

# US Mini Data Challenges (& ATLAS ADC Planning)

Shawn McKee / University of Michigan

ATLAS ADC TIM (<https://indico.cern.ch/event/1472836>)

January 23, 2025



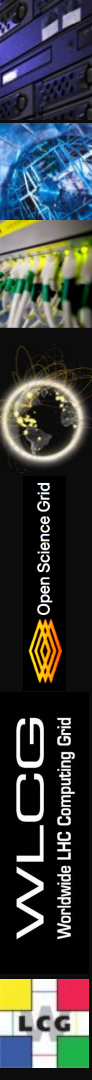
# Overview

For this ADC TIM, I wanted to cover plans from USATLAS regarding WLCG Data Challenges and associated “mini-challenges”

This work is being tracked by the new **TCB** (chaired by Alessandro de Girolamo and James Letts) and coordinated by **WLCG DOMA** (chaired by Johannes Elmsheuser and Katy Ellis).

The US has been very active in planning, preparing and executing both the Data Challenges and the intervening mini-challenges.

In this presentation I want to present planned activities and discuss the **ATLAS ADC perspective** on the path forward.

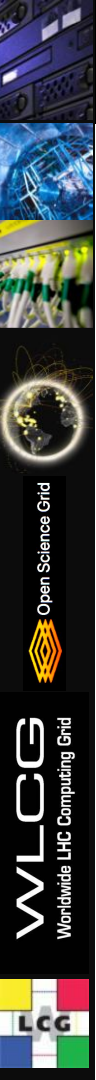


# Planning for Data Challenges

As we worked on the first WLCG Data Challenge in 2021, we discovered that the preparation and testing **prior** to the actual challenge was very beneficial.

We also learned that beneficial testing came in two forms: verifying the **capacity** of the data transfers (identifying bottlenecks) but also validating the **capabilities** we foresee as helping us to improve our infrastructure and eventually meet HL-LHC goals.

After DC21, we implemented the idea of mini-challenges to allow smaller and even “on-demand” testing to better understand if changes to our infrastructure were helpful and to better quantify impact.



# Fall 2024 US Mini Challenges

Hiro already covered the results of the US load testing from Fall 2024 as previewed in the Nov 13, 2024 [WLCG DOMA](#) in the previous talk.

Both USATLAS and USCMS undertook some capacity mini-challenges, designed to benchmark our current infrastructure.

These were simple load-tests where we wanted to evaluate the capacity limits for our various sites.

We were not trying to identify where we might adversely interact with other activities, as we do when we run the regular data challenges.

The important thing to note is that we intend to continue running “capacity” mini-challenges as a way to benchmark progress and status.

Next capacity date is early-to-mid summer 2025 (co-schedule with others)

# Capability Challenge Planning

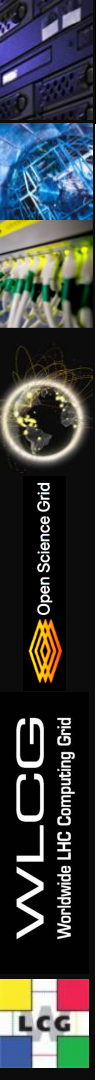
As part of the IRIS-HEP planning process, we discussed mini-challenges at the September 2024 retreat.

We targeted fall capacity tests, followed by capability tests early in the new year.

It is important that capability testing is down from the ground up, not imposed by the TCB or DOMA.

**This means we need proponents to self-organize within a rough timeline.**

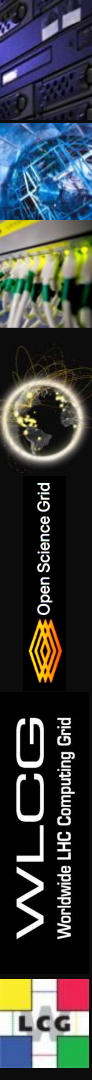
The plan is now to have “capability” tests during February of 2025.



# What is the Current Menu of Capability Challenges?

We have a number of capabilities that we believe would benefit our infrastructure, including:

- **Traffic visibility** in the network via **SciTags**
  - Includes both Fireflies and packet marking
- **Jumbo frames** in operation to improve WAN data transfer performance and reduce hardware resources
- **Software Defined Network** using projects like SENSE-RUCIO to orchestrate our networks to optimize and prioritize various workflows
- **Network monitoring validation** to ensure we have needed visibility into our network use and that the data we are gathering is accurate and well understood.
- Exploring **alternative network congestion control protocols** to achieve better WAN data throughput (e.g., BBRv3 or others compared to default protocols for EL9 systems)
- **IPv6-only sites** to reduce the burden on site administrators and improve the security of sites by only requiring a single data plane
- **Network flow monitoring** for sites to allow them better understanding of their collaborative use of the network and to better diagnose problems, identify security threats and to suggest improvements in distributed workflows for the experiments.
- **Increase in the size of data files**: Test performance and operations with an increased average file size to determine impact and benefit for WLCG.
- Interactions with **Cloud Storage**, e.g., [SEAL Storage Technologies](#).
- **Tape I/O** performance and use-cases?



# Organizing the Capability Challenges

We have targeted “February 2025” as the time for advocates to run various capability challenges.

To support self-organization, we have set up a **Google directory** and some templates to identify the various plans, personnel and timelines:

<https://drive.google.com/drive/folders/1Af7hWa0Zm30EuqsV1PbekSjb--gXAsVG?usp=sharing>

There is an overview document, a capability template and initial draft versions for all the currently know capabilities we may want to test.

Advocates should plan to update the Data Challenge Google calendar as they schedule their testing.



Open Science Grid



WLCG  
Worldwide LHC Computing Grid



# Capability Tests

We already have one capability we will test in February:

- Jumbo Frames (AGLT2 (Shawn), MWT2(Lincoln), NET2 (Eduardo))

There will likely be more

- Storage tokens, monitoring validation, SciTags, Tape optimization

We don't need to test ALL possible capabilities in February and we should plan to track all interesting capabilities until we get to HL-LHC.

For capabilities that show a good cost-benefit result, we should move them into our production infrastructure.

Some capabilities may not currently be viable and we will need to revisit them regularly with additional testing.

We can use the Google directory to track results and update planning.



# Topics for Discussion

Concerns or suggestions regarding “capability” vs “capacity” tests?

Are there priorities for ATLAS ADC that need (capability) testing and are not on our list yet?

- FYI, we added “file size” based upon input from James Letts / CMS. This seems like a useful thing to explore and capability testing can be a way to understand costs and benefits. Any volunteers to push this?

What can we do to increase participation in capability testing? Are there other activities underway that could be beneficially connected to capability testing?

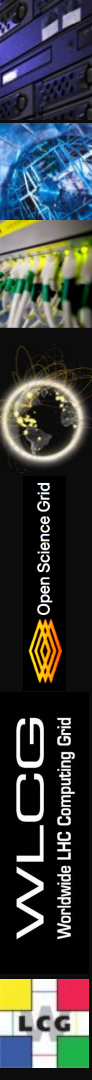
# Summary & Plans

We need to prepare activities for February regarding tests of potentially useful capabilities.

These tests will be driven by proponents and the Google docs will allow us to coordinate and inform the broader community.

This is a longer term activity in support of the WLCG Data Challenges and will need ongoing work...

## Questions or Discussion?



# Acknowledgements

Thanks to **Hiro Ito** for his contributions to the slides and for running the USATLAS tests!!

We would like to thank the **WLCG**, **HEPiX**, **perfSONAR** and **OSG** organizations for their work on the topics presented.

In addition we want to explicitly acknowledge the support of the **National Science Foundation** which supported this work via:

- IRIS-HEP: NSF OAC-1836650 and PHY-2323298



# Background Material

Here are some resources we know about:

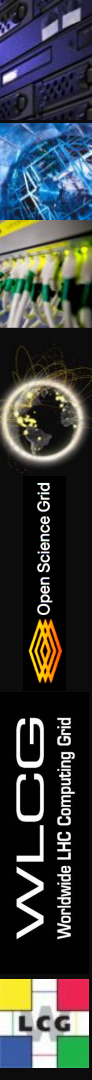
**WLCG DOMA wiki page:** <https://twiki.cern.ch/twiki/bin/view/LCG/DomaMiniChallenges> (includes link to [Google Calendar](#) to track activities)

## Presentations

- [WLCG Data Challenge 2024 \(DC24\) Status and Plans Related to ATLAS DDM](#) (Jun 2023)
- [DC24 Planning and Near Term Activities](#) (Jul 2023)
- [USATLAS Data Challenge 2024 Take-aways](#) (Feb 2024)
- [Medium to Long Term Network Plans for ATLAS and CMS](#) (Mar 2024)
- [DC24 Network Activities & Results](#) (May 2024)

## Some Google Docs

- [WLCG/DOMA Data Challenge 2024: Final Report](#)
- [USATLAS Milestones/MiniChallenges for Next WLCG Data Challenge in 2024](#)
- [Planning Mini-Challenges for US ATLAS Facilities and Distributed Computing](#)
- [NOTES: USATLAS Facility Status and Evolution Discussion](#)



# Backup Slides

# OSG-LHC/IRIS-HEP Planning

At the [IRIS-HEP retreat](#) in September 2024, we discussed how to prepare for DC26  
As mentioned, mini-challenges are an important tool that we want to enable

Goals for the next DC:

- Move the majority of our data via IPv6 and have one or more sites **IPv6-only**
- Have 80%+ of our traffic identified by SciTags
- Have SENSE/Rucio used in production at one or more sites
- Improve site network monitoring to identify traffic by LHCONE, LHCOPN, R&E and commodity

The plan:

- (DONE) Before the end of 2024 rerun capacity tests for US sites to determine current values
- (NEXT) February 2025, execute a joint USATLAS-USCMS **capabilities** mini-challenge: scitokens, SciTags, SENSE, jumbo frames
- Early-to-mid Summer 2025, execute a joint USATLAS-USCMS **capacity** mini-challenge

