

EGU22-8262

<https://doi.org/10.5194/egusphere-egu22-8262>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Data Access Made Easy: flexible, on the fly data standardization and processing

Mathias Bavay¹, Charles Fierz¹, and Rodica Nitu²

¹WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

²World Meteorological Organization, Geneva, Switzerland

Automatic Weather Stations (AWS) deployed in the context of research projects provide very valuable data thanks to the flexibility they offer in term of measured meteorological parameters, choice of sensors and quick deployment and redeployment. However this flexibility is a challenge in terms of metadata and data management. Traditional approaches based on networks of standard stations can not accommodate these needs and often no tools are available to manage these research AWS, leading to wasted data periods because of difficult data reuse, low reactivity in identifying potential measurement problems, and lack of metadata to document what happened.

The Data Access Made Easy (DAME) effort is our answer to these challenges. At its core, it relies on the mature and flexible open source Meteolo meteorological pre-processing library. It was originally developed as a flexible data processing engine for the needs of numerical models consuming meteorological data and further developed as a data standardization engine for the Global Cryosphere Watch (GCW) of the World Meteorological Organization (WMO). For each AWS, a single configuration file describes how to read and parse the data, defines a mapping between the available fields and a set of standardized names and provides relevant Attribute Conventions Dataset Discovery (ACDD) metadata fields, if necessary on a per input file basis. Low level data editing is also available, such as excluding a given sensor, swapping sensors or merging data from another AWS, for any given time period. Moreover an arbitrary number of filters can be applied on each meteorological parameter, restricted to specific time periods if required. This allows to describe the whole history of an AWS within a single configuration file and to deliver a single, consistent, standardized output file possibly spanning many years, many input data files and many changes both in format and available sensors. Finally, all configuration files are kept in a git repository in order to document their history.

A basic email-based interface has been developed that allows to create new configuration files, modify an existing configuration file or request data on-demand for any time period. Every hour, the data for all available configuration files is regenerated for the last 13 months and stored on a shared drive so all are able to access the current data without even having to submit a request. A table is generated showing all warnings or errors produced during the data generation along with some metadata such as the data owner email in order for the data owner to quickly spot

troublesome AWS.