

Cultural and Technological Issues and Solutions for Geodynamics Software Citation

Eric M. Heien^[1], Lorraine Hwang^[1], Allison E. Fish^[2,3], Laura Soito^[3], MacKenzie Smith^[3], Joseph Dumit^[4], Louise H. Kellogg^[1,5]



1) Computational Infrastructure for Geodynamics
3) University Library

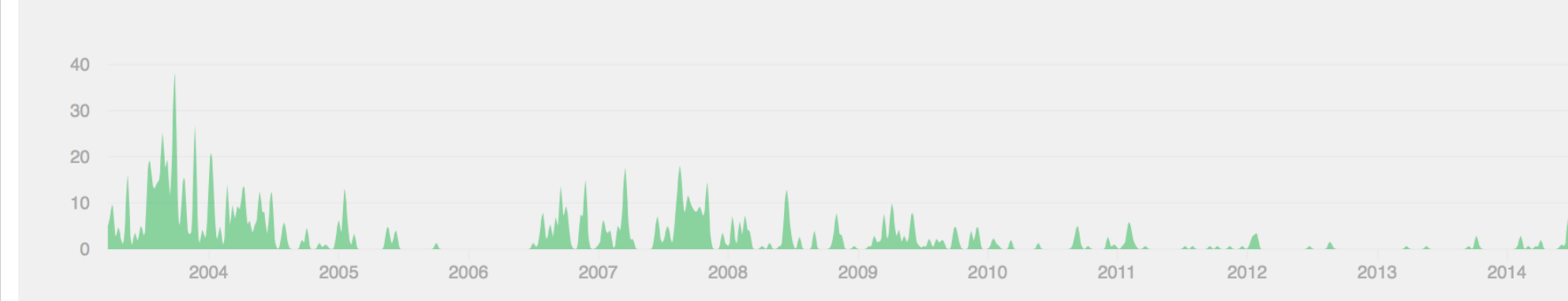
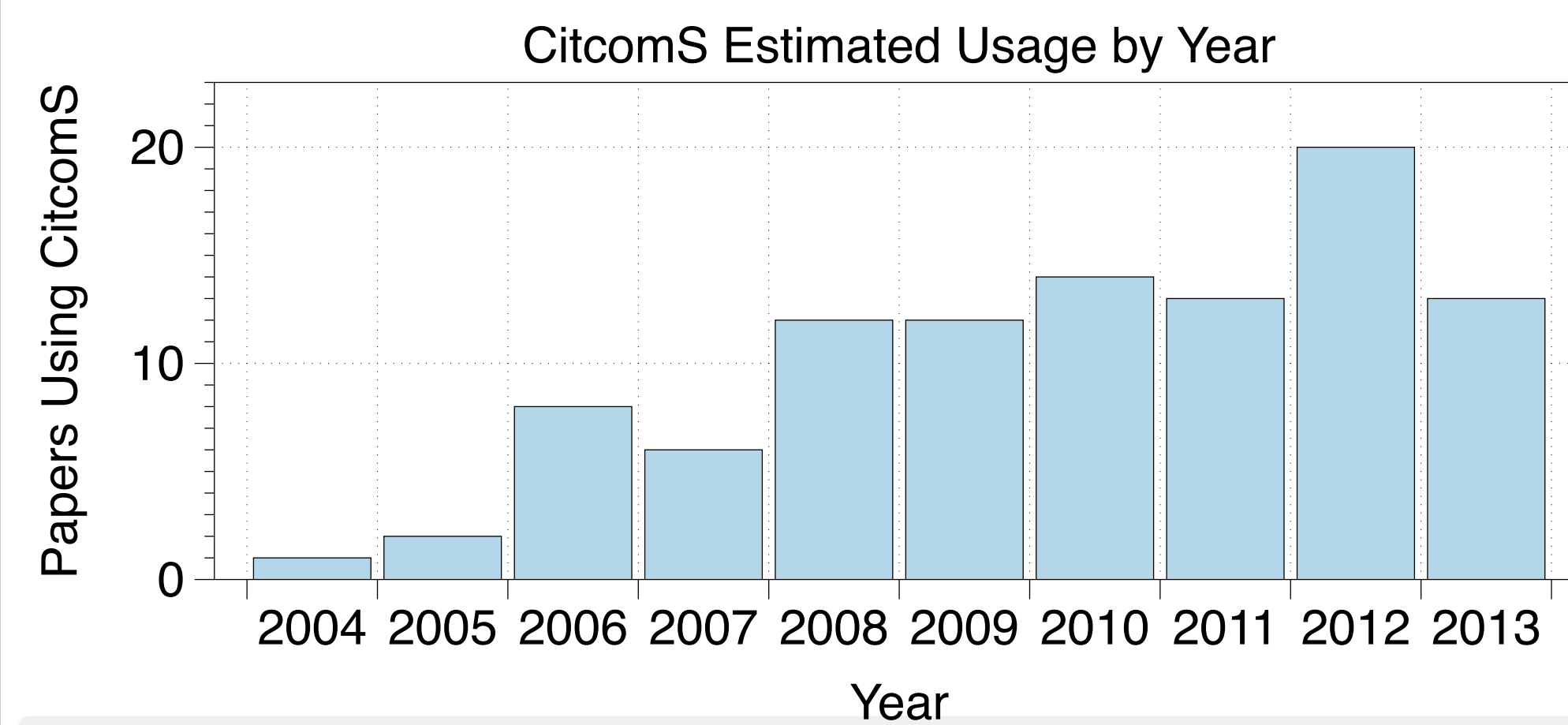
2) Center for Science and Innovation Studies
4) Department of Anthropology

5) Department of Earth and Planetary Sciences
University of California, Davis CA 95616 USA



Computational software plays a key role in scientific research. However, development and use of this software is hampered because there is no well-defined way for authors to receive credit or professional recognition. This discourages researchers from publishing codes or making them easier to use and results in the same functionality being written in multiple codes.

We investigate issues involved in citing software in a scientific context, and introduce features that should be components of a citation infrastructure, particularly oriented towards codes and scientific culture in geodynamics research.



Ten years of development activity and usage of CitcomS mantle convection code.

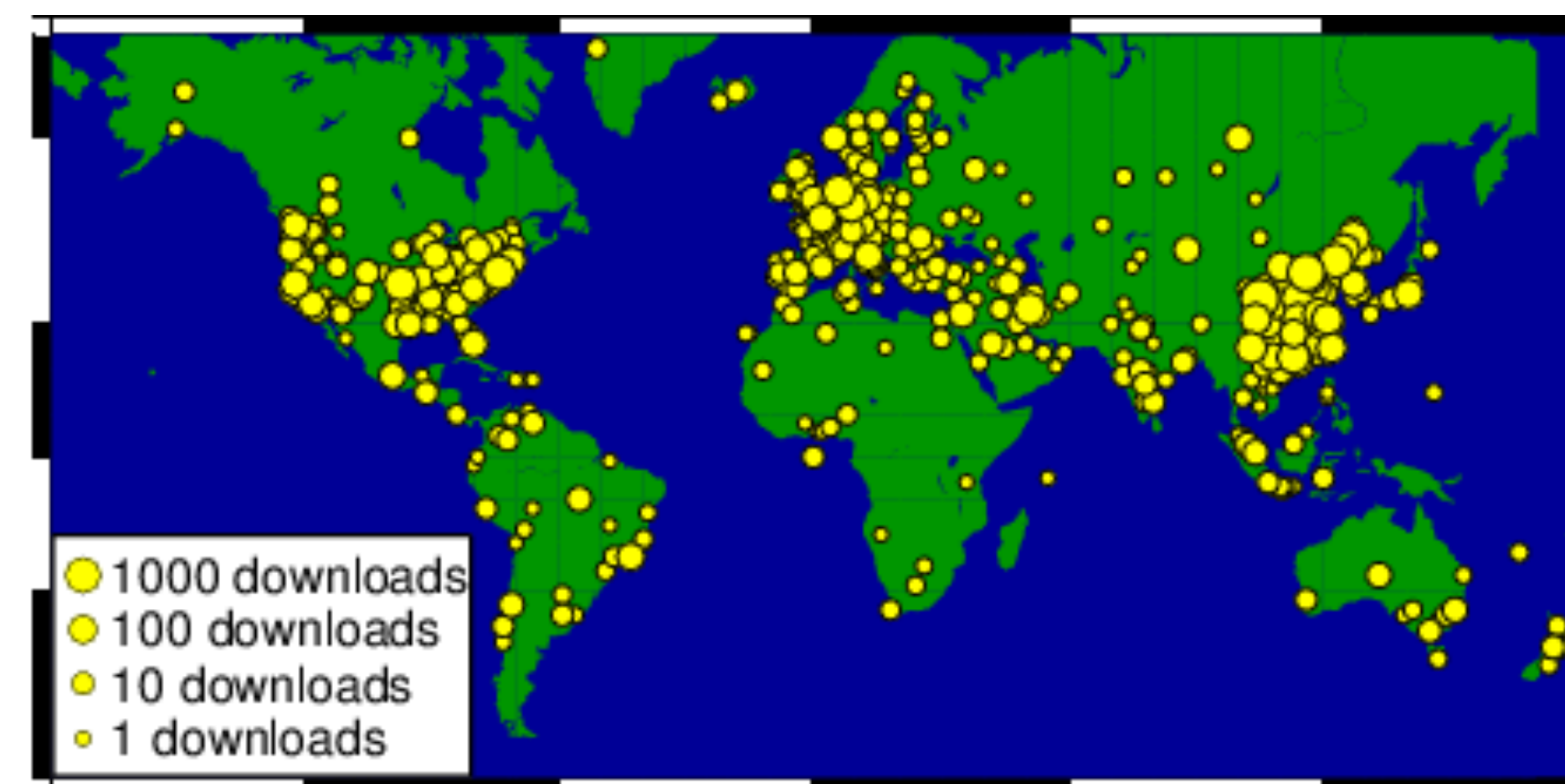
Current Issues

- Is there a compelling reason to cite scientific software?
- What software is cite-worthy?
- What are the expectations for software with multiple revisions (e.g. reproducibility, authorship, provenance)?
- Are available systems (e.g. DOI) sufficient for software complexity?

Cultural Barriers

What prevents citing/sharing code?

- Is only published research citable?
- Should journals enforce tracking/citation of software? There is a lack of accepted standards and protocols.
- Should researchers get credit for creating software and research tools?
- Scientists don't want responsibility for misuse of released software/tools.



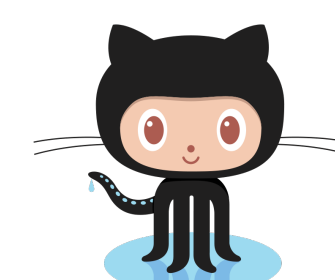
Download map of all CIG codes over a one year period. Even with widespread usage there is little support for citation or acknowledgement.

Existing Approaches

Cite Paper. Software is often referenced by citing the paper it was first described in. There are multiple issues with this, including 1) lack of details about software, 2) lack of clarity about software authorship, 3) no assurance of software availability.

Embedded Citation Generation. Some software or libraries generate standard citations, and give details on the runtime environment, functionality, and version.

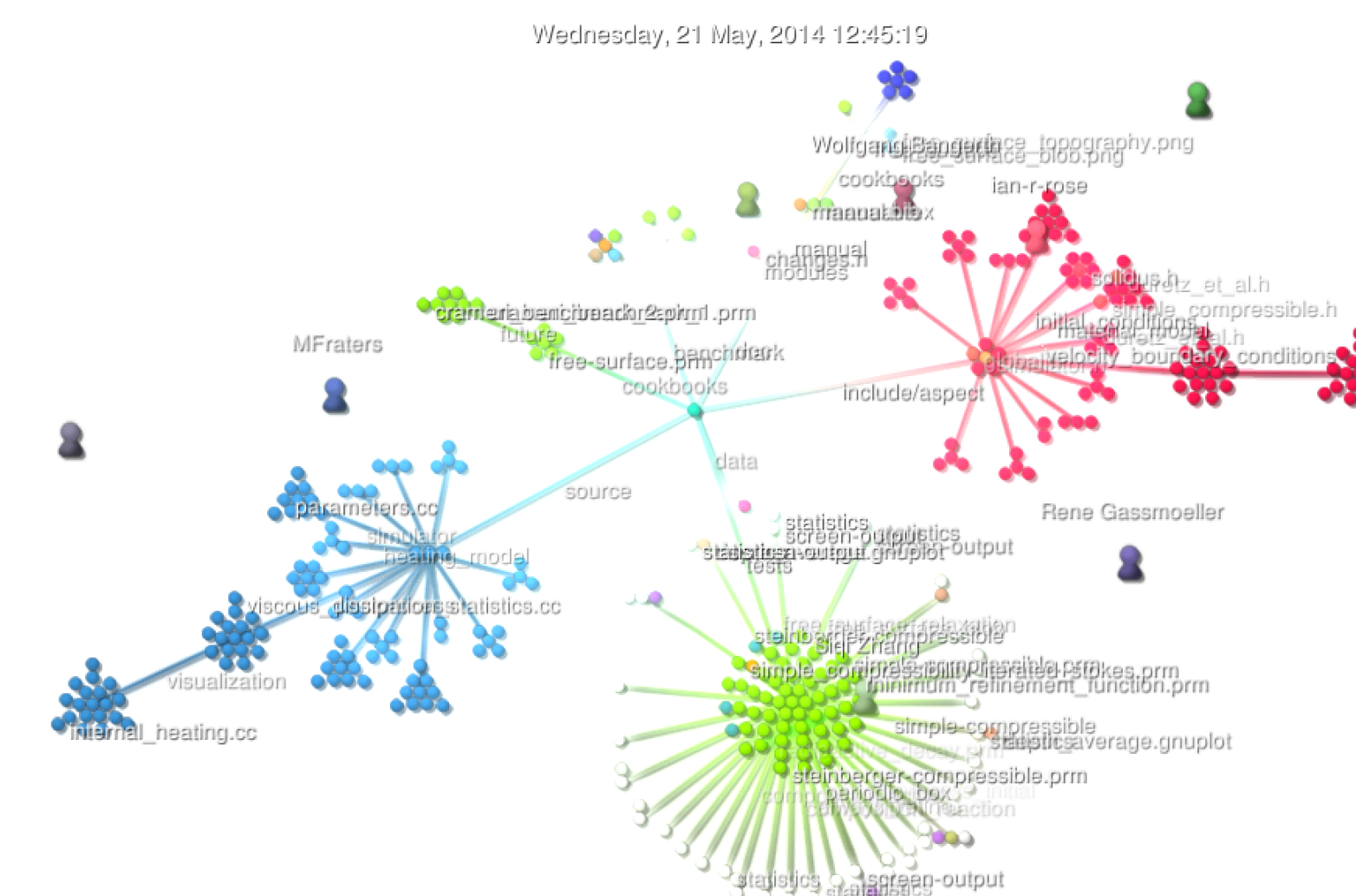
Zenodo (zenodo.org), Github/Mozilla/figshare Collaboration. Generate software citations with a DOI and stores code.



Proposed Solution

What should a successful citation system include? It should:

- allow users to easily register software
- allow queries for existing software
- find relations between software packages
- generate citations that are compatible with major journals and publication formats.



Development activity of ASPECT code, as visualized by Gource. Numerous developers are simultaneously working on shared parts of the code.

Proposed Minimal Metadata

Citation Data	Details
Unique Identifier	Uniquely identifies both the software and the version.
Software Authors	Identifies principal developers based on contribution over time. Linked to author identification tools such as ORCID, Google Scholar, or Web of Science.
Dates	Dates of creation or modification of citation.
Ontology Identifier	Links to established ontologies to help researchers find existing software, such as AGU Index Terms.
Software license	e.g. GPL, MIT, BSD. Ensure proper use and legal rights upheld.
Code access	Method of accessing the code, e.g. on a website, version control repository, commercial entity, contact information, code availability. Linked to existing access methods such as Github.
Code version	Source control revision, released code version number, or other method of uniquely identifying version of code.
Compiler version, options, platform	Depending on the code, compiler and other options can affect the results. Helps scientific reproducibility.
Software ancestor	Enables tracking of software history, revisions, branching.
Dependent libraries or code	Tracks libraries, other code this software depends on and ensures credit is assigned appropriately.

For more information or to give feedback:
geodynamics.org/cig/projects/saga



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