OBIS 2.0: Real-time integration, quality control and analysis of rich marine data streams

Pieter Provoost and Ward Appeltans, OBIS Secretariat

For almost two decades, the Ocean Biogeographic Information System (OBIS) has played a key role in the mobilisation and sharing of marine biodiversity data. The increasing demand for near real-time rich datasets to support conservation, human health, and the blue economy, as well as rapidly increasing data volumes due to technological advancements in observing systems, have triggered improvements in the data model and a reengineering of the OBIS platform. The new data platform can now handle complex high-volume datasets and make them available to users with very little delay.

The new OBIS data portal (https://obis.org) provides statistics on species (and higher taxa), datasets, OBIS nodes, areas (EEZ, LME, IHO Sea Area, EBSA, World Heritage Sites), data provider institutions and countries (based on origin of data providers), while the new interactive OBIS mapper (https://mapper.obis.org) allows the user to visualise, filter and download the data and metadata.

Real-time data integration

Traditionally, the OBIS database was rebuilt on a quarterly schedule. This caused significant delays between the initial publishing of data and these data being available as part of the integrated dataset. Also, because quality control was often performed months after publishing, issues often remained unaddressed. The OBIS 2.0 data system architecture allows for continuous harvesting of data sources, and datasets are now fetched, processed and made available as part of the integrated dataset in less than an hour after publishing. The data flow is illustrated in Figure 1.

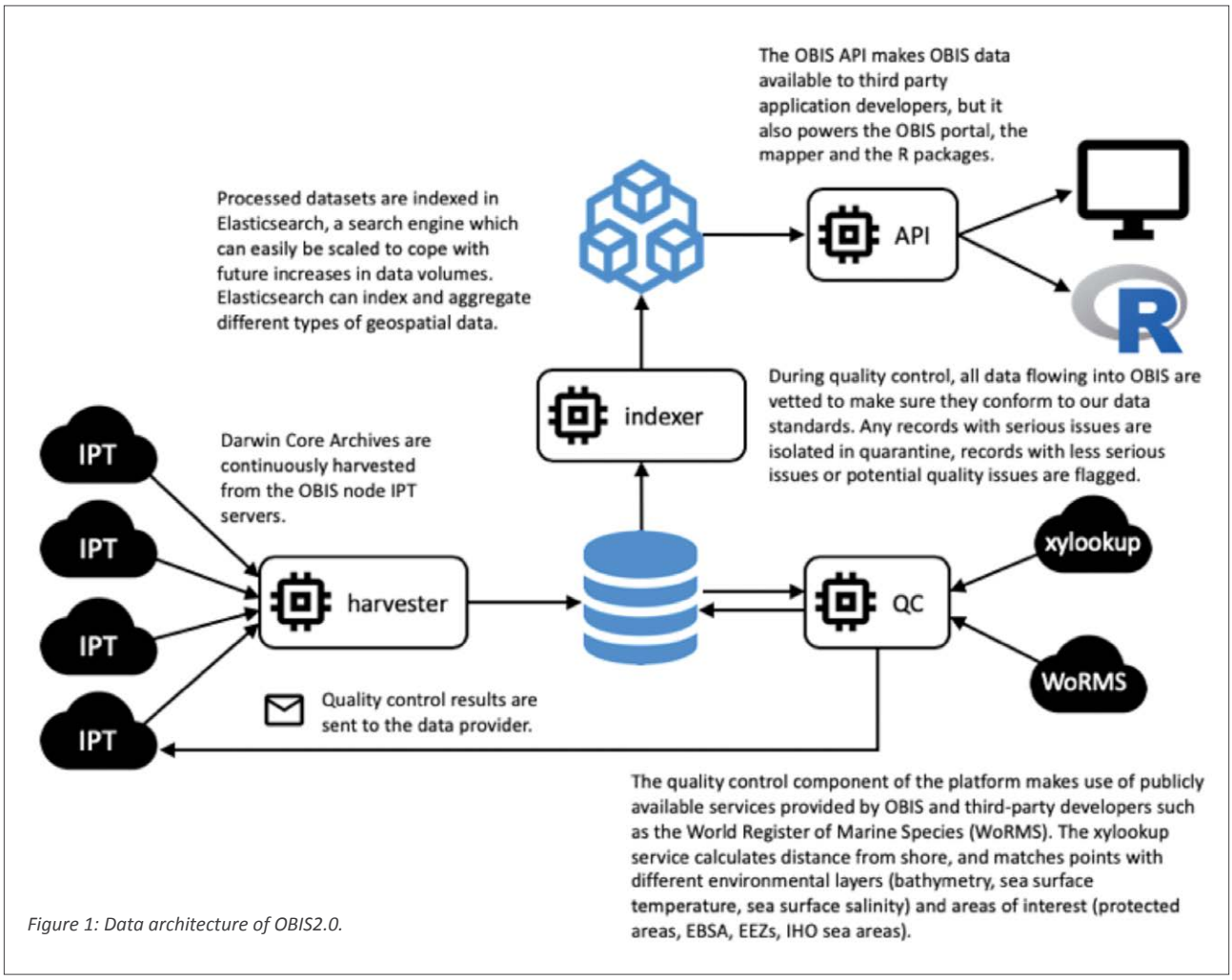

Figure 1: Data architecture of OBIS2.0.

Structured data

Since its inception, OBIS has focused on recording species occurrences to document what lives where in our oceans. However, OBIS data providers also record sampling methodology and measure biotic and abiotic variables. Storing these data in a machine-interpretable way is essential for OBIS to become an effective data-sharing platform which supports the development of Essential Ocean Variable (EOV) data products. To make this possible, the OBIS network has created a Darwin Core Extension which supports exchanging rich datasets containing a variety of measurements and facts linked to community-developed vocabularies (Figure 2).

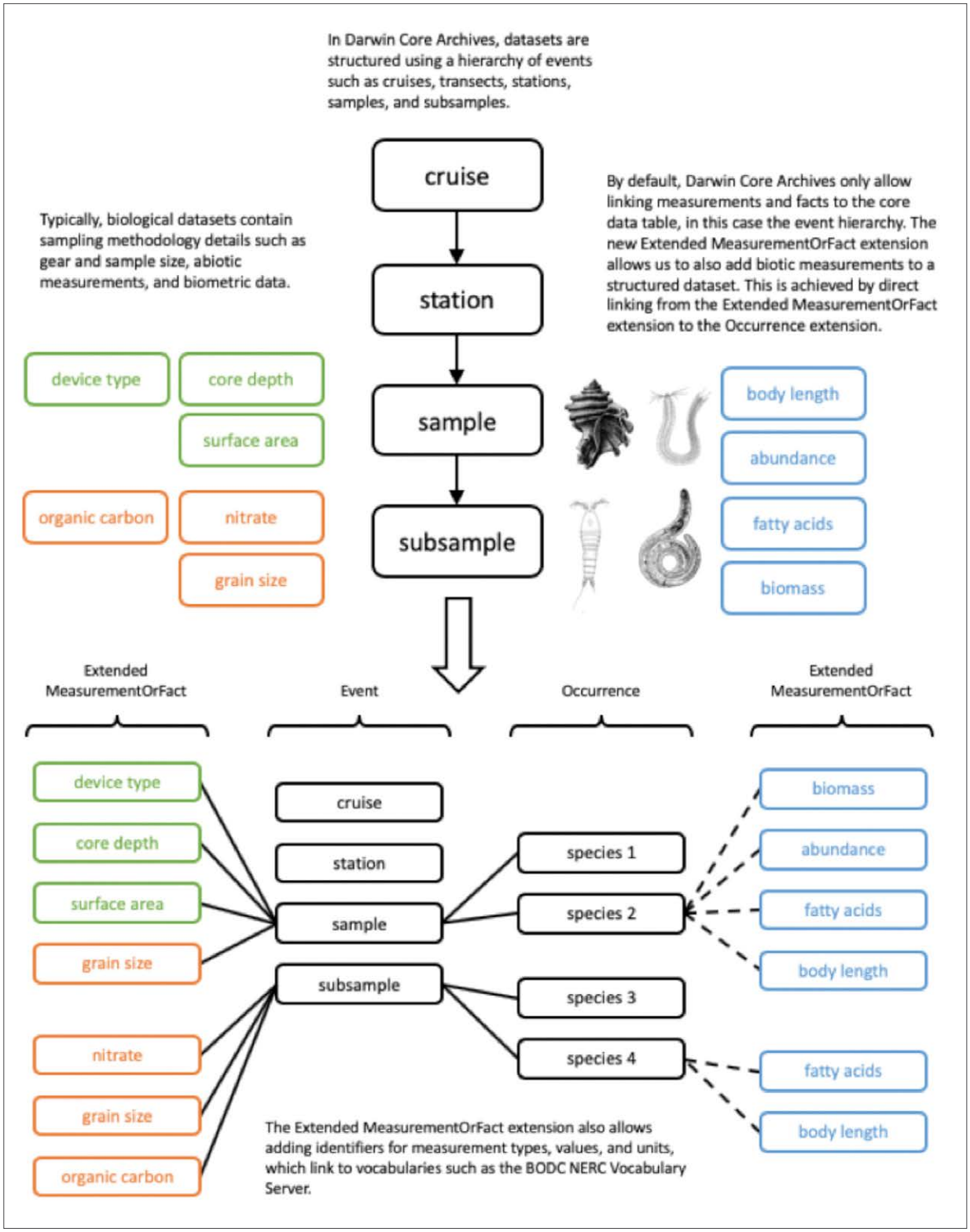

Figure 2. The OBIS-ENV-DATA standard allows the documentation of qualitative and quantitative information about both the sampling event and the species observation within that sampling event.