NAUTILOS Data Management Infrastructure

Gabriele Pieri [0000-0001-5068-2861], Institute of Information Science and Technologies, National Research Council of Italy (CNR-ISTI) (Italy), Gabriele.Pieri@isti.cnr.it
Marco Tampucci [0000-0001-5550-9881], CNR-ISTI (Italy), Marco.Tampucci@isti.cnr.it
Federico Volpini, CNR-ISTI (Italy), federico.volpini@isti.cnr.it
Marco Alba, ETT Solutions (ETT) (Italy), marco.alba@ettsolutions.com
Francesco Misurale [0000-0003-4004-3155], ETT (Italy), francesco.misurale@ettsolutions.com
Antonio Novellino [0000-0002-3263-5583], ETT (Italy), antonio.novellino@ettsolutions.com

Introduction and problem description

NAUTILOS Project (New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation) is an H2020 project devoted to fill-in marine observation and modelling gaps for chemical, biological and deep ocean physics variables through the development of a new generation of cost-effective sensors and samplers. It brings together many partners with multidisciplinary expertise ranging from ocean instrumentation development and integration, ocean sensing and sampling instrumentation, data processing, modelling and control, operational oceanography and biology and ecosystems and biogeochemistry such, water and climate change science, technological marine applications, and research infrastructures. In this context, the data generated within the NAUTILOS project are meant to be highly complementary to the existing observing systems and thus with high impact and value for any NAUTILOS stakeholder.

One specific goal of the project is to facilitate interoperability towards stakeholders (e.g. National Oceanographic Data Centres and thematic and international data assembly repositories, EMODnet, SeaDataNet, Copernicus Marine Environmental Monitoring Services), hence the project has organized data management according to a workflow that makes its data as ready as possible. This includes the adoption and application of standardized and harmonized standards and formats.

Data management Back-end

NAUTILOS infrastructure is organized with a data layer designed to manage data and data products, a service layer to organize them to offer the services, and an application layer i.e., the end-user interface (and features) to access and use the developed and provided services. The data layer is in charge of ingesting datasets and is able to make available data in different formats, like csv, txt, netcdf, etc.

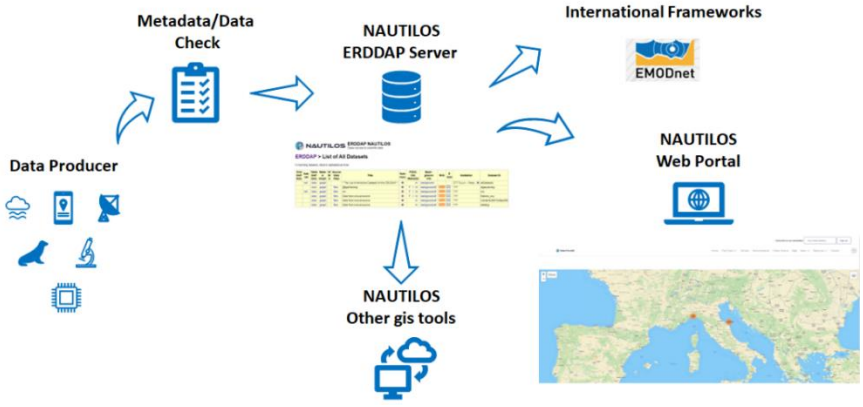

Figure 1: NAUTILOS Data flow and infrastructure.

Figure 1 shows NAUTILOS data publication workflow that processes data towards a multi-step tier approach where a first round is visible to internal users and permit to validate the new flow with the proper application of harmonized standards and quality checks/quality flags, etc. The next round makes this workflow operational, hence there is a direct streamless flow from the deployed platform

towards the NAUTILOS backend. The level of discovery is still limited to NAUTILOS users for a second level of checks. The last step is the full publication of the data (as well as the organization of the data into collection that can be easily consumed by stakeholders). A single dataset source may be integrated in more products and the service layer organizes data within the NAUTILOS datasets and data products publication services and catalogues (e.g. ERDDAP).

While implementing these three tiers, NAUTILOS applies special measures to improve and ensure data interoperability: data format, metadata and data services. As an example, to improve data provenience visibility and NAUTILOS data citation, it adopts digital object identifier (doi). Notably a DOI is also important when datasets are used from large databases that are subjected to changes, e.g., annual reprocessing, versioning, etc.

Data management Front-end

The development of the NAUTILOS Front-end is based on a web portal[1] providing an end-to-end data and information management, search, discovery and access system. The portal implements a WebGIS viewer with a dynamic map that provides the user with zooming and panning features, as well as with features to include and select data and data-products. Each dataset is enriched with metadata, and tools are integrated to pre-view these metadata contents. The front-end also provides several functionalities to filter the data and change their rendering behaviour. For each dataset, the front-end provides a set of dedicated options that can be used to filter data to be shown. Moreover, selecting the "time series" option changes the data representation providing a view of the data that evolves over depending on registered time. In order to prevent a performance drop the front-end is able, for each dataset, to automatically group near points of interest and explode them when zoom-in. For specific datasets, users can change the rendering colour depending on stored data, i.e. for the "mini-drifter" datasets, users can change the polyline segments depending on drifter speed or measured temperature (Figure 2). Finally, displayed data can be downloaded as reports containing all the activated layer data and relative diagrams.

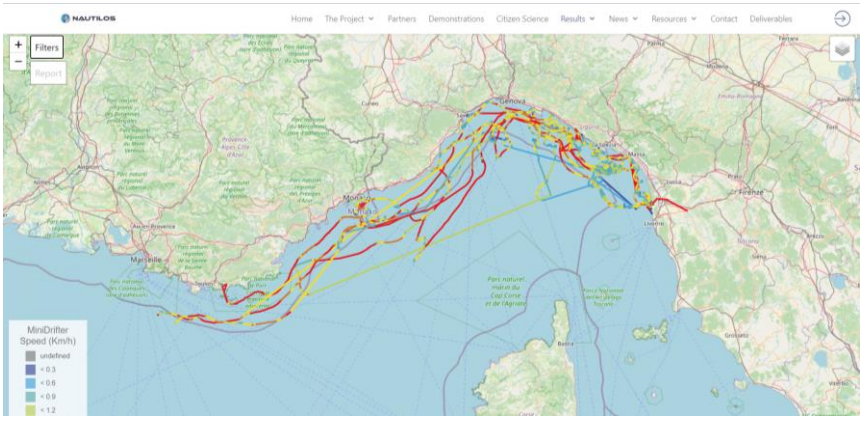

Figure 2: NAUTILOS front-end example.

Conclusions

NAUTILOS is developing new tools to unlock measurement and sharing of chemical, biological and deep ocean physics variables with a focus on places and parameters that are still missing in the standard "networks" and programs. The NAUTILOS data management system is adopting the latest development from the international community and is developing updates and new workflows to advance these standards further. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS).

[1] https://nautilos-h2020.eu/data-portal/