ɭData Availability Principles and Practice

Science requires evidence. Making data available lets other scientists replicate one's analyses, confirm results, uncover errors, or find new insights. Moreover, gathering data can be expensive and time consuming. Because the same data can be used for a range of purposes, making data available can be an efficient use of limited research resources. Doing so can also improve traceability and, thus, accountability when it comes to research findings.

These reasons and more lie behind recent efforts to promote data availability. The American Meteorological Society (AMS) recently updated its data policy guidelines (https://www.ametsoc.org/index.cfm/ams/publications/ethical-guidelines-and-ams-policies/data-policy-and-guidelines/) to require, among other things, that papers in its journals include a Data Availability Statement. Data do not need to be "open." Authors simply need to explain how to find and use the data or why, in some limited circumstances, the data are not available.

Weather, Climate, and Society (*WCAS*) combines environmental, natural, and social sciences. Data about humans and society often require different treatment and protection than do data about natural science topics, such as weather and climate. A one-size-fits-all requirement may, thus, have unintended consequences. Here we describe how the editors of *WCAS* will put the AMS data policy into practice.

To address these concerns, the AMS data policy has a specific section titled "Social Science Data and Data Restrictions," noting that access to some social science data "may be subject to restrictions, up to and including being unavailable." Authors in such cases should explain such limitations in their Data Availability Statement. Limitations are fine and expected, as illustrated in items 5 and 6 of AMS's example Data Availability Statements (https://www.ametsoc.org/index.cfm/ams/publications/authors/journal-and-bams-authors/formatting-and-manuscript-components/data-availability-statement-examples/). At the same time, to help prevent abuse of this provision, *WCAS* will ask authors to provide justification for these limitations upon submission.

Here we provide some examples to illustrate *WCAS*'s expectations. AMS recognizes that data that concern personal information or culturally or politically sensitive topics need to be carefully handled. Item 5 includes two pertinent examples:

- Example 1: "Due to its proprietary nature <or ethical concerns>, supporting data cannot be made openly available. Further information about the data and conditions for access are available at the [*repository name*] at [*insert DOI here*]."
- Example 2: "Due to confidentiality agreements, supporting data can only be made available to bona fide researchers subject to a nondisclosure agreement. Details of the data and how to request access are available from [*data manager contact info*] at [*institution where data reside*]."

While certain kinds of human-subjects research may already be subject to oversight intended to address concerns about privacy and abuse of sensitive information, other kinds might not be but perhaps should still be protected. Greater attention may need to be paid to

ɭ Denotes content that is immediately available upon publication as open access.

the ways in which separately innocuous data can be combined in unanticipated ways that might enable identity theft or an invasion of privacy. These issues could be situation dependent: for example, data on traffic patterns or roadside litter that may be reliably anonymous in a big city could be easily traced to households or individuals in a rural area or specific community. Any data-archiving system should have provisions to handle sensitive data, and in some cases human-subjects research ethics may require that only aggregate data be kept and that individual or individually identifiable data be destroyed. The Data Availability Statement can describe when such steps have been taken. A justification could explain the risks that making data available would entail or cite the relevant human-subjects research requirements.

Another concern involves data that are provided in the either explicit or implicit expectation of confidentiality. Item 5 provides another pertinent example:

- Example 3: "Due to privacy and ethical concerns, neither the data nor the source of the data can be made available."

In some economic and finance studies, for example, merely specifying in print the company that provided the data may harm that company's interests and may deter companies from providing any access. In such cases, the Data Availability Statement should state that the data cannot be made available. A justification could provide the reasons why data cannot be placed in a repository and a contact cannot be identified.

In addition to ethical and legal requirements, data availability is subject to practical limitations as well. Data-archiving resources may be insufficient to support those producing huge datasets. This is a particular issue for modelers, where the results of model runs can amount to many terabytes or more of data and could include many separate runs of the model. Reasonable accommodation can be made in these cases—for example, by providing documentation of the model parameters and methods used. Item 6 gives three relevant examples, the third of which addresses the model output problem:

- Example 1: "The dataset on which this paper is based is too large to be retained or publicly archived with available resources. Documentation and methods used to support this study are available from [*data manager contact info*] at [*institution*]."
- Example 2: "The authors were unable to find a valid data repository for the data used in this study. These data are available from [*data manager contact info*] at [*host institution*]."
- Example 3: "The numerical model simulations upon which this study is based are too large to archive or to transfer. Instead, we provide all the information needed to replicate the simulations; we used model version [*V#.#*]. The model code, compilation script, initial and boundary condition files, and the namelist settings are available at [*DOI or permanent URL*]."

The Data Availability Statement should make clear what has been archived and what steps have been taken to provide information about the data that could not be kept. A justification could describe the efforts undertaken to find a suitable repository or a comparison of dataset size with available archiving space or funds, as well as the degree to which documentation and methods allow evaluation and replication of the study.

Thoughtful data availability requirements such as AMS's benefit both the scientific community and society. Consistent policies and practices can help to reduce misunderstanding and divergent interpretations. As editors of *WCAS*, we do not wish or intend the data availability requirement to become a barrier to publication, whether because of the sensitivity of the data or because of limited institutional resources. At the same time, the exceptions to making data available should not be used by researchers as a way to evade their responsibilities. We welcome contributors to and

readers of *WCAS* to read the AMS data policy and to contact us with any questions or concerns.

Henry P. Huntington
Chief Editor

Emma Archer
Editor

Walker S. Ashley
Editor

Susan L. Cutter
Editor

Michael A. Goldstein
Editor

Carla Roncoli
Editor

Tanya L. Spero
Editor