

An all-in-one web tool to apply CTD quality control, format data, and generate metadata under SeaDataNet criteria

Pablo Otero, Instituto Español de Oceanografía, Centro Oceanográfico de Vigo (Spain),

pablo.otero@ieo.es

Gonzalo González-Nuevo, Instituto Español de Oceanografía, Centro Oceanográfico de Vigo (Spain),

gonzalo.gonzaleznuevo@ieo.es

Elena Tel, Instituto Español de Oceanografía, Servicios Centrales, Madrid (Spain), elena.tel@ieo.es

I. INTRODUCTION

Professional data centres of the Pan-European region active in marine data collection, in particular National Oceanographic Data Centres (NODCs), routinely share diverse datasets through SeaDataNet, a distributed Marine Data Infrastructure. This network provides on-line integrated databases of standardized quality.

Sharing marine data through SeaDataNet is an optimal way of ensuring FAIR principles: Findable, Accessible, Interoperable and Reusable. However, submission of data to this infrastructure demands a set of technical tasks that cover quality control processing, adoption of common vocabularies, implement file format standards and preparation of associated metadata. Although common software tools are made available to NODCs to facilitate data and metadata preparation (see <https://www.seadatanet.org/Software>) these tasks continuous to be complex and time-consuming. For example, the technician must use MIKADO software to generate INSPIRE-complaint metadata and adopt SeaDataNet Common Vocabularies; NEMO software to transform the original file to a common data transport format; and OCTOPUS software to check the compliance of a file. Additional software must be used to quality check and properly flag each individual record.

The Instituto Español de Oceanografía (IEO) acts as NODC and submits diverse datasets collected by the Spanish oceanographic fleet to SeaDataNet infrastructure. Until the present, more than 35,000 CTD vertical profiles have been processed and submitted. Taking into account that this kind of data usually follows the same format and involves similar processing, a web application has been developed to perform all these tasks straightforward. The aim is to save processing time and to reduce human interaction that could lead to errors or lack of uniformity in data.

II. METHODOLOGY

A Python application (<http://ctdcheck.ieo.es/>) has been coded in combination with Flask, a lightweight WSGI web application framework, to get the app up and running. Parts of the code are based on the CoTeDe^[1] and Python-CTD^[2] packages. The source code is available at <https://github.com/PabloOtero/CTDChecker>.

The software parses a set of csv files and homogenise variables (names, column order, precision, etc.). Different quality test are applied to each vertical profile. Most of the tests simply reproduce the procedure recommended by GTSP^[3] and SeaDataNet manual^[2]. For example, the application checks if the date is valid, the station is at sea, records are in the expected range (globally and regionally), compares with climatology, detects spikes and performs a density inversion test among others. Depending on the success during the test, a flag is assigned to each individual record and also an overall flag to each parameter and to the entire profile. The application also plots temperature and salinity profiles and their comparison with climatologic values. Bad or suspicious points are highlighted.

Processed records are formatted to the SeaDataNet MEDATLAS auto-descriptive ASCII format. Finally, Common Data Index (CDI) metadata are generated, one per vertical profile.

As input, the software requires a Cruise Summary Report (CSR) metadata file associated to the dataset. In this way, software extracts main info from the cruise (e.g., ID, principal investigator, vessel, dates, location, project...) and ensures consistency when generating CDI metadata and correct inclusion of this information in the header of the output data file.

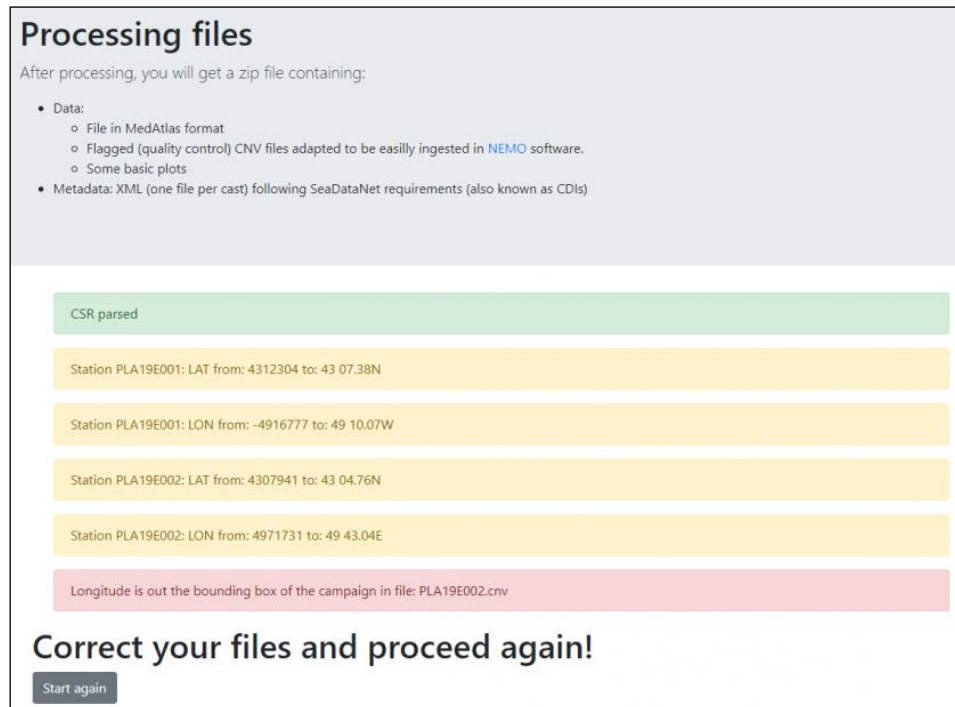


Figure 1: Screenshot example of an output of the application, showing some warnings (geographical coordinates have been reformatted and therefore requires further revision) and error (location is out the bounding box of the campaign informed in the CSR metadata file).

III. CONCLUSIONS

An all-in-one web tool to apply quality control to CTD vertical profiles, reformat data, and generate metadata under SeaDataNet criteria has been created. The objective of this tool is saving time and reduce human errors. The reason to implement a web-based instead a desktop application is to avoid any installation process, skip upgrades and use cross-platform. The application always requires the previous existence of a CSR file in the SeaDataNet infrastructure with the aim of create completely coherent metadata (CDIs) and full-descriptive header files.

REFERENCES

- [1] Castelao, G. P., (2020). A Framework to Quality Control Oceanographic Data. Journal of Open Source Software, 5(48), 2063, <https://doi.org/10.21105/joss.02063>
- [2] Fernandes, F. (2014, August 25). python-ctd v0.2.1 (Version v0.2.1). <http://doi.org/10.5281/zenodo.11396>.
- [3] UNESCO–IOC, 2010. GTSP Real-Time Quality Control Manual. First revised edition. UNESCO–IOC. United Nations Educational, Scientific and Cultural Organization 7, Place de Fontenoy, 75352, Paris 07 SP. IOC/2010/MG/22Rev.
- [4] SeaDataNet, 2010. Data Data Quality Control Procedures. Version 2.0 May 2010. Available at <https://www.seadatanet.org>