GREG LANDSBERG, CHIARA MARIOTTI, AND ANDREAS MEYER

THOUGHTS ON PUBLISHING SOFTWARE PAPERS (IN CMS)

SOFTWARE CITATION AND RECOGNITION IN HEP MINI-WORKSHOP 22.11.22

CURRENT PRACTICES

- Theorists are very good in publishing papers related to software frameworks and packages
- GEANT4 Collaboration has standard journal references for full simulation
- ML tool and Root tool developers also provide better or worse defined citations: TMVA, Root, RooFit, MINUIT, RooStats, Keras, TensorFlow, ...
- Similar to ATLAS, we cite the recommended references (usually found on HEPFORGE) for various generators, such as: MadGraph, Pythia, POWHEG, SHERPA, MC@NLO, HPT, HYDJET, CASCADE, EPOS, SUPERCHIC, STARLIGHT, JHUGen, MELA, MCFM, and many others
- At the same time, we are pretty bad in citing our own software - most of these packages are not described even in the CMS Notes/PAS documents, let alone papers

PROS AND CONS

- Pros:
  - Gives software developers extra visibility and recognition (some of them do not sign our physics papers) - may help retaining more of them in academia despite generally lower pay than in industry
  - Documents the core infrastructure of our simulation, reconstruction, and physics analysis tools (in some cases we are forced to use conference proceedings, as for example for FastSim) - well cited references too!
  - Helps to make it easier for the general HEP community to analyze CMS Open Data
  - Many high-visibility journals (Nature, Science) now strongly pressure authors to make code public; git reference to the CMSSW repository is hardly a useful reference
- Cons:
  - None we could think of...

HAS IT BEEN DONE BEFORE?

- You've just heard about ATLAS and LHCb experiences so let's focus on non-LHC experiments
- Most of the publicly available HEE software references are conference proceedings
- However, there are a few examples of software papers:

The Belle II Core Software #2

Belle-II Framework Software Group • T. Kuhr (Munich U.) et al. (Sep 12, 2018)

Published in: *Comput.Softw.Big Sci.* 3 (2019) 1, 1 • e-Print: 1809.04299 [physics.comp-ph]

pdf DOI cite claim reference search 146 citations

ProtoDUNE-DP Light Acquisition and Calibration Software #2

D. Belver (Madrid, CIEMAT), J. Boix (Barcelona, IFAE), E. Calvo (Madrid, CIEMAT), C. Cuesta (Madrid, CIEMAT), A. Gallego-Ros (Madrid, CIEMAT) et al. (Mar 3, 2021)

Published in: *IEEE Trans.Nucl.Sci.* 68 (2021) 9, 2334-2341 • e-Print: 2103.02415 [physics.ins-det]

pdf DOI cite claim reference search 4 citations

Software and Computing for Small HEP Experiments #1

FASER and ATLAS and LZ and Fermi-LAT and H1 and T2K and SBND Collaborations • Dave Casper (UC, Irvine)(ed.) et al. (Mar 15, 2022)

Contribution to: 2022 Snowmass Summer Study • e-Print: 2203.07645 [hep-ex]

pdf links cite claim reference search 6 citations

MAUS: the MICE analysis user software #14

R. Asfandiyarov (Geneva U.), R. Bayes (Glasgow U.), V. Blackmore (Imperial Coll., London), M. Bogomilov (Sofiya U.), D. Colling (Imperial Coll., London) et al. (Dec 6, 2018)

Published in: *JINST* 14 (2019) 04, T04005 • e-Print: 1812.02674 [physics.comp-ph]

pdf links DOI cite claim reference search 3 citations

WHERE ONE COULD PUBLISH

- There are several journals that are very appropriate for publishing software papers:
  - Computer Physics Communications (impact factor 4.4)
  - Computing and Software for Big Science (impact factor 5.5)
  - Nuclear Instruments and Methods in Physics Research (impact factor 1.5)
  - IEEE Transactions on Nuclear Science (impact factor 1.7)
  - JINST (impact factor 1.4)

HOW TO PUBLISH?

- In CMS, there are two routes to publish such papers
- First, and most obvious, is a CMS Collaboration publication with extra authors added as exceptions
  - In my opinion this is the most preferred way of publishing on major software components, such as CMSSW or FastSim
- Second approach is to have a limited authorship paper from a non-PAG group, e.g., Software and Computing Group of CMS
  - Perhaps appropriate for more technical papers, such as batch job submission management software, data set publishing tools, various GRID tools

WHY HAVE NOT WE BEEN PUBLISHING SOFTWARE PAPERS?

- In fact, we had - mainly early in the CMS history, via limited authorship:

Design and development of a graphical setup software for the CMS global trigger #3

P. Glaser (Vienna, OAW), T. Nobauer (Vienna, OAW), H. Bergauer (Vienna, OAW), M. Padrta (Vienna, OAW), A. Taurok (Vienna, OAW) et al. (Dec, 2005)

Published in: *IEEE Trans.Nucl.Sci.* 53 (2006) 1282-1291

pdf links DOI cite claim reference search 1 citation

Implementation of a software feedback control for the CMS monitoring lasers #2

Liyuan Zhang (Caltech), Kejun Zhu (Caltech), David Bailleux (Caltech), Adolf Bornheim (Caltech), Ren-yuan Zhu (Caltech) (2008)

Published in: *IEEE Trans.Nucl.Sci.* 55 (2008) 637-643

DOI cite claim reference search 3 citations

OUR VIEW

- We definitely think that it's important to document significant software development done in a large experiment, such as CMS
- This is not just because of the "publish or perish" principle, but also because core developers move on and it is hard to keep history via online resources, such as Twiki
- Some of the architecture we use may be reused by other experiments - after all our main CMSSW code is public and in order for other to use it efficiently, it should be well documented

OUR WISHLIST

- We definitely want to see at least three software components documented as peer-reviewed papers that we could cite in the majority of our physics publications:
  - CMSSW Framework (Reconstruction, Simulation, Alignment and Calibration) - no public reference apart from the git repository
  - FastSim Framework - only a conference proceeding reference
  - combine Statistical Analysis Tool - no public reference apart from the git repository
  - Selected physics analysis packages (e.g., COFFEA) - never reference at all
- I am sure there are many more more specialized packages that deserve to be written up and published
- There is no fundamental reason why we should not be publishing these papers!