Dear reviewer,

Thank you for your feedback and suggestions. We have carefully considered your comments and made the necessary revisions in the text. The corrected sections for each comment have been highlighted in red in the document.

Dear Editors,
Thank you for the interesting contribution and write-up of the proceedings. Apologies for the late review, I was only just appointed to review the work.
I have a few minor comments and questions below, and anticipate that we can iterate to an accepted state very quickly.

Kind regards

Pg1:
“ATLAS Experiment at the LHC[1] —- one of the largest”;
The em-dash appears too long, and as a combination of two dashes?

fixed

Pg2:
“Data in the WLCG are showed in the form of datasets”; “showed” does not work here; prefer something like “organised” ?

fixed

Pg2: “mc16\_13TeV … corresponding to the 2016 data-taking conditions…”; Strictly speaking mc16 is not just for the 2016 period, and in fact covers most (all?) of the run-2 campaign.
Was the dataset used actually reconstructed for 2016 data-taking (e.g. with the appropriate r-tag), or is this just a generalisation of the mc16 nomenclature?

Pg2 - a brief description or reference to DAOD and PHYS datasets might be beneficial, especially as PHYS is a fairly new format for ATLAS ?
Would it be useful to add in the total size and number of datasets considered in this study?

**We added three footnotes: for DAOD, mc16\_13TeV and PHYS. And added some information about number of datasets and length of time series.**

For this research a dataset group of Derived Analysis Object Data (DAOD)\footnote{DAOD is the format optimized for analysis and contains objects and information relevant to specific analyses}, belonging to the mc16\\_13TeV project\footnote{mc16\\_13TeV project contains Monte-Carlo simulated events corresponding to the run 2 data-taking campaign for pp collisions with an energy of 13 TeV} and PHYS format description\footnote{PHYS is the common derivation format intended for ~80\% of all physics analyses, not including non-standard analyses, like long-lived particles or signatures, custom jet collections} was chosen. The total length of time interval under investigation was about 2.5 years. The number of datasets included in the study group was 30 thousands.

Pg2: Section 2
From reading this section a few times, it is still not entirely clear what the inputs to the models are. I.e. how have you defined the popularity here? Is it the number of accesses per day, week, or month? Does it vary with the training period? I think a brief sentence would help for clarity.

**This sentence was added:**

Thus, the popularity of a group of datasets represents a time series with weekly calculated number of accesses to the group of datasets.

Pg 5:
“Up to 10 000 users utilise these resources for physics analysis”; for interest, where does this number come from; if it is based on the ~ number of ATLAS Users, then I suspect that the actual number of grid ‘users’ is practically much smaller?

**Yes, it was a mistake. We used the total number of users of the WLCG in general.**

Fixed: 5000

Pg 5: Definition of SMAPE; I believe there are a few ways to define SMAPE; would it be beneficial to the reader to include a footnote with the exact definition used?

\footnote{SMAPE quantifies the percentage difference between the observed and predicted values}

Pg5;
“Evaluation of user analysis payloads of a WLCG resource requires a relative metric showing is a relative metric showing the ability of a resource to process user analysis jobs in comparison to all other resources.”
Some extraneous words are present here.

fixed

Pg 6:
“they spent in queued and finalized states”
You defined three states previously, e.g. waiting, but now call it “queued”. Should amend to used consistent terminology.

fixed

For Waiting and Running, which of these in principle (if there is not direct-IO at a site) include the time to download a file to WN scratch space?

**“including time to download input files to a worker node” was added to the sentence describing all statuses.**

Jobs statuses were categorized into three types: waiting (assigned, activated, defined, pending, and other waiting statuses in a queue, including time to download input files to a worker node), running, and finalizing (statuses related to finishing jobs, such as merging or transferring).

Pg 6:
Eq1 and preceding text. The equation has “bars” above the terms, but the text does not include them in the definition of the variables. Are these the same definitions / uses of the symbols? If, so they should be made consistent.

Fixed

Do these values include only successor finished jobs, or are jobs that fail also included in the metric? If it is the later, does that affect how one should interpret the value of U\_t ? It should be mentioned in the text whether failed jobs contribute.

**The sentence was added:**

To avoid the contribution of failed payloads, in this research only successfully fiinished jobs were taken into account.

I assume these are all walltime values (should be defined). If so, is it ‘corrected’ for multi-core analysis workloads (e.g. multiplied for the number of cores), or, is that not relevant for this metric?

No, it is not walltime. It’s the average duration of jobs in different statuses at a resource in general, independently of the number of cores.

Pg 6 (incl. Fig 4.)
“There is obviously one site with insufficient analysis workload” : I think this needs some small clarification. Previously, a “site” was defined as synonymous with a “computing centre”. In Figure 4, what actually seems to be show are different Panda Queues, (e.g. T0 and T0\_ART), as well as different computing resources (e.g. T0 vs ‘normal’). Do you just mean that the T0\_ART queue was not running many jobs at certain periods? If so, perhaps the sentence can be rephrased slightly.

Pg6 “CERN-T0 is it” -> CERN-T0 it is ?

Figure~\ref{fig:3} demonstrates the evaluated analysis workload of three ATLAS-CERN sites over one randomly chosen week. **There is obviously one resource that was running not many analysis jobs ---** ANALY\\_CERN\\_T0\\_ART, the workload at CERN resource is normal, and at CERN-T0 it is normal as well, with short periods when this resource seems to be overloaded with analysis jobs.

Pg 7:
“closely at the this metric” : additional the / this

fixed

Pg 8:
“PanDA queues” this is the first mention of a PanDA queue, having previously been referred to as a Computing Center or ‘site’. Propose to use consistent terminology throughout.
In the introduction it would be good to mention that by ‘site’ (or similar), that you mean a PanDa queue at a site (or which there could be several queues per site).

**The following text was added at page 6 before the reference to figure 4.**

It should be noted that all evaluations were conducted on distributed queues associated with ATLAS sites. Each site comprises multiple queues, each serving a specific purpose.

Pg 8:
Data popularity forecasting:
Suggest to rephrase this sentence, as the “:” is redundant?

In terms of data popularity forecasting different methods, such as LSTM, Stacked LSTM, and Autoregressive Random Forest, were tested for assessing dataset popularity.

Do you have a target value for when these models will become useful, i.e. what SMAPE value after X days is required to make use of these models ?

**This sentence was added:**

A target accuracy of 5\% SPAME can be considered for the predictions. Therefore, our current models can forecast data popularity trends for a horizon of 1-3 days.

User analysis …
Suggest to rephrase this sentence, as the “:” is redundant?

An advanced job calculation method was proposed to estimate the workload analysis of WLCG resources. This method enables the estimation of jobs at PanDA queues within 1-hour intervals.