Current Practices for Software Citation in ATLAS

George Redlinger, BNL
Pierre Savard, Toronto

Software Citation and Recognition in HEP, 22-Nov-22

Introduction

- We attempt to provide a snapshot of current practices in ATLAS for citing software in our physics publications
- We do not discuss the broader topic of recognition in the community for work done on software
- We will stick to being descriptive, rather than analytical. We make no attempt to reconstruct the logic behind the choices we make.
- In some citations, it is ambiguous whether we are citing the software itself as a tool or if we are citing the physics or analysis methods behind the software. We try to be as inclusive as possible when classifying such citations.

Catch-all citation for ATLAS software/firmware

ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, url: https://cds.cern.ch/record/2767187

- This note contains the following text:

1 Software Availability and Use

Software developed by the ATLAS Collaboration is gathered in the ATLAS Community on Zenodo [1]. This community entry collects software projects approved by the collaboration, including:

- Athena, the ATLAS simulation, reconstruction, and analysis software. This software is built and distributed in several different "projects" (collections of packages), to fulfill a variety of use cases. The software is also available on CERN GitLab [2].
- The ATLAS TDAQ software. This software and firmware collection is used in the trigger and data acquisition systems of the experiment and is also available on CERN GitLab [3].

Several versions of some of these packages are included in the Zenodo Community entry, corresponding to stable, well-validated and heavily-used releases of the software. An individual data analysis or data taking run may use several versions of the software for different data-taking periods, different simulated data samples, or different aspects of the analysis.

The software is generally distributed under the Apache License 2.0 [4]. Specific license, copyright, and ownership information for each contribution is available within the corresponding code repository and Zenodo record.

References

Detector Simulation

ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det]

GEANT4 Collaboration, S. Agostinelli et al., *Geant4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250

Monte Carlo generators (PYTHIA)

T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph]

T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph]

- We have used these for years, but they actually do not strictly follow what is requested at the PYTHIA website

Citing PYTHIA

PYTHIA is the product of many years of work developing models and implementing them in an event generator. When you use PYTHIA for a publication, **both** the most recent PYTHIA manual (currently *A comprehensive guide to the physics and usage of PYTHIA 8.3*) **and** the paper(s) introducing the physics model you are using, **must** be cited. Note that in some cases the used physics models are not created by, or only partially created by, the PYTHIA authors, underlining the importance of proper citations.

- In the pdf manual, as well as in the HTML manual, you can find an extensive bibliography pointing to the correct physics paper(s) to cite, in addition to the PYTHIA manual, when using PYTHIA.
- If you use PYTHIA by proxy, ie. through another program which interfaces to PYTHIA, and your result relies on this interfacing, the above guideline still holds.
- When using PYTHIA, always specify the full version number, and any relevant parameter modifications.
- You are welcome to contact us if you are unsure what to cite in your paper.

Monte Carlo generators (Madgraph)

J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph]

- Recommendation on the Madgraph5_aMC@NLO website

> The standard reference for the use of the code is: J. Alwall et al, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", arXiv:1405.0301 [hep-ph]. In addition to that, computations in mixed-coupling expansions and/or of NLO corrections in theories other than QCD (eg NLO EW) require the citation of: R. Frederix et al, "The automation of next-to-leading order electroweak calculations", arXiv:1804.10017 [hep-ph]. A more complete list of references can be found here: http://amcatnlo.web.cern.ch/amcatnlo/list_refs.htm

Monte Carlo generators (Sherpa)

E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph]

- Guidance from the Sherpa website

References

The main reference for Sherpa 2.2.x is arXiv:1905.09127. Please cite this publication if you have used Sherpa 2.2.x for your studies. Additionally, each run will produce a file called `Sherpa_References.tex` which contains information about the modules used in this run and the corresponding publications which should be referred. If in doubt, please seek the advice of the Sherpa authors.

- Typically we cite several other Sherpa papers, but these would probably be classified as papers on the physics or the algorithms, rather than software itself

Monte Carlo generators (POWHEG)

- Citation instructions from the POWHEG website, which is what we follow. Some additional citations are sometimes made

Proper references back to Index

In addition to the specific subprocess reference, the following papers should always be cited while using the POWHEG BOX:

- P. Nason, *JHEP* **0411** (2004) 040, hep-ph/0409146 [paper]
- S. Frixione, P. Nason and C. Oleari, *JHEP* **0711** (2007) 070, arXiv:0709.2092 [paper]
- S. Alioli, P. Nason, C. Oleari and E. Re, *JHEP* **1006** (2010) 043, arXiv:1002.2581 [paper]

If you use the RES package, please quote also

- Tomáš Ježo and Paolo Nason, *JHEP* **1512** (2015) 065, arXiv:1509.09071 [paper]

For a pedagogical introduction to the underlying theory, look at Next-to-Leading-Order Event Generators, P. Nason and B. Webber, 2012

Statistical analysis

- We always cite the “fundamental” papers that describe the methods, e.g. for unfolding, profile-likelihood, BLUE, HeraAverager, BumpHunter, Markov chain MC, …
- We are not consistent about citing certain types of well-known analysis software that implement these methods, including:
  - RooFit
  - RooUnfold
  - HistFactory
  - Pyhf
  - HistFitter
  - TRexFitter
  - BumpHunter
  - Bayesian Analysis Toolkit (rare)
- More often than not, these are usually not cited

Multivariate analysis and machine learning

- We are not completely consistent in citing publications describing the methods.
  - The BDT algorithm/method is rarely cited these days
  - For machine learning, textbooks are often cited, but not always
- For the software, we sometimes cite TMVA for BDTs
- For ML software, typical citations are:

F. Chollet et al., *Keras*, 2015, url: https://keras.io

M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, url: https://www.tensorflow.org/

Summary

- Although we stated at the outset that we will not reconstruct the logic that led to our citation choices, some general trends can be seen
- “Catch-all” citation for ATLAS software tools
- Papers on detector simulation methods, physics processes taken into account during detector simulation
- Citations of MC generators, generally following the citation requirements set out by the generator authors
- For statistical analysis and multivariate techniques, we cite the papers/textbooks describing the methods, but we generally do not cite the software tools we use to implement them
  - The one exception to this are the new ML tools like Keras/TensorFlow

BACKUP