Software Citation: State of the practice, challenges, solutions

Stephan Druskat
German Aerospace Center (DLR), Institute for Software Technology & Humboldt-Universität zu Berlin, Dept. of Computer Science & de-RSE – Society for Research Software

Berliner Bibliothekswissenschaftliches Kolloquium
2020-06-02
Software citation: A problem?

How would you cite https://github.com/RoboMaster/RoboRTS?

For https://github.com/luogantt/xgboost_new, do you cite this?

@inproceedings{Chen:2016:XST:2939672.2939785,  
  author = {Chen, Tianqi and Guestrin, Carlos},  
  title = {{XGBoost: A Scalable Tree Boosting System}},  
  booktitle = {Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining},  
  ...  
}  
https://github.com/luogantt/xgboost_new/blob/master/xgboost-master/CITATION
Software Citation: General Challenges

Software is essential to research!

Software contributes to the research cycle …
1) … as infrastructure,
2) … as research product in its own right.

In research, contributions are measured through citation.

Cultural challenge: Software not recognized as research product
Systemic challenge: Current citation system is for papers and books
Overall challenge: Identify software used in research
### Software Citation Principles (Smith, Katz, Niemeyer, et al., 2016)

<table>
<thead>
<tr>
<th>Principles</th>
<th>Meaning: „Software citation …“</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>→ cites software like papers. [culture, system]</td>
</tr>
<tr>
<td>Credit and attribution</td>
<td>→ facilitates credit and legal attribution. [culture, system]</td>
</tr>
<tr>
<td>Unique identification</td>
<td>→ uses machine-actionable UIDs. [system]</td>
</tr>
<tr>
<td>Persistence</td>
<td>→ lets UID and metadata persist. [system]</td>
</tr>
<tr>
<td>Accessibility</td>
<td>→ facilitates access to software, metadata, etc. [system]</td>
</tr>
<tr>
<td>Specificity</td>
<td>→ facilitates identification of/access to versions. [system]</td>
</tr>
</tbody>
</table>
**Software Citation Principles: Are We There Yet?**

<table>
<thead>
<tr>
<th>Principles</th>
<th>Example Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>Policies, initiatives, guidance, spreading the word</td>
</tr>
<tr>
<td>Credit and attribution</td>
<td>Software journals, reference lists, policies, creditability</td>
</tr>
<tr>
<td>Unique identification</td>
<td>DOIs for software</td>
</tr>
<tr>
<td>Persistence</td>
<td>Archival repositories, Software Heritage</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Archival repositories</td>
</tr>
<tr>
<td>Specificity</td>
<td>Versioned DOIs</td>
</tr>
</tbody>
</table>

Also: Different software types! Metadata! Tooling! Stakeholder challenges! Etc. (Katz et al., 2019)

---

Are we there yet? Not quite! We’re **making progress**, but there are still (tough) **challenges**, including culture change (which is hard).
Software Citation - What Works Today?
The Citation File Format (CFF) (Druskat et al., 2017)

How to provide complete and correct metadata? Provision of software citation metadata, in machine- and human-readable format, following the Software Citation Principles.

cff-version: 1.1.0
message: If you use this software, cite it with the metadata below.
authors:  
  - family-names: Druskat
given-names: Stephan
  orcid: https://orcid.org/0000-0003-4925-7248
title: “My Research Software”
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2017-12-18

citation-file-format.github.io
Software Citation - What Works Today? biblatex-software style extension (di Cosmo, 2020)

How to use complete and correct metadata?

@softwareversion {delebecque:hal-02090402-condensed,  
    title = {Scilab},  
    author = {Delebecque, Fran\c{c}ois and Gomez, Claude},  
    url = {https://www.scilab.org/},  
    year = {1994},  
    month = jan,  
    license = {Scilab license},  
    version = {1.1},  
    repository= {https://github.com/scilab/scilab},  
    abstract = {Software for Numerical Computation.}  
}
Software Citation – Towards a better system for (software) citation
Impact of dependencies in res. software for citation (Druskat, 2020 & forthcoming)

How to record citation information for software dependencies?

Extend the research citation graph to record transitive credit for software dependencies.

Transitive credit? Credit maps!

Citation: paper A → software B → software C

Credit metric: B (A): .2, C (B): .3 -> C (A): .06
References


Figures

- Screenshot tex.stackexchange.com: Content under CC BY-SA 3.0 Unported

Thank you!

@stdruskat
stephan.druskat@dlr.de
https://orcid.org/0000-0003-4925-7248
https://de-rse.org