knc7                1/1     Running     0              4d7h
postgres-postgresql-0                     1/1     Running     0              4d6h
server-rucio-server-77797b7bc8-srl57      2/2     Running     0              4d7h
xrd1-7dbcfcbdc6-wdchp                     1/1     Running     0              4d7h
xrd2-67d96d44c5-kgx46                     1/1     Running     0              4d7h
xrd3-58bc96b5bd-fd7wn                     1/1     Running     0              4d7h
```

RUCIO pods and services at the K8s cluster

```
NAME                        TYPE           CLUSTER-IP      EXTERNAL-IP    PORT(S)             AGE
fts                         ClusterIP      10.43.53.70     <none>         8446/TCP,8449/TCP   4d23h
ftsdb                       ClusterIP      10.43.213.8     <none>         3306/TCP            4d23h
postgres-postgresql         ClusterIP      10.43.124.166   <none>         5432/TCP            4d23h
postgres-postgresql-hl      ClusterIP      None            <none>         5432/TCP            4d23h
server-rucio-server         LoadBalancer   10.43.92.207    141.34.51.80   80:31450/TCP        4d23h
xrd1-service                LoadBalancer   10.43.216.70    141.34.51.80   1094:31948/TCP      4d23h
xrd2-service                LoadBalancer   10.43.204.214   141.34.51.80   1095:30788/TCP      4d23h
xrd3-service                LoadBalancer   10.43.114.200   141.34.51.80   1096:32340/TCP      4d23h
```

```
[root@client user]# openssl s_client -CAfile /etc/grid-security/certificates/74df993b.0 -connect ruciobdms.cta-test.zeuthen.desy.de:443
CONNECTED(00000003)
depth=1 CN = DPPS Development CA
verify return:1
depth=0 CN = rucio-server
verify return:1
---
Certificate chain
 0 s:CN = rucio-server
   i:CN = DPPS Development CA
   a:PKEY: rsaEncryption, 2048 (bit); sigalg: RSA-SHA256
   v:NotBefore: Sep 19 15:39:51 2024 GMT; NotAfter: May 11 15:39:51 2049 GMT
---
Server certificate
-----BEGIN CERTIFICATE-----
```

REDACTED

```
-----END CERTIFICATE-----
subject=CN = rucio-server
issuer=CN = DPPS Development CA
---
No client certificate CA names sent
Peer signing digest: SHA256
Peer signature type: RSA-PSS
Server Temp Key: X25519, 253 bits
---
SSL handshake has read 1373 bytes and written 431 bytes
Verification: OK
---
New, TLSv1.2, Cipher is ECDHE-RSA-AES128-GCM-SHA256
Server public key is 2048 bit
Secure Renegotiation IS supported
Compression: NONE
Expansion: NONE
No ALPN negotiated
SSL-Session:
    Protocol  : TLSv1.2
    Cipher    : ECDHE-RSA-AES128-GCM-SHA256
    Session-ID: F5CB52757240FF3281F4E39822C8E3D87301F1609A8369C66966DF60F849128D
    Session-ID-ctx:
    Master-Key: 4AE686358340AF6444248BD10D2A18ADA7CB4DF6AC87D29596B92FB73CEC1EBE6FEE1F77C03CAD0AD4C9E62DCB163FAB
    PSK identity: None
    PSK identity hint: None
    SRP username: None
    Start Time: 1727189736
    Timeout   : 7200 (sec)
    Verify return code: 0 (ok)
    Extended master secret: yes
```

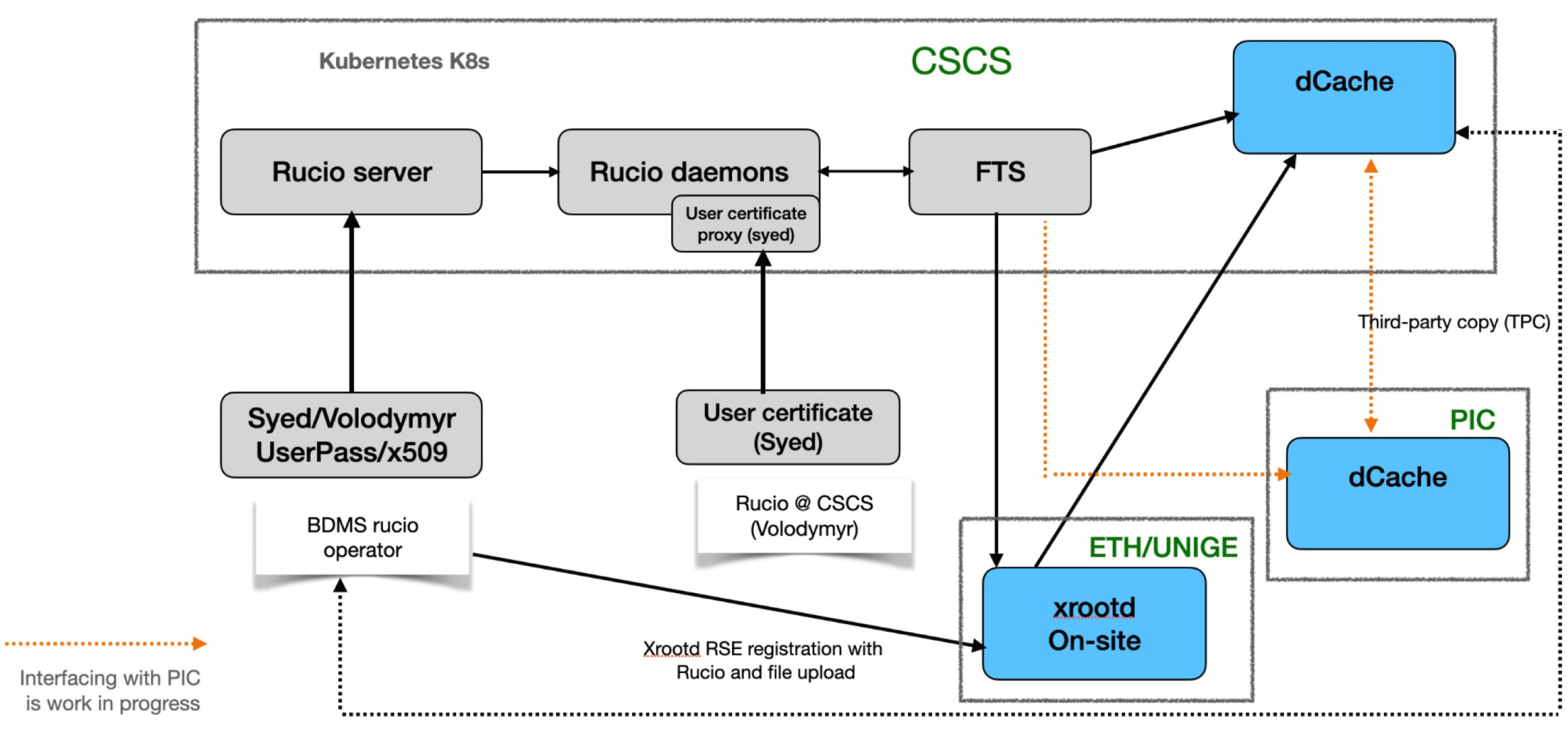
Unit testing to evaluate RUCIO functionality

```
[root@client user]# pytest -vvv test_rucio_operations.py
============================= test session starts ==============================
platform linux -- Python 3.9.18, pytest-8.3.3, pluggy-1.5.0 -- /usr/bin/python3
cachedir: .pytest_cache
rootdir: /home/user
collected 6 items

test_rucio_operations.py::test_server_version PASSED                     [
test_rucio_operations.py::test_authentication PASSED                     [
test_rucio_operations.py::test_rses PASSED                               [
test_rucio_operations.py::test_add_dataset PASSED                        [
test_rucio_operations.py::test_upload_file PASSED                        [
test_rucio_operations.py::test_replication PASSED                        [

============================= warnings summary ===============================
test_rucio_operations.py::test_upload_file
test_rucio_operations.py::test_replication
  /usr/local/lib/python3.9/site-packages/urllib3/connectionpool.py:1063: InsecureRequestWarning: Unverified HTTPS request is being made to host 'ruci
cta-test.zeuthen.desy.de'. Adding certificate verification is strongly advised. See: https://urllib3.readthedocs.io/en/1.26.x/advanced-usage.html#ssl
ngs
    warnings.warn(

-- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html
======================== 6 passed, 2 warnings in 91.67s (0:01:31) ========================
```

Helm chart of Rucio server, Rucio daemons, and Rucio Postgres DB

```
NAME       NAMESPACE      REVISION   UPDATED                                   STATUS     CHART                      APP VERSION
daemons    ruciodebug5    1          2024-09-19 17:43:32.336285373 +0200 CEST  deployed   rucio-daemons-35.0.0
postgres   ruciodebug5    1          2024-09-19 17:41:07.314714873 +0200 CEST  deployed   postgresql-15.5.32         16.4.0
server     ruciodebug5    1          2024-09-19 17:42:26.804087476 +0200 CEST  deployed   rucio-server-35.0.0
```

BDMS prototype at CSCS Kubernetes Cluster



- Realistic set-up of Rucio storage element with dCache at CSCS K8s cluster and XrootD deployed at UniGe/ETH
- Tested replication with RUCIO between UNIGE and CSCS using a VO proxy and different protocols (ssh, https, root)

Conclusion

- **BDMS prototyping**
  - Generating self-signed and CA-signed host certificates helped us to realize replication successfully
  - RUCIO workflow is identical on both Docker and Kubernetes, but we have to note the following
    - Prototyping is advantageous on Kubernetes as it has the same infrastructure on off-site datacenters
    - DPPS Release 0 requires a successful prototype in a Kubernetes Test cluster (DESY)

- **Next Steps**
  - Work on BDMS plugin code once it is agreed by the management
  - BDMS deployment on a Test data center for on-site (CTAO-N, La Palma) data to be transferred to an off-site data center (PIC)
  - Extending BDMS with monitoring, specially during data-transfer between data centers
  - Investigate (i) Token-based RUCIO, (ii) Storages and Protocols (dCache, webdav)

Thank you

Developers and Community

CTAO

www.ctao.org