

Recommendations for format of data associated with publications

This document recommends a standard for data and metadata that will comply with new open data requirement. Software is available in IDL, MATLAB and PYTHON which follow this standard.

Approach adopted:

1. We use the HDF5 scientific data format
2. Figures are represented as groups of 1D, 2D, 3D Arrays in HDF5 (terms in *italics* are HDF5 entities)
 - One HDF5 *file* per figure
 - Multiple HDF5 *groups* (or actually *subgroups* of the *root group*) are used to represent distinguishable elements in figure.
 - Array data itself is stored as HDF5 *datasets*
 - *Dataset* names – describe the x, y, z in each data group
 - Metadata, as listed below, is represented as HDF5 *attributes* attached to *group* (*root group* or otherwise) or to a *dataset*
 - We define metadata at file(figure), group and data levels (as shown below)

Y(X) (file/figure level metadata shown here in blue)

- file = name of hdf5 file (without extension) text that should include the figure number/identifier
- fig_description = could be the figure caption or a short form of the caption
- fig_source = string that identifies the manuscript or publication (examples: 'JA-14-21' or 'NF 55 023012 2015')
- date = Date/time written (automated)
- user_fullname = Writer's full name
- user_id = Writer's account name (automated)
- n_groups = Number of data groups
- comment = (attribute of root group) string, anything you want to say, including description of plot type (e.g. polar, rectilinear)
- **For each dataset or data group (group/set metadata/attributes in red)**
 - group_name = Data set (string, could be legend or similar text)
 - plot_graphics = plotting info (color, symbol, line type)
 - **Dataset (data metadata in green)**
 - **x = x data values**
 - x_axis = text, short label
 - x_name = text, longer description (optional)
 - x_type = string (data type optional?)
 - x_units = units of x data
 - nx = number of x data values
 - **y = y data values**

- `y_axis` = text, short label
- `y_name` = text, longer description (optional)
- `y_type` = string (data type optional?)
- `y_units` = units of y data
- `ny` = number of y data values