



An Open Data Standard for Cloud Particle Images and Reference Software to Produce and Validate Compliant Files

Graeme Nott and Dave Sproson
FAAM Airborne Laboratory, Cranfield, UK

The use of airborne cloud imaging probes has resulted in decades of in situ particle-by-particle data taken across the gamut of pristine and anthropogenically-modified cloud types around the globe. Image data from such probes is recorded in proprietary and instrument- or system-specific formats. Binary formats have evolved to minimise the stress on, now possibly outdated, hardware and communication systems that must operate in the difficult aircraft environment. This means that there is a significant knowledge and technical barrier to new users, particularly for those that are not from fields that have traditionally used such cloud data. Processed image data is generally available, however this precludes the application of more advanced or specialised processing of the raw data. For example, historical cloud campaigns of the 1970s and 80s used imaging probes for cloud microphysical measurements at a time when satellite measurements of those regions were sparse or nonexistent. Fields such as atmospheric processes modelling, climate modelling, and remote sensing may well benefit by being able to ingest raw cloud particle data into their processing streams to use in new analyses and to address issues from a perspective not normally used by those in the cloud measurement community.

The Single Particle Image Format (SPIF) data standard has been designed to store decoded raw binary data in netCDF4 with a standardised vocabulary in accordance with FAIR Guiding Principles. This improves access to this data for users from a wide range of fields and facilitates the sharing, refinement, and standardisation of data processing routines. An example is the National Research Council of Canada (NRC) Single Particle Image Format (SPIF) conversion utility which converts binary data into SPIF files. In a similar fashion to the Climate and Forecast (CF) Conventions, SPIF defines a minimum vocabulary (groups, variables, and attributes) that must be included for compliance while also allowing extra, non-conflicting data to be included.

The ability to easily check files for compliance to a data standard or convention is an important component of building a sustainable and community supported data standard. We have developed a Python package called *vocal* as a tool for managing netCDF data product standard vocabularies and associated data product specifications. *Vocal* projects define standards for netCDF data, and consist of model definitions and associated validators. *Vocal* then provides a mapping from netCDF data to these models with the Python package *pydantic* being used for compliance checking of files against the standard definition.

We will present the *vocal* package and the SPIF data standard to illustrate its use in building

standard compliant files and compliance-checking of SPIF netCDF files.