

Aligning Standards Communities: Sustainable Darwin Core MlxS Interoperability

Raïssa Meyer[‡], Pier Luigi Buttigieg[§], John Wiczorek^l, Thomas Stjernaard Jeppesen[¶], William D. Duncan[#], Yi-Ming Gan[□], Maxime Sweetlove[□], Saara Suominen[«], Task Group Sustainable Darwin Core MlxS Interoperability[»]

[‡] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremen, Germany

[§] Helmholtz Metadata Collaboration, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

^l Museum of Vertebrate Zoology, University of California, Berkeley, United States of America

[¶] Global Biodiversity Information Facility, Secretariat, Copenhagen, Denmark

[#] Lawrence Berkeley National Laboratory, Berkeley, United States of America

[□] Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[«] Ocean Biodiversity Information System (OBIS), Intergovernmental Oceanographic Commission of UNESCO, IOC Project Office for IODE, Oostende, Belgium

[»] TDWG Genomic Biodiversity Interest Group (GBWG), San Francisco, United States of America

Corresponding author: Raïssa Meyer (raissa.meyer@awi.de)

Abstract

Biodiversity is increasingly being assessed using omic technologies (e.g. metagenomics or metatranscriptomics); however, the metadata generated by omic investigations is not fully harmonised with that of the broader biodiversity community.

There are two major communities developing metadata standards specifications relevant to omic biodiversity data: [TDWG](#), through its Darwin Core ([DwC](#)) standard, and the Genomic Standard Consortium ([GSC](#)), through its Minimum Information about any (x) Sequence ([MlxS](#)) checklists. To prevent these specifications leading to silos between the communities using them (e.g. [INSDC](#): an internationally mandated database collaboration for nucleotide sequencing data [from health, biodiversity, microbiology, etc.] using the MlxS checklists; [OBIS](#) and [GBIF](#): global biodiversity data networks using the DwC standard), there is a need to harmonise them at the level of the standards organisations themselves.

To this end, we have brought together representatives from these standardisation bodies, along with representatives from established biodiversity data infrastructures, domain experts, data generators, and publishers to develop sustainable interoperability between the two specifications. Together, we have:

1. generated a [semantic mapping](#) between the terminology used in each specification, and [syntactic mapping](#) of their associated values following the Simple Standard for Sharing Ontology Mappings ([SSSOM](#)), and
2. created an example [MlxS-DwC extension](#) showing the incorporation of unmapped MlxS terms into a DwC-Archive.

To sustain these mechanisms of interoperability, we have proposed a Memorandum of Understanding between the GSC and TDWG. During our work, we also noted a number of key challenges that currently preclude interoperation between these two specifications.

In this talk, we will outline the major steps we took to get here, as well as the future activities we recommend based on our outputs.

Keywords

metadata standards, omics, biodiversity, eDNA, semantics, data policy

Presenting author

Raïssa Meyer

Presented at

TDWG 2021

Acknowledgements

We would like to thank all members of the TDWG/GBWG [Sustainable DarwinCore MlxS Interoperability Task Group](#) for their contributions to the work presented here.

Conflicts of interest