

# Good Scientific Practice at the MPI-M

Bjorn Stevens (on behalf of the MPI-M Collegium) April 28, 2015

## Summary

This document outlines guidelines for good scientific practice at the Max Planck Institute for Meteorology (MPI-M). It has been developed to add specificity to similar guidelines ([ScientificPractice.pdf](#)) adopted by the Senate of the Max Planck Society.

The MPI-M requires that all *primary data* for an investigation be archived for a period of at least ten years. The MPI-M accepts responsibility for maintaining this archive and making it available upon request. The MPI-M also maintains the highest standards for authorship of scientific papers. The corresponding author for a published study is responsible for ensuring that their study complies with these guidelines for good scientific practice.

This policy is effective July 1, 2015. The policy will be reviewed and updated in the first half of 2016.

## Ensuring Access to Primary Data

Primary data is that which is collected in the field, in the form of notes or, more typically, measurements. For numerical investigations the primary data is the source code of the model and analysis scripts that were used in a study, as well as model input and/or configuration specifications (e.g., namelist files). The corresponding author of a manuscript is responsible for preparing an archive for the primary data on which the manuscript is based. This archive should be submitted to the institute upon final publication of the manuscript. The institute will assume responsibility for access to this archive for a period of at least ten years subsequent to the publication of the manuscript.

The primary data archive should include all primary data that is not otherwise archived. In the case that elements of the primary data are archived by others, pointers to where primary data is externally archived should be included in the primary data archive. To maintain a manageable size for the primary data archive, in the case where a primary dataset used in a study is larger than about 10 Mbytes this dataset, should be archived independently (Central IT Services, CIS, can advise on different solutions) in a manner that guarantees the institute's ability to access and distribute it for at least ten years. In many cases primary data for a study will be archived by a third party. For instance, official versions of one of our models are archived as part of the version management system and need not be independently archived. In other cases data (e.g., model output by other centers, field measurements, or satellite or paleo archives) may be taken from an acknowledged source, which if the data is described in a scientific publication, is then responsible for maintaining access to the data. In these cases the primary data need not be included in a study's primary data archive, but information as to how to access it must be provided. Duplicating information explaining how to access externally archived primary data in the README file of the primary data archive is encouraged. The MPI-M also encourages investigators to consider archiving additional (secondary) data that might be helpful in subsequent studies, for instance the numerical fields that might be used to construct particular figures, if they

think doing so could be useful for others, or for themselves, if they are later asked to reproduce some aspect of their study.

## Authorship

Authors of a study are responsible for the content of that study. This usually implies that when considering authorship the *ideas* or the *substantive contributions* of a particular study should be clear and these will determine the selection of the authors. Individuals who contribute to a study by providing the (financial) means to conduct the study, or the rendering of technical assistance (for instance by aiding in scripting, plotting or technical writing) should be acknowledged, but such acknowledgment should not take the form of co-authorship. Institute contributions (for instance through IT or administrative services or even technical help) are to be acknowledged through the author's affiliation. The provision of data, model output, or the use of a model, that has been published in a previous study does not constitute a basis for authorship.

## Ensuring good scientific practice

To ensure good scientific practice, beginning in the first half of 2015, the MPI-M will establish an archive for the primary data for all studies for which an institute member is the corresponding author. The responsibility of the author is to provide the MPI-M information manager a tape archive file containing any primary data as describe above. To guide in the preparation of the archive the institute maintains an [example archive](#), which is available upon request. The final manuscript should also contain a statement in the acknowledgment that alerts the reader to the presence of the archive by stating the following:

*Primary data and scripts used in the analysis and other supplementary information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting [publications@mpimet.mpg.de](mailto:publications@mpimet.mpg.de)*

In addition, the internal review of all papers will include a check that a proposed publication conforms with the institute policies regarding authorship and the archival of primary data.

One part of this check, is to ensure good scientific writing. Ideally a scientific paper should fully describe the methods used to enable the study to be reproduced by others. This includes, as appropriate, documentation of the model version used, sources of other primary data, configuration and input data, analysis algorithms and methods. The institute archive of model primary data is not a substitute for good scientific writing. For this reason the institute also requires that the reproducibility of a study be considered as a factor in the internal review. The reviewer should ask if the paper sufficiently detail the methods employed for someone else to reasonably replicate the study. If information that would be required to reproduce a study is not adequately documented in the draft paper, then the author should be asked why. In the case that key information cannot (for whatever reason) be provided in the paper documenting the study itself, then it is the obligation of the internal reviewer to note this in the internal review and request this information as part of archive of primary and supplementary data that the institute will maintain.

## Further thoughts on good scientific practice

Because our investigations often lead us to modify the models we use it can be helpful to document code changes by forming a versioned branch for each model configuration that is used. In this case, simply documenting the branch, or branches, upon which calculations are based will ensure that the primary data for an investigation is maintained by the MPI-M versioning system. Alternatively, changes to a standard branch can be documented separately. All models are encouraged to be placed within a versioning system that can be maintained by the institute. In some cases this may be impractical. In these cases it is important to save the model code itself, any input data it requires, and configuration specifications that might be externally specified, as part of the documentation of an investigation.

Authorship can often be a grey area. There is history of including data providers, or tool developers, in the publication of a paper as a way of recognizing their contributions. Care should be taken in this regard as it can be considered a form of honorary authorship, in which case it is not good scientific practice. When a model, or data, that is not previously documented in the literature is used in a study, then it is incumbent on the authors of a study to describe the model or the data. In this case the model and data become part of the scientific contribution of a particular study and co-authorship by the model developers and data providers would be appropriate. But if the model or data being used is described elsewhere, then the model or data should be appropriately cited, and further technical help should be acknowledged, but not through co-authorship. In those cases when the model, or the data, constitutes the main idea of the study (for instance in model description papers) the core contributors to the development of the model or the collection and curation of the data merit consideration as authors.

There will always be other grey areas, where the basis for authorship is unclear. In these cases the corresponding author should ask whether or not the proposed author contributed to important ideas that are presented in the study. Can the proposed author be considered as a co-owner of the ideas? If so then the individual in question should be considered as a co-author. Authors may also consider the possibility of delineating contributions for a study in the acknowledgment section of a paper, particularly when the contributions of the various authors are limited to only a subset of a particular study. In all cases the corresponding author is expected to discuss his or her basis for deciding who to include as authors, and the internal reviewer should feel free to ask whether or not the varied contributions to the study were appropriately acknowledged.

The institute ombudsperson provides a further resources ([website](#)) for questions regarding good scientific practice.