



Towards a community standard for DAS metadata: Latest advances within the Geo-INQUIRE project

Javier Quinteros¹, Gilda Currenti², Michele Prestifilippo², John Clinton³, Pascal Edme³, Christopher Wollin¹, Diane Rivet⁴, Shane Murphy⁵, Jonathan Schaeffer⁶, Helle Pedersen⁶, and Angelo Strollo¹

¹Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, 2.4 Seismology, Potsdam, Germany (javier@gfz-potsdam.de)

²Istituto Nazionale di Geofisica e Vulcanologia (INGV), Rome, Italy

³Swiss Seismological Service (SED), ETH Zürich, Switzerland

⁴Université Côte d'Azur, Observatoire de La Côte d'Azur, CNRS, IRD, Géoazur, Valbonne, France

⁵Ifremer, Plouzane, France

⁶Centre National de la Recherche Scientifique (CNRS), France

In the last 5 years, the seismological community has experienced an impressive growth in the novel Distributed Acoustic Sensing (DAS) technique, in terms of both the number of experiments and the generated data volume. DAS experiments can generate data at a much finer resolution in space and time, than is seen with standard acquisition techniques. This creates challenges not only for data centres regarding the data management, but also for users that need to access and process this data. The current seismological standards for data and metadata formats, as well as community services specifications, are not capable of handling these datasets in an effective way - not unexpected considering that data volumes and 'station' numbers are orders of magnitude larger than typical broad-band experiments.

Within the context of the "Geosphere INfrastructures for QUestions into Integrated REsearch" project (Geo-INQUIRE, <https://www.geo-inquire.eu/>), we have defined a roadmap to advance towards community standards for some of these aspects. The main objective of improving the FAIRness of these datasets was separated in 3 steps. First, we defined how to archive downsampled versions of the datasets in standard community formats (i.e. miniseed and StationXML). Second, we wanted to support the definition (and foster the adoption) of a new metadata standard for DAS experiments based on the outcome of the DAS Research Coordination Network group (DAS-RCN), an initiative led by US researchers. And finally, we wanted to work on the definition of a data format capable of providing fast processing on the data centre side, as well as being able to provide the data to the user to be processed elsewhere.

We worked with 3 datasets from the Global DAS Month (February 2023), acquired by INGV, ETH and GFZ. These datasets had been published and made available in different non-standard formats. We used these experiments as test cases to later apply this workflow to the datasets generated by the Transnational Access Calls of this project at a variety of Research Infrastructures across Europe (e.g. at Etna, Bedretto, Ligurian, Madeira, Irpinia, and others).

Regarding the data volume and lack of standardisation, we have improved “dastools”, a software package developed at GFZ, to read DAS data in proprietary formats from different manufacturers and convert it to standard miniseed. Downsampling in time and space it provides a reduced version ready to be archived in seismological data centres.

Regarding metadata formats, we included in “dastools” the support for the DAS-RCN proposal, discussed and agreed within the community during the last 3 years. We can generate a first draft version of the metadata based on the information available in the raw data of the experiment. We also added a converter to StationXML (still beta) in order to support each step of the archival of a downsampled version of the DAS data.

We plan to work soon on the definition of a data format for this type of experiment as it is a key part of our project.

In parallel, we’ve just started the development of a Seedlink plugin (real-time transmission) to be deployed and tested at interrogators.