

Codemeta: A Rosetta Stone for Software Metadata

EARLY-CONCEPT Grants for Exploratory Research (EAGER)

Carl Boettiger, UC Berkeley

Abstract

Research relies heavily on scientific software, and a large and growing fraction of researchers are engaged in developing software as part of their own research (Hannay et al 2009). Despite this, *infrastructure to support the preservation, discovery, reuse, and attribution of software* lags substantially behind that of other research products such as journal articles and research data. This lag is driven not so much by a lack of technology as it is by a lack of unity: existing mechanisms to archive, document, index, share, discover, and cite software contributions are heterogeneous among both disciplines and archives and rarely meet best practices (Howison 2015). Fortunately, a rapidly growing movement to improve preservation, discovery, reuse and attribution of academic software is now underway: a recent NIH report, conferences and working groups of Force11, WSSSPE & Software Sustainability Institute, and the rising adoption of repositories like GitHub, Zenodo, figshare & DataONE by academic software developers. Codemeta is a distributed open source project to improve how these resources can talk to each other.

Example JSON-LD

```
Branch: master codemeta / examples / example-codemeta-rdataone.json
Find file Copy path
gothub Change order of properties in examples
48e5255 19 days ago
2 contributors
120 lines (119 sloc) 4.49 KB
Raw Blame History
1 {
2   "title": "dataone: R Interface to the DataONE REST API",
3   "description": "Provides read and write access to data and metadata from the DataONE network of data repositories. Each Data
4   "identifier": "https://dx.doi.org/10.5063/F1M61HSX",
5   "@context": "https://raw.githubusercontent.com/codemeta/codemeta/master/codemeta.jsonld",
6   "@type": "SoftwareSourceCode",
7   "agents": [
8     {
9       "@id": "http://orcid.org/0000-0003-0077-4738",
10      "@type": "Person",
11      "email": "jones@nceas.ucsb.edu",
12      "name": "Matt Jones",
13      "affiliation": "NCEAS",
14      "mustBeCited": true,
15      "isMaintainer": true,
16      "isRightsHolder": true,
17      "role": {
18        "namespace": "http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/gmxCodelists.xml#CI_RoleCode",
19        "roleCode": [
20          "originator",
21          "resourceProvider"
22        ]
23      }
24    }
25  ]
26}
```

A Metadata CrossWalk

18 contributors met in Portland, OR on April 15-17, 2016 to agree on a draft crosswalk between major software metadata schema.

Schemas in Crosswalk

DataCite, OntoSoft, Zenodo, GitHub, Figshare, Software Ontology, Software Discovery Index, Dublin Core, R Package Description, Debian Package Metadata, Python Distutils (PyPI), Trove Software Map, Perl Module Description (CPAN::Meta), JavaScript package descption (npm), Java (Maven), Octave, Ruby Gem, ASCL, Schema.org

Organizations



A JSON-LD Context for Software Metadata

doi 10.5063/schema/codemeta-1.0

- A linked-data representation of codemeta concepts
- Basic types from Schema.org, also Dublin Core and XSD
- 26 new terms introduced in codemeta namespace

Project Deliverables

Project Website: codemeta.github.io

JSON-LD Context file:

github.com/codemeta/codemeta/codemeta.jsonld

Crosswalk Table: github.com/codemeta/codemeta/crosswalk.csv

codemeta.json examples: github.com/codemeta/codemeta/examples

Coming soon

Utilities for automatically generating codemeta.json for common software package types (e.g. R packages)

Zenodo parsing of codemeta.json into Zenodo metadata and on to DataCite records.

Partner organizations extend metadata representation

Workshop Participants

- Carl Boettiger, UC Berkeley
- Matt Jones, NCEAS
- Arfon Smith, GitHub
- Yolanda Gil, USC ISI
- Martin Fenner, DataCite
- Krzysztof Nowak, Zenodo
- Mark Hahnel, figshare
- Abby Mayes, Mozilla Science Lab
- Luke Coy, RIT & MSL
- Kyle Niemeyer, Oregon State
- Alice Allen, ASCL
- Mercè Crosas, Harvard, IQSS
- Ashley Sands, UCLA
- Neil Chue Hong SSI
- Peter Slaughter, NCEAS
- Patricia Cruse, DataCite
- Dan Katz, NCSA
- Carole Goble, University of Manchester

See more contributors at

github.com/codemeta/codemeta/graphs/contributors and

github.com/codemeta/codemeta/network/members