



# WLCG Monitoring Task Force target goals for DC24

WLCG Monitoring Task Force

Alessandra Forti, Borja Garrido, Derek Weitzel, Fabio Andrijauskas, Julia Andreeva, Katy Ellis, Shawn McKee

WLCG DOMA Data Challenge 2024 Workshop <https://indico.cern.ch/event/1307338/>

November 9, 2023



# Index

- WLCG Monitoring task force
- DataChallenge requirements
- Reliable XRootD monitoring
- Site Network information
- WLCG common dashboards for DC



## WLCG Task Force

The WLCG Monitoring Task Force was set up following the outcome of the Data Challenge activity in 2021

The highest priority task is to address current issues with transfers and Site monitoring, following the needs expressed by the Data Challenge activity



## **What are the DC activity needs? What is the Task Force Providing?**

- Reliable XRootD monitoring
  - That can be plot together with FTS transfers data
- Site networking information
  - In order to correlate with other network information
- Common dashboard across experiments
  - Specific dashboard for experiment collaboration?



# Reliable XRootD Monitoring

- Current integration (GLED) covers only XRootD servers
  - Lots of the data is lost in transport (unreliable)
  - [New flow](#) will aim to fix this issues.
    - Main change is replacing UDP protocol by message queue.
    - First implementation done by OSG colleagues (thanks to Derek Weitzel) has been adopted for the European part of WLCG
- XCache monitoring
  - Covered by new flow as data streams are identical
- Monitoring for dCache with xrootd door use case
  - Not covered under the WLCG scope



## Reliable XRootD Monitoring (New flow) I

- Flow has been validated and the data we require for DC is as expected
  - Detected some fields produced by the new flow which might not be reliable but are not important for the exercise
    - (i. e : throughput, since we do approximate it over using the bytes read/write and time of the plot)
- New components are packaged and have been validated
  - Instructions for site managers on shoveler configuration can be found [here](#)
- Limited to single VO servers
  - Currently we need to hardcode VO information so this setup can only be done for non multi-vo servers
- [Grafana dashboard](#) to validate flow configuration (data arriving and being processed)
  - Should not be used for monitoring, that will be provided as part of other dashboard



# Reliable XRootD Monitoring (New flow) II

## Minimum requirements for DataChallenge 2024:

- Collector and shoveler deployments at CERN
  - Missing to validate scalability (More data required)
- EOS servers monitoring using the new flow
  - Waiting for restart campaign happening soon
- Access to enriched “raw” data (enriched with CRIC topology, not aggregated)
  - To be done by the MONIT team

## Nice to have:

- Other sites integrated. Part of US sites are already using message queue based implementation for a while. Campaign for European sites used by CMS, ATLAS and LHCb starts this week. ALICE sites will be included later
- Aggregation of data (in bigger bins with less labels for long term storage), not for DC per-se but good for accounting purposes



# Reliable XRootD Monitoring (dCache with xRootd door) I

- Currently not covered by WLCG monitoring
- Billing information for dCache can be enabled by the site. Not done everywhere, not integrated on the global level.n
- Python script to get data from Kafka and push it to CERN MONIT system
  - Developed at DESY (Thanks Sandro Grizzo)
  - Released as part of dCache 9.2
- Resolution of IPs to site and enrichment required
  - To happen on the MONIT site
  - Based on CRIC data and APIs





# Reliable XRootD Monitoring (dCache with xRootd door) II

Minimum requirements for DataChallenge 2024:

- Integration script released as part of dCache
- Validation of the workflow with Desy
- FNAL sending data to CERN activeMQ for monitoring
  - This will cover pileup data access
- Access to enriched “raw” data (enriched with CRIC topology, not aggregated)
  - To be done by the MONIT team

Nice to have:

- Other sites integrated, campaign to be started after validation of the workflow with Desy
- Aggregation of data, not for DC per-se but good for accounting purposes



# Site Network Monitoring

- **GOAL:** To **instrument** and **document** site networks to understand usage and identify bottlenecks, filling in missing information identified during [DC21](#).
  - [Presentation](#) from April WLCG Ops Coordination covers details
  - See DOMA [project description](#)
- Project is hosted on CERN Gitlab:  
<https://gitlab.cern.ch/wlwg-doma/site-network-information>
  - Currently have Python snmp monitoring example + [new GO version](#) to be added
- Site network monitoring visible (by RCSITE or NetSite) at  
<https://monit-grafana-open.cern.ch/d/Mwuxgoglk/wlwg-site-network?orgId=16&from=now-7d&to=now&var-site=All>

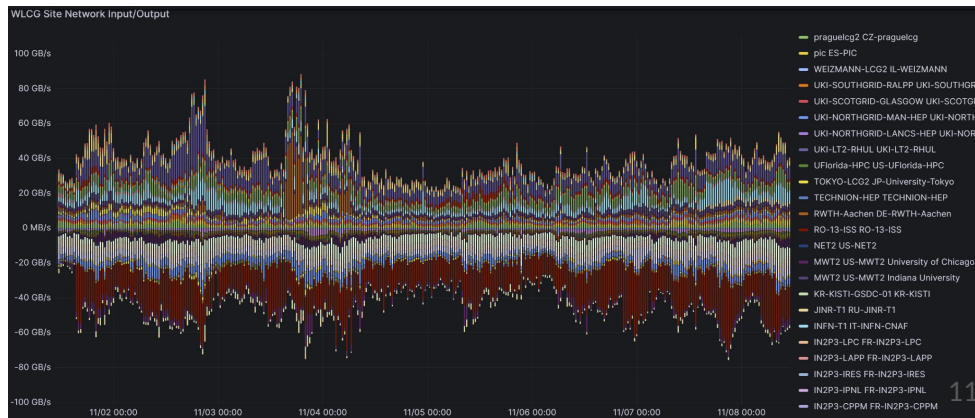
# Site Network Monitoring II

Minimum requirements for DataChallenge 2024:

- We expect ALL sites that plan to participate in DC24 to provide the requested network description AND total IN/OUT traffic for the whole site as described in CERN Gitlab.
- We currently have ~50% of the campaign identified sites complete ([32/64](#)).
- May need ticket follow up for non-responsive sites or new tickets for missing DC24 sites.

Nice to have:

- Comprehensive net description
- Site network diagrams





## Common dashboard across experiments

- [DC21 dashboard](#) still working
  - ATLAS, CMS and LHCb will continue to use this dashboard
  - Needs to be updated to DC24 requirements
    - Include XRootD data
    - Extend with network plots



# Common dashboard across experiments

## Minimum requirements for DataChallenge 2024:

- Usable dashboard for overview across experiments
  - Integrate FTS and XRootD information
    - XRootD contains same fields and a subset of labels so possible to plot them together
    - How to tell apart XRootD transfers from FTS for sites using the new flow
  - Integrate Site Network Monitoring information

## Nice to have:

- Experiments should tell us



## Overview for the Data Challenges

- Dashboard used for 2021 DC keeps working and can be reused/extended
- Site network monitoring campaign in progress to bring in all T1s and some T2s
- New components required for XRootD monitoring are ready
  - We will need to select some specific extra sites for testing purposes to confirm issues seen in RAL
  - Main focus will be to cover CERN (XRootD) and FNAL (dCache)
    - Other sites will be added progressively/in parallel
- Dependency on Scitags task from the “Research Network Technical Working group”
  - Until we get VO as part of the packet marking, we won’t be able to assign the proper one to “multi vo sites”
  - Activity label will also be needed to tell apart “Data Challenge” related transfers as it was done with FTS

**What is missing from the experiment or various project perspective!? Let’s Discuss!**

---

# Backup slides



## Dashboard for new functionalities?

It is critical that the Monitoring Task Force knows about needed monitoring from the various DC24 related activities.

We do have the WLCG DOMA [CERNbox location](#) but we don't know what the monitoring requirements actually are. Need to have efforts engage with our Task Force especially where we need to compare behavior/metrics with/without a capability.

Examples:

- Packet and flow marking information
- Jumbo frame vs non-jumbo frame sites
- Storage token use