ID23- SIMPLIFYING QUALITY CONTROL AND STANDARDIZATION OF CTD DATA UNDER SEADATANET REQUIREMENTS

PABLO OTERO[1], GONZALO GONZÁLEZ-NUEVO[2] AND ELENA TEL[3]

1 Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, Subida a Radio faro 50, 36390, Vigo (Spain), +34 98649211, pablo.otero@ieo.es
2 Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, Subida a Radio faro 50, 36390, Vigo (Spain), +34 98649211, gonzalo.gonzaleznuevo@ieo.es
3 Instituto Español de Oceanografía, Servicios Centrales, c/ Corazón de María 8, 28002, Madrid (Spain), +34 913421100, elena.tel@ieo.es

Abstract
Sharing marine data through SeaDataNet infrastructure ensures preservation and promotes reusability. 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 Data Centres and/or End Users to facilitate data and metadata preparation, these tasks continuous to be complex and time-consuming. To speed-up this process, a Python-Flask web application is presented here to quality check and create metadata and data according to SeaDataNet requirements. The web tool focuses on CTD vertical profiles, although code could be easily adapted to process other type of records.

Keywords
CTD, Python, SeaDataNet, quality control, interoperability.

INTRODUCTION
SeaDataNet infrastructure manages the large and diverse datasets collected by the oceanographic fleets and the automatic observation systems in the countries bordering the European seas. This infrastructure agglutinates the National Oceanographic Data Centres (NODCs) of 34 countries, active in data collection, and provides integrated datasets of standardized quality on-line.

The Instituto Español de Oceanografía (IEO) acts as NODC and submits to SeaDataNet both data and metadata coming from different types of sensors. 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 edit and generate XML metadata entries that follow the ISO 19139 Schema, are INSPIRE-complaint and adopt SeaDataNet Common Vocabularies; NEMO software to convert from any type of ASCII format to common data transport formats; and OCTOPUS software to check the compliance of a file. Additional software must be used to quality check and properly flag each individual record.

Taking into account that hundreds of CTD vertical profiles are recorded by IEO each year and that their associated 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.

METHODS AND RESULTS
Software to parse, quality check, flag and format CTD data and their associated metadata has been coded in Python combined with Flask. Flask is a lightweight WSGI web application framework, easy to get started because there is little boilerplate code for getting a simple app up and running.

To proceed, the user must provide as input to the application a XML file containing information of the cruise, also known as Cruise Summary Report (CSR). The software parses the CSR metadata file and extracts information about the cruise, dates, marine region, the name of the researcher, the project associated to the data acquisition and so on. This information will be used later to ensure that common information is translated unequivocally to formatted data files and also to new metadata.

On the other hand, the user provides a set of CTD files in ASCII format. The software parses CTD files and split the profile to select downcast, check depth, geographical position and time, rename parameters according with common vocabularies, check pressure and apply different quality test. Most of the tests done by the application simply reproduce the procedure recommended by GTSPP[1] 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. Some of these tests have been implemented in the code by adapting the CoTeDe[3] and Python-CTD[4] packages. Depending of 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. Data are also formatted to the SeaDataNet MEDATLAS auto-descriptive ASCII format.

Finally, ISO 19139 XML metadata are created for each individual profile following a template. These metadata are known as Common Data Index (CDI) and gives users a highly detailed insight in the availability and geographical spreading of marine datasets that are managed by the SeaDataNet data centres.

CONCLUSIONS
The development of a web application to create CTD data and metadata according with SeaDataNet standards has demonstrated to save time and reduce human errors. With the user experience in mind, a web-based application was chosen to avoid any installation process, skip upgrades and use cross-platform. Obviously, this kind of app requires connectivity to internet and could be useless on board some research vessels. However, the app pursues to create SeaDataNet-conformant data and metadata files; and to do this, a CSR file must already exists in the SeaDataNet infrastructure. This implies processing necessarily in delay mode, minimizing the problem of internet dependency.

REFERENCES
[1] UNESCO–IOC, 2010. GTSPP 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.
[2] SeaDataNet, 2010. Data Data Quality Control Procedures. Version 2.0 May 2010. Available at https://www.seadatanet.org
[3] http://cotede.castelao.net
[4] Fernandes, F. (2014, August 25). python-ctd v0.2.1 (Version v0.2.1). http://doi.*org/10.5281/zenodo.11396*.