Lightning Talk: Software Citation: Process, Principles, and Implementation

Daniel S. Katz*, Kyle E. Niemeyer†, Arfon M. Smith‡, FORCE11 Software Citation Working Group§

*University of Illinois Urbana-Champaign, Urbana, Illinois, USA, Email: d.katz@ieee.org
†Oregon State University, Corvallis, Oregon, USA, Email: Kyle.Niemeyer@oregonstate.edu
‡GitHub, Inc., San Francisco, California, USA, Email: arfon@github.com
§FORCE11, https://www.force11.org/group/software-citation-working-group

Abstract—**Software is a critical part of modern research and yet there is little support across the scholarly ecosystem for its acknowledgement and citation. Inspired by the activities of the FORCE11 working group focused on data citation, FORCE11 started a Software Citation Working Group (SCWG). The group initially sought members, and currently has about 55–60 members. The working group reviewed existing community practices, developed a set of use cases, and drafted a software citation principles document.**

This presentation will discuss the principles (in brief: importance, credit and attribution, unique identification, persistence, accessibility, and specificity), how they will impact the practice of research, and they can be implemented by researchers, publishers, librarians and others who build and maintain repositories, scholars of science, university administrators, and research funders. It should also spark discussion in Track 2 of WSSSPE4 about both the next steps related to software citation and the community goals related to software credit, reproducibility, and sustainability.

I. LIGHTNING TALK

Software is a critical part of modern research and yet there is little support across the scholarly ecosystem for its acknowledgement and citation. Inspired by the activities of the FORCE11 working group focused on data citation, FORCE11 started a Software Citation Working Group (SCWG). The group initially sought members, and currently has about 55–60 people (researchers, developers, publishers, repositories, librarians) as members, including members of WSSSPE3 software credit breakout group who joined en masse in October 2015.

The working group reviewed existing community practices, including those of groups such as the Software Sustainability Institute, WSSSPE, Project CRediT, Ontosoft, and CodeMeta, and in domains such as astronomy and astrophysics, life sciences, geosciences. The group then developed a set of use cases (collaborative via a Google Doc [1]). Finally, the group drafted a software citation principles document, starting with the FORCE11 data citation principles, and then updated the equivalent software principles based on the software use cases and related work, a set of working group discussions, community feedback and review of the draft, and feedback and discussion in a one-day workshop at FORCE2016 in April 2016.

Seventeen discrete use cases related to software citation were established to help understand the necessary requirements for citation. These involve 13 stakeholder types, including researchers, research software engineers, publishers, indexers, domain groups, libraries, archives, repositories, funders, policy makers, evaluators, and citation managers. The use cases helped identify the basic metadata needed in a citation of software: unique identifier, software name, author(s), version number, release date, and location. Interestingly, only unique identifier was needed for all use cases. Beyond the principles of software citation themselves, the document contains extensive discussion about the principles and related topics. These topics include which unique identifiers should be used (DOIs are recommended), what software should be cited, the role of software papers, how to cite derived software, basic elements of a citation format in reference lists, the elimination of citation limits, the types of software that should be cited (all), and what an identifier should resolve to. In addition, the document contains a discussion of past and other work related to software citation, both in specific domains and the general research software community.

The FORCE11 Software Citation Working Group then made a set of recommendations for Software Citation Principles at the full FORCE2016 meeting April 2016. This has the goal of encouraging broad adoption of a consistent policy for software citation across disciplines and venues. The group also presented a discussion of the motivations for developing the principles, reviews of existing community practice, and a discussion of the requirements these principles would place upon different stakeholders.

Working examples and possible technical solutions for how these principles can be implemented will be discussed in a separate paper that is being developed.

After the FORCE2016 events, the working group modified the principles document to reflect issues that were raised, and then published the draft document on the FORCE11 website [2] and called for public comments, which were made via Hypothes.is (http://hypothes.is) annotations, GitHub issues, and emails. All comments were compiled into a document on GitHub [3], and explanation of the working group chairs' responses (e.g., document changes, further discussion, etc.) were also recorded there. All previous work by the group was

also done in the open, and recorded on GitHub, so that all decisions/discussions can be traced.

The final draft paper was then submitted to *PeerJ Computer Science*, published in *PeerJ Preprints*, and posted on the FORCE11 website [2]. Initial review comments were returned 8 July 2016, and the working group chairs recently (1 August 2016) submitted a revised version, and updated the *PeerJ Preprints* document [4]. Our goal is to have a final version of the paper accepted and published by WSSSPE4.

The remaining plans of the Software Citation Working Group involve promotion and distribution of the Software Citation Principles, specifically an endorsement effort to get both individuals and organizations to sign on to the principles and related, creation of some publicity material such as an infographic and 1–3 slides. After this, the Software Citation Working Group will have completed its work and will end.

Next, we expect FORCE11 to spin up of a new working group focused on implementing the software citation principles. This group will work with institutions, publishers, funders, researchers, etc., to implement the principles established by the first group, and will write an implementation examples paper.

This presentation will discuss the principles (in brief: importance, credit and attribution, unique identification, persistence, accessibility, and specificity), how they will impact the practice of research, and they can be implemented by researchers, publishers, librarians and others who build and maintain repositories, scholars of science, university administrators, and research funders. It should also spark discussion in Track 2 of WSSSPE4 about both the next steps related to software citation and the community goals related to software credit, reproducibility, and sustainability.

REFERENCES

[1] FORCE11 Software Citation Working Group, "Software citation use cases," https://docs.google.com/document/d/1dS0SqGoBIFwLB5G3HiLLEOSAAgMdo8QPEpjYUaWCvIU, 2016. Accessed: 2016-07-10.

[2] A. M. Smith, D. S. Katz, K. E. Niemeyer, and FORCE11 Software Citation Working Group "Software Citation Principles," FORCE2016 Website, https://www.force11.org/software-citation-principles, 2016. Accessed: 2016-07-10.

[3] FORCE11 Software Citation Working Group, GitHub repository, https://github.com/force11/force11-scwg. Accessed: 2016-07-10.

[4] A. M. Smith, D. S. Katz, K. E. Niemeyer, and FORCE11 Software Citation Working Group, "Software Citation Principles," *PeerJ Preprints* 4:e2169v3, 2016. https://doi.org/10.7287/peerj.preprints.2169v3.