Linnaeus meets eDNA: crosslinking open data between metagenomics infrastructures and GBIF.org to address spatial and taxonomic bias and advance understanding of biodiversity

Kyle Copas (1), Dmitry Schigel (1), Thomas Stjernegaard Jeppesen (1), Urmas Kõljalg (2), and R. Henrik Nilsson (3)

(1) GBIF Secretariat, Copenhagen, Denmark (kcopas@gbif.org), (2) University of Tartu Natural History Museum, (3) University of Gothenburg, Gothenburg, Sweden

Together, fungi, bacteria, archaea, protists, viruses and phyto- and zooplankton are to the Earth's biosphere as dark matter is to the galaxy—a pervasive, ubiquitous and undifferentiated mass that is little observed and less understood. The fact that untold myriads of these organisms lack of formal scientific names remains one reason that they are underrepresented in global biodiversity observation and monitoring data. Without taxonomic identities, the elements of cryptic biodiversity are often overlooked or ignored in important scientific and policy contexts.

GBIF—the Global Biodiversity Information Facility—coordinates a network comprised of dozens of countries and hundreds of institutions worldwide to produce GBIF.org. This free, FAIR and open online resource brings together more than 1 billion species occurrence records, which provide evidence about where and when life forms occur on Earth.

In 2018, GBIF started working with researchers from University of Tartu Natural History Museum and the University of Gothenburg responsible for UNITE, a community-curated database for molecular identification of fungi. This collaboration piloted how to add 'species hypotheses'—the sequence-based Operational Taxonomic Units (OTUs) defined in UNITE—and georeferenced eDNA occurrences to the traditional Linnaean-based records available through GBIF.org. It also led to the development of a new prototype tool, Sequence ID, which enables users to match fungal DNA sequences to UNITE OTUs and prepare sequence-based occurrence records for publication on GBIF.org.

GBIF is now working with European Bioinformatics Institute (EMBL-EBI) and other partners to building on these efforts and increase the metagenomic data available through GBIF.org, both georeferenced occurrence data and molecular reference libraries (e.g. Silva, GLOMICON). In this session, we will provide an update on how this work bridges between the molecular and morphological biodiversity data domains to address major spatial, temporal and taxonomic gaps and biases in biodiversity data and offer researchers new perspectives on the seen and unseen patterns of biodiversity.