Software Attribution for Geoscience Applications in the Computational Infrastructure for Geodynamics

Louise H. Kellogg[1], Lorraine J. Hwang [1], Allison Fish [2], MacKenzie Smith[1], Laura Soito [3], and Joseph Dumit[1]


The Problem: Scientific software, largely developed by individual scientists, represents a significant intellectual contribution. The increasing expectation that scientific software be open-source creates a need for mechanisms to cite software, to provide credit and recognition to developers, promote discoverability of software, and improve scientific reproducibility.

We investigate the requirements for software citation of a scientific community that develops much of its own scientific software for modeling: the Computational Infrastructure for Geodynamics (CIG).

Computational Infrastructure for Geodynamics (CIG) Community: A community-driven, NSF-supported consortium of 70+ universities and research institutions that supports development of scientific software for computational geophysics. Community involves 900+ individuals, spanning the domains of geophysics, computational science, applied mathematics, and computer science. Most software is forward modeling; CIG software is open-source, documented, and tested. For more information about CIG see: www.geodynamics.org

CIG Example 1: Data assimilation into a model of seismic wave propagation using CIG software SPECFEM3D

CIG Example 2: Rayleigh, a massively parallel spherical harmonic code for seismology. Running on ALCF Mitra

CIG Example 3: Aspect, an adaptive-mesh refinement FEM code for mantle convection. Aspect plume model courtesy of J. Dannberg

A strawman proposal for attribution: Software is cited more frequently when a citation tradition has been built in the literature, essentially showing the new user what and how to cite. Survey results indicate that scientists do not know how - what information do I include?, what – what is the citable unit & reference format? or when - what software needs to be cited? In response, we propose to provide attribution information through both CIG’s website and embedded in the software.

Sample citations:

In the body of the manuscript: “We use PyLith 2.1.0 for Darwin (Aagaard et al., 2013a; Aagaard et al., 2013b; Aagaard et al., 2015) published under the open source MIT expat license.”

In the acknowledgements: “We thank the Computational Infrastructure for Geodynamics (geodynamics.org) which is funded by the National Science Foundation under award NSF-0949446.”

In the References Cited or Bibliography:


Supported by the U.S. National Science Foundation under Grant No. SMA-1448633 and EAR-0949446.

CIG Community Software Practices

Defining the need: We used a mixed method approach that included an analysis of practices in the published literature, interviews, and surveys to examine the social and technical barriers to attribution experienced by members of the Computational Infrastructure for Geodynamics (CIG) community.

Survey respondents from the CIG community were mostly mid-level to senior level researchers.

- Do you need to write or modify software code during the course of your research?
  Yes: 94%
  No: 6%

Respondents indicated that software should be cited when its use was a significant factor to advance science. Exactly what this means and what should be recognized in citation were up for debate.

- Do you cite software in your publications?
  Yes: 93%
  No: 7%

Respondents indicated that software should be cited when its use was a significant factor to advance science. Exactly what this means and what should be recognized in citation were up for debate.

Current Practice: We examined more than 300 publications, provided by the CIG community, whose authors reported using CIG software. Software was mentioned by name 500 times, but without sufficient information to support reproducibility. Most publications named the software used and cited the software in the manner requested by developers; usually by citing a scientific publication that used or described the software or its methods. Few attributed CIG as the software repository.

Scientific software developers most frequently request that users cite a scientific publication using or describing the code. Many legacy codes provide no citation information.

Proposed Metadata: Of the schemes examined, the metadata needed to create the attribution database maps best into the dataCite metadata scheme. However, for many of the primary fields additional values in a controlled list need to be defined or created (*). In addition, the specification lacks a property for storing supplemental citation data (italics below).

Title	Rights	PublicationYear	RelatedIdentifier**
Version	Identifier* Description* Publisher** CitationList** Creator** Contributor** CitationSpecialText**

*Primary field with subproperties
*Require definition of additional values in controlled list

For more information: https://geodynamics.org/cig/projects/saga/draft-attribute

Feedback: saga@geodynamics.org