

Killed By A Thousand Paper Cuts?

A Newcomer's Perspective On Possibilities And Gaps In Software
Citation Workflows

Katrin Leinweber

orcid.org/0000-0001-5135-5758

Technische Informationsbibliothek
CC-BY-3.0-DE

- ▶ Leibniz Information Center for Science & Technology
- ▶ (FAIR) RDM expertise since 2013 (Angelina Kraft)
- ▶ currently establishing support capabilities for software projects
 - ▶ Software/Data/Library Carpentries.org workshops
 - ▶ “Train yourself and your team” (Hong 2018)

About me

- ▶ studied biochemistry, arctic ecology & geology
- ▶ PhD: method¹ development² for biofilm research
 - ▶ Git, R, Unix shell, L^AT_EX, etc.
- ▶ 7 quarters in pharma-LIMS design & implementation
- ▶ joined TIB in Nov'17 to support software projects

¹coupled light & electron microscopy doi:10.7717/peerj.858

²assay robotisation with KNIME doi:10.1186/s12866-016-0676-9

Developer Options To Expose Citation Info

Developer options to expose citation info (by convenience)

1. BibTeX snippet on website or in CITATION file (Wilson 2013)
 2. get Zenodo-DOI, -backup & -landing-page for GitHub repo³
 - ▶ since Sperber (2017): DOI versioning
 3. adhere to community standards (DESCRIPTION in R package)
 4. add metadata to landing pages (GitHub topics, Zenodo keywords, etc.) manually
- ▶ follow codemeta.json or CITATION File Format proposals

³[guides.GitHub.com/activities/citable-code](https://guides.github.com/activities/citable-code)

codemeta.json (Jones et al. 2017)

- ▶ exchange format⁴ for software archives
- ▶ generate top-down (by archive) or bottom-up (by build)?⁵
- ▶ codemetaR⁶ converts R package DESCRIPTION

⁴codemeta.github.io/

⁵[GitHub.com/force11/force11-sciwg/issues/36](https://github.com/force11/force11-sciwg/issues/36) &
[GitHub.com/ropensci/codemetar/issues/57](https://github.com/ropensci/codemetar/issues/57)

⁶[GitHub.com/ROpenSci/codemetar](https://github.com/ROpenSci/codemetar)

JSON...? It's a machine's work!

```
{ "@context": [
  "http://purl.org/codemeta/2.0",
  "http://schema.org"
],
"@id": "https://doi.org10.5281/zenodo.1048320",
"@type": "SoftwareSourceCode",
"identifier": "codemetar",
"description": "...defines a 'JSON-LD' format for...",
"issueTracker": "https://github.com/ropensci/codemetar/is
...}
```

CITATION File Format (Druskat 2016)

- ▶ alternative to implementing missing Bib(La)TeX items/fields
- ▶ alternative “source format rather than a target”⁷ where no community standard exists
- ▶ .cff can be parsed in Java & Ruby⁸ & converted to .bib

⁷[GitHub.com/force11/force11-sciwg/issues/45](https://github.com/force11/force11-sciwg/issues/45)

⁸[GitHub.com/Citation-File-Format](https://github.com/Citation-File-Format)

Machine- & human-read- & -writable YAML

```
cff-version: 1.0.3
message: If you use this software, please cite it as below
references:
  - type: software
    authors:
      - family-names: Druskat
        given-names: Stephan
        orcid: 0000-0003-4925-7248
    title: My Research Tool
    version: 1.0.4
    doi: 10043/zenodo.1234
```

Up-and-coming alternatives suggest low confidence, that Bib(La)TeX, biber, etc. will (be able to) implement Software Citation Principles (Smith, Katz, and Niemeyer 2016).

Risk of chism: users stays with L^AT_EX, while software citation implemented elsewhere.

User Options To Import & Use Citation Info

User options to import citation info (by convenience)

1. 1-click import into reference manager during web-search
2. .bib or other machine-readable snippet
 - ▶ CiteAs.org builds database & prototype applications around citing “alternative scholarly products”
 - ▶ SoftwareHeritage.org automatically ingests & archives public source code (Di Cosmo and Zacchiroli 2017)
 - ▶ might replace Zenodo as best citation enabler (Katz 2017)

However, more inconveniences along the way...



(c) NDR (2017) Autobahn 20: Wenn die Straße plötzlich versinkt

Example: All you need is... A less strict reg-ex.

zotero / translators

<> Code 185 Pull requests 57 Projects 0 Insights

support other CRAN mirrors #1552

Merged zuphilip merged 8 commits into zotero:master from katrinleinweber:patch-1 14 minutes ago

27. Feb. 2018, 09:41 MEZ

Conversation 16 Commits 8 Files changed 1

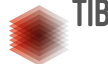
Changes from all commits Jump to... +326 -44

Unified Split Review changes

370 R-Packages.js

	@@ -2,20 +2,16 @@
2	2 "translatorID": "24a10ebf-ada1-4b8d-8f76-5a29e24d3e78",
3	3 "label": "R-Packages",
4	4 "creator": "Sebastian Karcher",
5	- "target": "r-project\\.org/web/packages/(\\.*/index\\.html available_packages_by_(name date)\\.html)",
5	+ "target": "(cran\\.\\.\\.+ cloud\\.\\.r-project.org CRAN)/web/packages/",
6	6 "minVersion": "3.0",
7	7 "maxVersion": "",

GitHub.com/zotero/translators/pull/1552



Example: CRAN “hides” DESCRIPTION in <frame>

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)

scientoText: Text & Scientometric Analytics

scientoText: Text & Scientometric Analytics

Zotero

Datei Bearbeiten Werkzeuge Hilfe

Titel	Ersteller	hinz
Comprehensive R Archive Network		29.11
scientoText: Text & Scientometric Analytics	Uddin	29.11

GitHub.com/zotero/translators/issues/1487

Example: Zenodo can't expose non-standard CSL⁹ type

The screenshot shows a web browser window with the URL <https://zenodo.org/record/1048320>. The page title is "ropensci/codemetar: codemetar: Generate CodeMeta Metadata for R Packages". The author list includes Carl Boettiger, Maëlle Salmon, Noam Ross, Arfon Smith, and Anna Krystalli. The publication date is November 13, 2017. The DOI is 10.5281/zenodo.1048320. The related identifiers section shows a link to the GitHub repository. The license is listed as "Other". The file list shows a zip file named "codemetar-0.1.2.zip" and a directory named "ropensci-codemetar-931d89b" containing files like ".Rbuildignore", ".gitignore", ".travis.yml", and "CONDUCT.md". A green box highlights the "Open Access" button, and a red box highlights the "License" dropdown menu.

November 13, 2017

Software Open Access

Available in

GitHub

Publication date:
November 13, 2017

DOI:
DOI: 10.5281/zenodo.1048320

Related identifiers:
Supplement to:
<https://github.com/ropensci/codemetar>

License (for reuse):
Other

Preview

codemetar-0.1.2.zip

ropensci-codemetar-931d89b

- .Rbuildignore 260 Bytes
- .gitignore 88 Bytes
- .travis.yml 204 Bytes
- CONDUCT.md 1.4 kB

[forums.Zotero.org/discussion/comment/301250](https://forums.zotero.org/discussion/comment/301250)

⁹Citation Style Language: CitationStyles.org

Is it a tragedy?



"Le soldat de Marathon" by Luc-Olivier Merson (Public Domain via Wikimedia Commons; 10678715)

Summaries

What is already possible?

Groundwork & conventions

- ▶ manually creating & importing CITATION with BibTeX snippet
- ▶ getting a DOI for software & generating a landing page¹⁰
- ▶ Python modules: add & extract citation info with DueCredit¹¹
- ▶ Debian Science can generate .bib from installed packages¹²
- ▶ 1-click Zotero import from GitHub, PyPI, CRAN, zbMath, etc.
 - ▶ Do Citavi, Mendeley, EndNote, etc. enable software citations?

¹⁰[guides.GitHub.com/activities/citable-code](https://guides.github.com/activities/citable-code)

¹¹[GitHub.com/DueCredit](https://github.com/DueCredit)

¹²[GitHub.com/force11/force11-sciwg/issues/45](https://github.com/force11/force11-sciwg/issues/45)

What needs doing?

Developer options: exist, but can be expanded

- ▶ discuss & promote GitLab-Zenodo integration¹³
- ▶ test codemeta.GitHub.io/tools
- ▶ test .cff Java & Ruby readers, as well as BibTeX converter¹⁴
- ▶ both: integrate into IDEs, build-tools, archives, etc.

User options: import hurdles, usage problems

- ▶ keep parsers/translators¹⁵ for software-relevant sites up-to-date
- ▶ help push Mr. Sperber's software type additions upstream
- ▶ help push software types in CSL¹⁶
- ▶ both: update styles with software-relevant fields¹⁷

¹³[GitLab.com/GitLab-org/GitLab-CE/issues/35023](https://gitlab.com/GitLab-org/GitLab-CE/issues/35023) & [GitHub.com/Zenodo/Zenodo/issues/1404](https://github.com/Zenodo/Zenodo/issues/1404)

¹⁴[GitHub.com/Citation-File-Format](https://github.com/Citation-File-Format)

¹⁵[Zotero.org/support/dev/translators](https://zotero.org/support/dev/translators)

¹⁶[GitHub.com/force11/force11-sciwg/issues/6](https://github.com/force11/force11-sciwg/issues/6)

¹⁷[GitHub.com/Citation-Style-Language/styles](https://github.com/Citation-Style-Language/styles)

Summary: general tactics and strategies

- ▶ problem: funding schemes favour many projects trying to be perfect, while ignoring ongoing lack of 80%-solution
- ▶ upvote, discuss, submit feature requests & pull/merge requests to existing tools
- ▶ enable contributions to your projects (Hong 2018)
- ▶ put metadata, documentation, website¹⁸, etc. all into one bask. . . git repo
 - ▶ thus into versioned Zenodo bundle =less re-gathering work
- ▶ shared ecosystem¹⁹ =everyone can & should help care for it

¹⁸pages.GitHub.com

¹⁹ScholarlyCommons.org

Summary: general tactics and strategies

Open Source Friday

[Contributors](#)

[Businesses](#)

[Maintainers](#)

[Sign up](#)

English

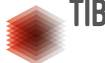


Build software with help from others

Every Friday, give your software project some love and make it easier for others to contribute.

[Sign up with GitHub](#)

OpenSourceFriday.com



Thank you for your attention, and to...

- ▶ Mila Runnwerth & Bastian Drees for reviewing the abstract
- ▶ MMS Days organisers for accepting it
- ▶ SCIWG members for advice & introductions
- ▶ TIB & its Lab Non-Textual Materials (NTM) for funding & support

Contact: Katrין.Leinweber@TIB.eu or 0511 762 14693

Bibliography I

Di Cosmo, Roberto, and Stefano Zacchiroli. 2017. "Software Heritage: Why and How to Preserve Software Source Code." In *iPRES 2017: 14th International Conference on Digital Preservation*. Kyoto, Japan. <https://hal.archives-ouvertes.fr/hal-01590958>.

Druskat, Stephan. 2016. "Lightning Talk: A Proposal for the Measurement and Documentation of Research Software Sustainability in Interactive Metadata Repositories." In.

Hong, Neil Chue. 2018. "Managing Research Software Development – Better Software, Better Research."
[doi:10.6084/m9.figshare.5930662.v1](https://doi.org/10.6084/m9.figshare.5930662.v1).

Jones, Matthew B, Carl Boettiger, Abby Cabunoc Mayes, Arfon Smith, Peter Slaughter, Kyle Niemeyer, Yolanda Gil, et al. 2017. *CodeMeta: An Exchange Schema for Software Metadata*. KNB Data Repository. <https://github.com/codemeta/codemeta/>.

Bibliography II

Katz, Daniel S. 2017. "Software Heritage and Repository Metadata: A Software Citation Solution." *Daniel S. Katz's Blog*. September 25. <https://danielskatzblog.wordpress.com/2017/09/25/software-heritage-and-repository-metadata-a-software-citation-solution/>.

Smith, Arfon M., Daniel S. Katz, and Kyle E. Niemeyer. 2016. "Software Citation Principles." *PeerJ Computer Science* 2 (September): e86. doi:10.7717/peerj-cs.86.

Sperber, Wolfram, dir. 2017. *Lost in the Web?* Technische Informationsbibliothek (TIB). doi:10.5446/21905.

Wilson, Robin. 2013. "Encouraging Citation of Software – Introducing CITATION Files." August 30. <http://blog.rtwilson.com/encouraging-citation-of-software-introducing-citation-files/>.