**Reporting Standards for Paleoceanographic/Paleoclimatic data**

Michal Kucera, Deborah Khider, Lorraine Lisiecki 12/07/2013

Discussed and adjusted during PMIP Ocean Workshop: Understanding changes since the Last Glacial Maximum in Corvallis OR, 12/06/2013

**Introduction**

The purpose of the template is to provide guidance as to the minimum requirements on documentation of material, methods, procedures and their results with respect to paleoclimatic data. The aim is to develop this draft into a formal community-endorsed recommendation. The motivation for the creation of this template is to (1) ensure replicability and transparency in scientific studies and to (2) facilitate data mining.

**Types of data**

There are three principle types of data that are relevant for the paleoceanography/paleoclimatology community:

*Original data values* – this would include raw values of all variables measured or otherwise determined on the substrate of interest, **recorded against an unambiguous sample identifier**. This includes age control points (ACP) used to derive the age model of the paleo-time series, including radioisotopes (14C, excess 210Pb), and benthic stack comparison. All data that are reported for the first time in a scientific publication should be treated as original. The procedure of obtaining original data should be documented such as to allow full replication of the result.

*Derived data values* – this category refers to values of variables derived from other variables, derived or raw, original or compiled, reported against an unambiguous sample identifier and accompanied with a documentation of the derivation procedure such as to allow full replication of the result. This category includes algorithms of any kind relating depth to age that have been used to assign ages to values of raw or derived variables.

*Compilation of data values* – this refers to raw values of all variables from published sources, gathered for the purpose of quantitative treatment, recorded against an unambiguous sample identifier and labelled with an unambiguous reference to the publication where such data were originally presented. Reporting of data compilations is required wherever these are used to derive patterns or variables from the data other than those reported in the original publication.

When reporting original raw data, the following information should be given (information in italics should ALWAYS be included):

|  |  |  |
| --- | --- | --- |
| **Information type** | **Example** | **Notes** |
| *Unambiguous sample identifier* | - *Identifier, coordinates and water depth of the sediment core or sample—Lat/Lon values should be reported in decimal format.*  *- Position of the analysed sediment segment in the core (include position as top and bottom rather than mid-point)* | All data, original, derived or compiled must be reported against depth in core. |
| Collection Method | - Gear type (multicorer, piston corer)  - Volume of sediment analysed  - Sampling methods (u-channel, syringe, slice) | Wherever any of the items below is the same for all samples in the study, then a reference to the methods section of the study is sufficient. |
| Method of sample processing | - Method of sediment disintegration and washing  - Sieve fraction analysed  - *Cleaning methodology* |  |
| Method of obtaining raw values | - Information on the taxonomy that has been followed  - Instrumentation  - *Assign a unique identifier to the sample* (create a number for each measurement, the identifier should also provide information about the laboratory in which the analysis was carried out).  - Information on preservation state |  |
| Raw data values | -Total number of specimens counted/picked in each sample  - *Report units and correction applied*  *- Report analytical/sample uncertainty* |  |

When reporting derived data, the following additional information should be given:

|  |  |  |
| --- | --- | --- |
| Information type | Example for assemblage counts of planktonic foraminifera | Notes |
| *Method of obtaining derived data values* | *Method of derivation from the original data* | Documented sufficiently to allow full replication |
|  | Unambiguous interpretation of the derived variable | For example, to label the outcome of a transfer function reconstruction as “SST” is insufficient, because it does not state to which season or depth this variable refers to. |
| *Derived data values* | *Value of the derived variable(s) with its associated uncertainty* |  |

When reporting a compilation of data, the following information should be given in addition to the information provided for raw or derived data:

|  |  |  |
| --- | --- | --- |
| Information type | Example for assemblage counts of planktonic foraminifera | Notes |
| *Reference to source of information* |  |  |
| *Whether raw data are available* | In many cases, percentages rather than raw counts have been reported. Such data are useful, but they represent derived data. |  |

**Archiving**

The data should be archived in easily accessible database administered by an editorial board. The database should included on on-line system for (1) the reporting of errors, (2) a community-driven quality flag system, and (3) the update of datasets as new data become available. Updates to a particular dataset should be assigned a version number with a link to the old and new references.

The editorial board should also be tasked in identifying and archiving older datasets.