

ABOUT AGU

AGU's Data Policy: History and Context

PAGE 337

Modern science is driven by data, and the implications are many: The ability to collect data more broadly and efficiently and to reuse and mine data in new ways is leading directly to new discovery, research, and understanding.

In turn, connecting data to publications is essential for ensuring the integrity of the results and facilitating reproducibility. Access to published data allows other scientists to build upon results. This access also allows the public and policy makers to have confidence in using results to ensure public safety, enforce regulations, build new policies, and foster national and international security.

In response to these trends, in addition to wanting to serve society better, major science funders have been striving to meet the aspiration of fully open data. In 2013 the U.S. Office of Science and Technology Policy mandated opening research data as well as the scientific literature. The European Union has similarly included open-data initiatives as part of its Horizon 2020 plan. The U.S. National Science Foundation (NSF) and other funders, public and private, are now requiring that proposals for funding have robust data management plans.

AGU recognized early on the importance of data to the advancement and integrity of science and to important societal issues and decisions. AGU first developed guidelines for data associated with publications in 1993, just a few years after the Web was developed and before most online publishing even began. AGU was also one of the first publishers to include supplements to its papers, which allow data archiving in the absence of a repository, and AGU staff led the effort to set community-wide standards for these supplements.

Similarly, in 1997, AGU was one of the first societies to adopt a position statement on data sharing: "AGU supports and encourages the full and open sharing of Earth and space science data for research and education. Such sharing enhances advancement of scientific and technical knowledge for education, economic advancement, public safety, and national and international security. Since measurements of geophysical quantities cannot in general be repeated, Earth and space data should be preserved, documented, and archived for future generations." This core principle has persisted in revisions and updates to this statement since then, including the most recent statement adopted in 2012 (see <http://bit.ly/data-position>).

The new data policy for AGU publications, adopted by the council in December 2013 (see

http://bit.ly/agu_datapolicy), builds on AGU's longstanding leadership role in fostering the availability of data. It brings procedures more directly in line with those of other leading journals presenting research in Earth and space sciences and aligns well with the spirit of the AGU position statement.

For example, *Science*, *Nature*, *Proceedings of the National Academy of Sciences*, and the Royal Society publications have been implementing firm policies requiring data availability upon publication for some time. These policies all mandate sharing related materials as well, including (except for *Nature*) computer codes. The Public Library of Science recently adopted a similar policy for its journals, including *PLoS One*.

An important addition to the AGU policy is that papers should include a concise statement in the acknowledgments telling the reader where core data supporting the paper's conclusions can be found, for example, in a supplement to the paper or preferably in a data facility. Such a statement helps readers and scientists easily identify the underlying data, and recent studies have shown that including such a statement increases data availability. The expectation is that all data and references will be available upon first publication of papers online.

AGU also encourages authors to provide references for specific data sets and software, and we support efforts to provide credit and recognition for these contributions to scientific knowledge. A keystone in this effort is AGU's new open-access journal, *Earth and Space Science*, which specifically encourages papers describing important data sets, methods, and models.

Similarly, AGU's policy, like those of several of the other publishers noted above, includes sharing codes used in computer programs.

The goals of sharing code are to support understanding how data were processed, interpreted, and generated; to facilitate transparency; and to accelerate further discovery in the spirit of AGU's position statement. Programmers should also receive credit for their contributions.

The availability of data or code does not require that data, software, or data products be free or that they must be passed along freely if they were purchased as part of a research project or are commercial products; indeed, many widely used data sets and software require purchase. Rather, proper attribution and pathways to find the data are required in the spirit of transparency, availability, and recognition. Similarly, there is no mandate or

expectation for training or assistance in using the data or for extensive documentation or commenting of software.

Given the complexity and volume of data, in some cases it may not be possible to comply fully with the policy. If data sets are too large to include as a supplement and if a data facility is not available, our policy requests that authors make an effort through their institution or lab to ensure the longevity of the data (for at least 5 years). This is good practice and has already been implemented by some other journals for some time. Several general data and code archives are now available. Researchers who obtain data from others or data facilities should reference the primary sources formally.

We recognize that different parts of the AGU research community have different and evolving expectations as to the core data required to support conclusions and that access restrictions may exist for legal reasons or because of confidentiality, national security, or privacy. When data or code cannot be made available for these reasons or because of the size of the archive, the reason should be indicated in the acknowledgment statement. AGU journal editors, selected from the wide community of Earth and space scientists, have discretion to approve exceptions.

The ease of data collection means that society as a whole and scientists in particular are acquiring more data than can be preserved even in the rapidly growing realm of cyber storage space. Many labs and instruments regularly collect terabytes or more of data daily. This amount of data means that communities will increasingly need to develop standards and best practices regarding which data are necessary to preserve and curate and how best to do so.

The broad Earth science community is engaged in a large effort to improve access to and use of data through efforts such as EarthCube and other projects. As types and amounts of data continue to grow, data storage and access will be an ongoing issue. AGU is working with other journals, funding agencies, and data facilities to improve communication about viable options and to help develop and promulgate best practices on formatting data and metadata.

AGU will be hosting an NSF-supported conference in early October to discuss data availability and archiving and to help develop and promulgate best practices across the Earth science journals. If you are an editor, publisher, data facility, or scientist and would like to help us in this effort, please send us an email. Our overall goal is to help researchers work productively in the evolving research environment, including before, during, and after submission of papers.

—BROOKS HANSON, Director, Publications, AGU; email: bhanson@agu.org; and ROB VAN DER HILST, Chair, Publications Committee, AGU