

Group Discussion Breakout

Software Citation and Recognition in HEP Workshop Discussion Questions

Attendees: Liz, Misha, Patrick, Alexandros, Eduardo, Mathew, Greg, George

Note taker: Matthew Feickert

Presenter (someone just to briefly summarize in main zoom room):

(**Block#1**) Questions to explore regarding software citation and credit for software

- a. What types of software should be cited in publications and how should they be cited?

In ATLAS where things are very conservative, there might be some resistance to citing tools. Eduardo, points out that we should maybe not be too concerned with what has historically been the case but we should perhaps state what we think we should be doing and let the experiments decide if they want to take these recommendations.

In terms of cascade citations, this also happens in physics with citing the PDG instead of a specific measurement. Patrick, for internal tools the citation is maybe different, but outside tools should be properly attributed. However, perhaps the software itself for collaboration tools can be cited as "Collaboration X" if there is some

- b. (For example, should we cite software directly, cite papers, cite both, and in what situations do these options make sense?)

Theorists ask for specific version citations. There is a lot of heterogeneous practices in this area, but we believe it should be both. Should we make a suggestion that code publication is important?

For reproducibility of software, perhaps ideally we can be creating a specific table of the software used (c.f. [Search for non-resonant Higgs boson pair production in the \$b\bar{b}\nu\bar{\nu}\$ final state with the ATLAS detector in pp collisions at \$\sqrt{s}=13\$ TeV](#) Table 1 for an example from ATLAS). Then for citations you are ideally also citing the version number, but as there is a very heterogeneous system of citation information available you are not always able to get a citation that is for the SemVer or CalVer release you used (e.g. there might be a paper for a release 2 years ago, but not for the version you actually used).

[Move this]: If you're citing only the original very old paper then this can be negatively impacting people who have done lots of development in the interim.

- c. If software developers say to cite their software using a specific method, should we follow their request strictly or loosely? How do we recognize contributors to software that have come in after a paper about the software was written?)

Most collaborations seem to be happy to go with the wishes of the authors.

Possible con that was brought up is if there are too many citations then sometimes journals do not like this and sometimes people who are reviewing the references are not happy about having to check a large number. How do experiments feel about cascade citations (Generally favourably).

George brings up that there should be some contrast to how the experiments handle hardware papers. Liz brings up the idea that the hardware journals actually have incredibly low citations. George also mentions that CERN is trying to push the idea that the real value of the publication is the number of citations to the article itself and not the journal.

- d. On harmonization/synchronization: What is the value of harmonizing practices of citing software within the experiments? How should the wishes of software authors and publishers be synchronized with the practices of those that publish HEP science (experiment or theory)? What are the prospects of such harmonization (a mechanism for doing so!) and what are the barriers?

It's difficult, there are clearly differences between external and internal software. Could get harmonization within the collaboration.

Matthew: we could provide incentives to authors to follow our field wide recommendations. If you make it clear to authors of software that if they want their software cited and they follow consistent methods there is a much better chance of this happening. Consistency is valued.

Misha: Having a common policy helps with adoption. If you can provide field wide recommendations / guidelines then it

Eduardo: There is a win-win situation. We have a uniform way of citing papers in journals which is a big win for everyone, and if we can have a common way of doing this for software that is also a win. DOIs are a common tool that we've started to move towards without explicitly discussing this, so if we can also move forward in a similar strategy with software there's obvious benefits.

- e. Is having a single paper that describes all the software used by an experiment good or bad? Should such papers be encouraged to be written and updated, and then cited? What are the advantages and disadvantages?

Patrick K: Similar to other situations : take the PDG where you try to dig into it and find the original publication of the number, at least for all external publication results.

Greg says that CMS does try to work with the authors. We agree that this is important.

Patrick: LHCb does not have a single collaboration wide paper on its software as ATLAS does

ATLAS has a single ½ page paper that just mentions the existence of the software and then lists the links to the individual software repositories that are ATLAS specific and additionally cites them.

(**Block#2**) Questions to explore regarding recognition and career development

- a. What steps can the HEP community take to ensure recognition of software efforts of those that develop and maintain software for HEP?
- b. What should we do to better support the career advancement of those that develop and maintain software for HEP?
- c. How should authorship of particular software be determined (i.e., who should be credited intellectually), and would this need to be approved by relevant experiments? (e.g., would it be similar to policies regarding authorship of papers? Should this take into account who needs credit for their career path?) What happens when there is a dispute of software authorship?