



**National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230**

NSF 12-058

Dear Colleague Letter - Data Citation

DATE: March 29, 2012

Subject: Data Citation in the Geosciences

Facilitating open and equal access to data and data sets is a fundamental operating principle of the Directorate for Geosciences (GEO), and the National Science Foundation (NSF) as a whole. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See [Award & Administration Guide \(AAG\) Chapter VI.D.4](#).

GEO believes that the benefit to the over-arching scientific enterprise from access to data and data sets far outweigh the burden of time and resources to an individual investigator and his or her host institution. GEO encourages data citation as a means to achieve the desired operating state for the geosciences with open and equal access to data available to all interested parties at a reasonable cost.

Principles of data citation are at various stages of maturity and adoption among scientific and engineering communities. In a 2009 report, for example, the American Meteorological Society (AMS) was urged by its *Ad Hoc* Committee on Data Stewardship Prospectus to "develop a plan for citing data referenced in publications and preserving data links for the long term." The American Geophysical Union (AGU) has taken the position that "the scientific community should recognize the professional value of data activities by endorsing the concept of publication of data, to be credited and cited like the products of any other scientific activity, and encouraging peer-review of such publications."

While many policy and practical challenges remain to be resolved and implemented, the Directorate for Geosciences encourages members of the community to lead an evolutionary transformation to establish data citation within the geosciences as the rule rather than the exception.

The Australian National Data Service lists many references to the benefits of and practices for data citation (http://ands.org.au/cite-data/resources.html#Data_Citation_Benefits). Benefits include the acceptance of research data as a legitimately citable contribution to the scientific record; permitting results to be verified and re-purposed for future study; and enabling data citation metrics to be tracked, as is done with publications. Also, data citation is one mechanism for complying with the long-standing NSF policy of data sharing (see Award and Administration Guide, [Chapter VI.D.4](#), http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4). An example of a data facility that currently assigns Digital Object Identifiers (DOIs) to datasets that investigators submit to its repository is the NSF-funded data facility Integrated Earth Data Applications (IEDA, www.iedadata.org). The DOI can then be used in publications to cite the data (e.g., see citation of the GMRT Synthesis dataset in Ryan, W.F.B., et al., G-Cubed, 2009). The DOI resolves to the bibliographic metadata of the dataset.

As transparency and reproducibility of scientific results are two established principles of science, these tenets apply to the data collected or produced in the course of a research project. However, currently few articles published in geoscience research journals cite the data used in the underlying research (e.g., *Data*

Now is the time for geoscientists to begin to meet the challenges of data citation. This may involve working with: (1) collaborators to decide which data sets are appropriate for citation; (2) data centers, libraries, repositories, and publishers to develop appropriate data citation methods and concomitant DOIs; and 3) research institutions to make data citation a common practice and a metric of value in institutional culture and practice. We urge principal investigators to discuss their efforts and suggestions about DOIs with their communities and program officers in order to accelerate progress toward data citation policies and standards. Further we encourage data citation for upcoming publications to provide transparency and opportunity to use and analyze data sets. Such openness will enrich and affirm valuable geosciences research. The result will be the development of a leading edge and robust practice of data citation that will improve and enrich the geosciences research and education enterprise.

Sincerely,
Tim Killeen, Assistant Director
Directorate for Geosciences (GEO)