

ATLAS Data Access Policy

May 21st 2014

Introduction

ATLAS has fully supported the principle of open access in its publication policy. This document outlines the policy of ATLAS as regards open access to data at different levels as described in the DPHEP [1] model. The main objective is to make the data available in a usable way to people external to the ATLAS collaboration.

The ATLAS policy for data preservation is described in a separate document. The collaboration's need to preserve data for its own use shares some requirements with making them open access. To support open access to data additional resources will be required to develop and support the tools to make the data available.

Policies for Different Data Levels

Open access to ATLAS data by people outside the collaboration can be considered at four levels of increasing complexity, listed below, with associated conditions, see Ref. [1]. This policy pertains to collision physics data (i.e. that are stored offline and intended for physics analysis) and the necessary associated metadata, along with associated simulated datasets and tools allowing to produce new simulated datasets based on an adequate simulation of the ATLAS detector.

Level-1. Published results

All scientific output is published in journals, and preliminary results are made available in Conference Notes. All are openly available, without restriction on use by external parties beyond copyright law and the standard conditions agreed by CERN.

Data associated with journal publications are also made available: tables and data from plots (e.g. cross section values, likelihood profiles, selection efficiencies, cross section limits, ...) are stored in appropriate repositories such as HEPDATA[2]. ATLAS also strives to make additional material related to the paper available that allows a reinterpretation of the data in the context of new theoretical models. For example, an extended encapsulation of the analysis is often provided for measurements in the framework of RIVET [3]. For searches information on signal acceptances is also made available to allow reinterpretation of these searches in the context of models developed by theorists after the publication. ATLAS is also exploring how to provide the capability for reinterpretation of searches in the future via a service such as RECAST [4]. RECAST allows theorists to evaluate the sensitivity of a published analysis to a new model they have developed by submitting their model to ATLAS.

Level-2. Outreach and Education

ATLAS recognizes the vital role of outreach and education, and participates in and encourages outreach and education activities, and makes selected data available for them. Typically a fraction of the complete ATLAS data-set is used, selected to provide a rich sample of events with interesting physics signatures but not adequate for a publication of a physics result. The data are provided in simplified, portable and self-contained formats for

educational and public understanding purposes. Lightweight environments to allow the easy exploration of these data are also provided.

This includes event displays of selected events, ntuple or similar level data for illustrating the calculation of invariant mass distributions, lifetimes, CP asymmetries, etc.. ATLAS collision data or simulated data or both may be used. ATLAS has already made such data public for a diverse set of purposes, ranging from analysis programs and event displays for high-school and university education to releasing selected subsets of simulated data events for Data Science related contests [5]. ATLAS level-2 data would also be suitable as an input to more universal, multi-experiment outreach interfaces. Examples of the full complexity of the analysis data and environment may also be provided for educational purposes.

The data are provided for educational and public understanding purposes only, and may not be used for publication of physics results. Results obtained using these data may be used for publications in other scientific areas, e.g. computer science or education, after confirmation with ATLAS.

Level-3. Reconstructed data

ATLAS recognizes the potential value of making its reconstructed (level-3) data openly available after reasonable embargo periods. Any release of the data must be accompanied by a release of adequate tools to perform a data analysis. In addition, there are very few scientific use cases for level-3 data that do not also require tools for the generation, simulation and reconstruction of newly simulated data. The development of such tools for the ATLAS Collaboration members already involves significant manpower to support the software and analysis infrastructure, and to make this stand-alone and simple enough for use by non-members would require a substantial effort. ATLAS is willing to engage with any external project that might wish to develop such tools. This would likely require significant additional resources beyond those required for ATLAS internal needs, and any implementation is contingent on ATLAS being able to identify these resources.

ATLAS is committed to preserving all the information required to support the use of level-3 data throughout the lifetime of the collaboration, which also means that the option to make it open access will be preserved. If and when the required tools for making the data available to non-ATLAS members, the Collaboration Board or its delegates would consider a staged and delayed release of the level-3 data, subject to the required additional resources being provided. There would be an embargo period on each dataset to allow the members of the collaboration a reasonable time to perform analyses. Given the complexity of particle physics analyses, this embargo period will be measured in years, not months.

By far the most practical means of conducting a specific new analysis that requires level-3 data is in partnership with the collaboration. In particular, the ATLAS collaboration has an established programme of Short Term Association (STA) for specific analyses [6]. Non-ATLAS members can submit a proposal, supported by an ATLAS member, to collaborate with ATLAS on a particular topic of interest. These proposals are considered on a case-by-case basis and if granted by the Collaboration Board give access to the ATLAS data and the internal information required for this analysis.

Level-4. Raw data

It is not practically possible to make the full raw data-set from an experiment of the scale of ATLAS usable in a meaningful way outside the collaboration. This is due to the complexity of the data, metadata and software, the required knowledge of the detector itself and the methods of reconstruction, the extensive computing resources necessary and the access issues for the enormous volume of data stored in archival media. It should be noted that, for these reasons, general direct access to the raw data is not even available to individuals within the collaboration, and that instead the production of reconstructed data (as discussed above) is performed centrally. Therefore ATLAS does not propose to devote resources to developing access to the full raw data-set for non-members of the collaboration during the active lifetime of the collaboration. Access to representative smaller samples of raw data might be considered if well motivated.

References

- [1] Status Report of the DPHEP Study Group: Towards a Global Effort for Sustainable Data Preservation in High Energy Physics, [arXiv:1205.4667](https://arxiv.org/abs/1205.4667) [hep-ex]
- [2] <http://hepdata.cedar.ac.uk/>
- [3] RIVET: <http://rivet.hepforge.org/>
- [4] Kyle Cranmer and Itay Yavin, RECAST: Extending the Impact of Existing Analyses, JHEP 1104:038,2011
- [5] www.kaggle.com/c/Higgs-boson
- [6] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/AtlasPolicyDocuments>