Session: **6802990** **7013886**

23/10/25 | 14:05 **Room:** Caldas

ORAL VIRTUAL • EN

A model for the provenance of marine biological samples and data

Presenter
Katrina Exter

Authors
Katrina Exter (Flanders Marine Institute (VLIZ) • Belgium)
Laurian Van Maldeghem (Flanders Marine Institute (VLIZ) • Belgium)
Christina Pavloudi (EMBRC • Belgium)
Ioulia Santi (EMBRC • Greece)
Cymon Cox (CCMAR • Portugal)

Abstract
We present our conceptual model for documenting provenance information for *biological material, observations, and their related and derived data*, with a focus on the marine biology domain. This model describes the provenance information necessary to describe the who, what, where, when, and how of collected/created/used biological material, the derived data, and observations. The model covers the activities carried out during typical marine biological field and experimental work, and is easily extendable to accommodate specialised areas of research (e.g. bioinformatics, AI-based analyses, complex imaging, etc.).

By providing this provenance information, one is providing: trustability; transparency and traceability; and reproducibility and reusability, thus ensuring FAIRness. Unfortunately, provenance metadata is not often provided at the level necessary to fully provide for "trust, trace, and reuse"; this can be due to lack of knowledge or neglect by individual scientists or due to the lack of standards' implementation by data repositories.

EMBRC has constructed this model, its associated schemas, and examples in plain text, EML, and JSON-LD, to make it easier for data producers to create provenance metadata for the data they publish and share. This work started in the Horizon project EOSC-Life, and is completing its journey in the Horizon project MARCO-BOLO.

We will present the conceptual model and its schemata, and explain how this provides the necessary provenance. We will demonstrate this in action by going through some example experiments and their datasets, and showing the provenance metadata that should accompany the resulting data. These examples and associated templates, can be used as resources by other scientists when creating their own provenance descriptions.