Integration of Biology Sensor Outputs in the European Marine Observation and Data Network

Véronique Créach, Cefas (UK), veronique.creach@cefas.co.uk
Luis Felipe Artigas, CNRS-LOG (France), felipe.artigas@univ-littoral.fr
Alain Lefebvre, Ifremer (France), Alain.Lefebvre@ifremer.fr
Markus Lindh, SMHI (Sweden), markus.lindh@smhi.se
Fabien Lombard, CNRS-LOV (France), lombard@obs-vlfr.fr
Klas Ove Möller, HZG (Germany), klas.moeller@hzg.de
Lennert Schepers, VLIZ (Belgium), lennert.schepers@vliz.be
Jukka Seppälä, SYKE (Finland), jukka.seppala@ymparisto.fi
Melilotus Thyssen, CNRS-MIO (France), melilotus.thyssen@mio.osupytheas.fr

JERICO-S3 is a new building-block in the "Joint European Research Infrastructure network of Coastal Observatories" concept. It has been initiated in 2011 and aims at addressing the challenge of observing the complexity and high variability of coastal and shelf seas from local to Pan-European level. An important target of ***JERICO-S3*** is to provide the research community with continuous, interoperable and more valuable coastal data, coupling physical, biogeochemical and biological information into an ecosystem approach, to support the EU Water Framework and Marine Strategy Framework Directives, Regional Seas conventions such as OSPAR and HELCOM, and to provide operational marine services of high societal value.

From Sensors to service access

Challenges involving data processing, quality control, standardisation are often hidden behind the exciting step of developing and applying new technologies. However, good data management is the key step which allows scientists to find easily the data, trust them and used them according to FAIR (Findable, Accessible, Interoperable, Reusable) principles in innovative approaches for environmental studies. During the last 15 years, the development of a new generation of biological sensors has drastically changed the studies of marine plankton from lab bench work to *in situ* and real-time observations. The new instruments have also demonstrated a great versatility. Many studies have already demonstrated installation in buoys, scientific vessels, and commercial ships, and that measurements can be made at high frequency. Target organisms, from bacteria to microphytoplankton and zooplankton, can now be optically characterized and/or photographed and archived. Consequently, scientists are now facing the difficulty to handle a large amount of data which need to be processed rapidly and harmonised before being stored in databases and then made accessible to different scientific/environmental management communities. One of the aims of JERICO-RI is to provide a framework for the data flow following the FAIR principles and to increase their visibility using the European Data Infrastructures. In order to draw up best practices in data management, we encourage data users and experts operating the sensors to (a) develop standardised protocol descriptions and minimal technical metadata elements for effective re-use;

(b) identify and extend the appropriate vocabularies; (c) identify tools for data integration and platforms for trust-worthy long-term archival; (d) map sensor-specific formats to standardized data formats to be ingested by European data infrastructures; (e) discuss the correct scale for meaningful spatial and temporal data aggregation. According to the size of the user community, the readiness of the technologies, and the relevance of information for the ecosystem system approach in the monitoring of coastal seas, three main types of biological sensors have been selected for data implementation in *JERICO-S3*:

- plankton imagery: different cameras and types of *in situ* or inflow machines, image acquisition and analytical tools have been improved over the last few years for addressing the biodiversity of plankton communities. Confronted with a huge number of images, platforms such as EcoTaxa have been developed to sort and annotate taxonomically the organisms using automated classification algorithms based on random forest or deep learning approaches.
- multispectral fluorometry: it had addressed the bulk phytoplankton community response, based on their pigment composition, to multi-wavelength excitation, rather than taxonomy. The technology is now more and more regularly implemented on board of research vessels and in FerryBox systems. Data produced has the same frequency as classical single excitation fluorescence sensor. If some approaches have been developed for automated estimation of pigment groups (from scientist or companies), the data have not yet been hosted by a European data infrastructure, and data format and vocabulary still need to be better defined.
- on-line automated flow cytometry: this single-cell/particle inflow technology becomes more and more popular for optically defining phytoplankton functional groups over the whole size range, due to the versatility and the autonomy of the instruments. Based on the work already done during previous European projects (DYMAPHY, JERICO-NEXT and SeaDataCloud), a first version of common vocabulary (BODC) and data format were established (SeaDataNet) and needs to be confirmed and accepted by a growing scientific community.

As these three technologies reach today different stages of data management, JERICO-S3 aims at providing the support to comply with international standards, to establish control quality procedures, to harmonise the data and build tools to help the data flow towards the data platforms before inviting the partners of the consortium to apply the different procedures to their biological data collected in the North Sea, English Channel, Cretan Sea, north west Mediterranean Sea and Baltic Sea. Combining high frequency biological data with physical and biogeochemical data with complementary approaches (e.g: remote sensing and modelling) is the key for an integrating monitoring and a better understanding of the changes in the ecosystems.