HAL
open science

Announcing biblatex-software: software citation made easy

Roberto Di Cosmo

To cite this version:

Roberto Di Cosmo. Announcing biblatex-software: software citation made easy. Software Engineering Notes, 2020, 45 (4), pp.22-23. 10.1145/3417564.3417570 . hal-02977711

HAL Id: hal-02977711

https://hal.science/hal-02977711v1

Submitted on 25 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Announcing biblatex-software: software citation made easy

Roberto Di Cosmo Software Heritage, Inria and University of Paris, France
roberto@dicosmo.org

Software, and software source code in particular, is widely used in modern research. It must be properly archived, referenced, described and cited in order to build a stable and long lasting corpus of scientific knowledge. And yet, up to now there was no specific support for citing software in the popular BibTeX format that we use for our bibliographies.
We got used to work around this lack of dedicated support by tweaking the `@misc` entry, citing the user manual, or an article related to a software artifact, or adding in the text or in a footnote urls pointing to a place where one could get the software: the project web page, or a code hosting platform.
These workaround were a manifestation of the fact that software artifacts themselves were not considered first class citizen in the scholarly world.
With the recent raise of interest on reproducibility of research results, from Artifact Evaluation Committees to the ACM Badges, we have started moving towards giving finally software the academic dignity it deserves.

1. THE BIBLATEX-SOFTWARE PACKAGE

Today, we are delighted to announce another step forward along this path, with the release of the `biblatex-software` package [8], available on CTAN [13], that allows to produce rich software bibliographies with ease. This package is a BibLaTeX *style extension* that adds support for these four *software entry types* to any other BibLaTeX style used in documents typeset in LaTeX, including in particular the ACM bibliographic style, making it straightforward to adopt.

This package support four kind of different entries:

`@software` for describing the general information about a software project

`@softwareversion` for describing a specific version or release of a software project

`@softwaremodule` for describing a module that is part of a larger software project

`@codefragment` for describing a fragment of code (full file, or selected lines of a file)

The **softwareversion**, **softwaremodule** and **codefragment** entries can inherit the missing fields from another entry designated by a *crossref* field.

The package also adds support for several new fields, and in particular the following ones:

license list (literal). The license/s in SPDX format (see [12]).

introducedin field (literal). For a software module or fragment, the version of the containing project where it has been first introduced.

repository field (uri). The url of the code repository (e.g on GitHub, GitLab, etc.).

swhid field (verbatim). The Software Heritage identifier of the software artifact. This intrinsic identifier is a SWHID (swh:cnt for a content, swh:dir for a directory, swh:rev for a revision, swh:rel for a release, etc.). See the SWHID specification [14] and the rationale [9].

Since an example is worth a thousand words, let's look at a couple of them.

1.1 Software, softwareversion and softwaremodule

The following entries describe the CGAL project [1], a sophisticated library of computational geometry algorithms, a specific version of it [2], and a module in the library [3]:

```
@software {cgal,
 title = {The Computational Geometry Algorithms Library},
 author = {{The CGAL Project}},
 editor = {{CGAL Editorial Board}},
 year = 1996,
 url = {https://cgal.org/}
}

@softwareversion{cgal:5-0-2,
 crossref = {cgal},
 version = {{5.0.2}},
 url = {https://docs.cgal.org/5.02},
 year = 2020,
 swhid = {
   swh:1:rel:636541bbf6c77863908eae744610a3d91fa58855;
   origin=https://github.com/CGAL/cgal/}
}

@softwaremodule{cgal:lp-gi-20a,
 crossref = {cgal:5-0-2},
 author = {Menelaos Karavelas},
 subtitle = {{2D} Voronoi Diagram Adaptor},
 license = {GPL},
 introducedin = {cgal:3-1},
 url = {https://doc.cgal.org/5.0.2/Manual/packages.html#
      PkgVoronoiDiagram2},
}
```

Here is the corresponding output produced using `biblatex-software`

[1] [Software] The CGAL Project, *The Computational Geometry Algorithms Library* (Coordinated by CGAL Editorial Board), 1996. URL: `https://cgal.org/`.

[2] [Software Release] The CGAL Project, *The Computational Geometry Algorithms Library* version 5.0.2 (Coordinated by CGAL Editorial Board), 2020. swhid: ⟨swh:1:rel:636541bbf6c77863908eae744610a3d91fa58855;origin=https://github.com/CGAL/cgal/⟩.

[3] [Software Module] Menelaos Karavelas, "2D Voronoi Diagram Adaptor", part of *The Computational Geometry Algorithms Library* version 5.0.2 (Coordinated by CGAL Editorial Board), 2020. lic: GPL. swhid: ⟨swh:1:rel:636541bbf6c77863908eae744610a3d91fa58855;origin=https://github.com/CGAL/cgal/⟩.

1.2 Software, softwareversion and codefragment

The following entries describe the Parmap library [4], a version of it [5], and a fragment of code of particular interest from one of the files that compose it [6]:

```
@software {parmap,
  title = {The Parmap library},
  author = {Di Cosmo, Roberto and Marco Danelutto},
  year = {2012},
  institution = {{University Paris Diderot} and {University
       of Pisa}},
  url = {https://rdicosmo.github.io/parmap/},
  license = {LGPL-2.0},
}

@softwareversion {parmap-0.9.8,
  version = {0.9.8},
  swhid = {
    swh:1:rev:0064fbd0ad69de205ea6ec6999f3d3895e9442c2;
    origin=https://gitorious.org/parmap/parmap.git;
    visit=swh:1:snp:78209702559384ee1b5586df13eca84a5123aa82},
  crossref = {parmap}
}

@codefragment {simplemapper,
  subtitle = {Core mapping routine},
  swhid = {
   swh:1:cnt:d5214ff9562a1fe78db51944506ba48c20de3379;
   origin=https://gitorious.org/parmap/parmap.git;
   visit=swh:1:snp:78209702559384ee1b5586df13eca84a5123aa82;
   anchor=swh:1:rev:0064fbd0ad69de205ea6ec6999f3d3895e9442c2;
   path=/parmap.ml;
   lines=101-143},
  crossref = {parmap-0.9.8}
}
```

[4] [Software] Roberto Di Cosmo and Marco Danelutto, *The Parmap library*, 2012. University Paris Diderot and University of Pisa. lic: LGPL-2.0. vcs: https://github.com/rdicosmo/parmap/.

[5] [Software Release] Roberto Di Cosmo and Marco Danelutto, *The Parmap library* version 0.9.8, 2012. University Paris Diderot and University of Pisa. lic: LGPL-2.0. vcs: https://github.com/rdicosmo/parmap/, swhid: ⟨swh:1:rev:0064fbd0ad69de205ea6ec6999f3d3895e9442c2;origin=https://gitorious.org/parmap/parmap.git⟩.

[6] [Software excerpt] Roberto Di Cosmo and Marco Danelutto, "Core mapping routine", from *The Parmap library* version 0.9.8, 2012. University Paris Diderot and University of Pisa. lic: LGPL-2.0. vcs: https://github.com/rdicosmo/parmap/, swhid: ⟨swh:1:cnt:d5214ff9562a1fe78db51944506ba48c20de3379;origin=https://gitorious.org/parmap/parmap.git;anchor=swh:1:rev:0064fbd0ad69de205ea6ec6999f3d3895e9442c2;path=/parmap.ml;lines=101-143⟩.

The package implements various mechanisms that provide fine grain control on the typesetting of the software-specific bibliographic entries, that are detailed in the documentation available in the `software-biblatex.pdf` file.

2. OBTAINING SWHIDS

The SWHIDs shown in the bibliography entries above are formatted as clickable links in the PDF version of this column, and clicking on them brings you into the Software Heritage archive [10], exactly on the page showing the software artifact that the author intended. This effect, which is best seen when following the links in the entry [6], greatly enhances the readability and usability of research articles that are dealing with software. The good news is that it's completely straightforward to have any publicly available source code archived in Software Heritage, and obtain the corresponding SWHID, in a few simple steps that are detailed in [7] as well as on the web page https://www.softwareheritage.org/save-and-reference-research-software/.

3. USING AND CONTRIBUTING

The `biblatex-software` package [8] is available on CTAN [13] and should be soon incorporated in the standard TEX and LATEX distributions. In the meanwhile, it is enough to copy in the directory that contains your article its key components: the extension package `software-biblatex.sty`, the references section style `software.bbx`, the data model extension `software.dbx` and the string localisation files `<language>-software.lbx`.

Contributions are welcome on https://gitlab.inria.fr/gt-sw-citation/bibtex-sw-entry

BibLaTeX users can finally properly handle software in their bibliographies. Let's spread the word!

References

[7] Roberto Di Cosmo. "Archiving and referencing source code with Software Heritage". In: *ICMS*. Lecture Notes in Computer Science. to appear, preprint available. 2020.

[8] [Software] Roberto Di Cosmo, *BibLaTeX stylefiles for software products*, 2020. url: https://ctan.org/tex-archive/macros/latex/contrib/biblatex-contrib/biblatex-software.

[9] Roberto Di Cosmo, Morane Gruenpeter, and Stefano Zacchiroli. "Referencing Source Code Artifacts: a Separate Concern in Software Citation". In: *Computing in Science and Engineering* 22.2 (Mar. 2020), pages 33–43.

[10] Jean-François Abramatic, Roberto Di Cosmo, and Stefano Zacchiroli. "Building the Universal Archive of Source Code". In: *Communications of the ACM* 61.10 (Sept. 2018), pages 29–31.

[11] Roberto Di Cosmo and Marco Danelutto. "[Rp] Reproducing and replicating the OCamlP3l experiment". OCaml. In: *ReScience C* 6.1 (2020).

[12] SPDX Workgroup. *Software Package Data Exchange Licence List*. https://spdx.org/license-list, retrieved 30 March 2020. 2019.

[13] *CTAN: the Comprehensive TeX Archive Network*. url: http://www.ctan.org/ (visited on 04/29/2020).

[14] Software Heritage. *SoftWare Heritage persistent IDentifiers (SWHIDs)*. Version 1.4. May 5, 2020. url: https://docs.softwareheritage.org/devel/swh-model/persistent-identifiers.html.