

Conference Abstract

# Mapping and Publishing Sequence-Derived Data through Biodiversity Data Platforms

Dmitry Schigel<sup>‡</sup>, Anders F. Andersson<sup>§</sup>, Andrew Bissett<sup>‡</sup>, Anders G Finstad<sup>¶</sup>, Frode Fossøy<sup>#</sup>, Marie Grosjean<sup>□</sup>, Michael Hope<sup>«</sup>, Urmaz Kõljalg<sup>»</sup>, Daniel Lundin<sup>^</sup>, R. Henrik Nilsson<sup>^, v</sup>, Maria Prager<sup>‡</sup>, Thomas Stjernegaard Jeppesen<sup>□</sup>, Cecilie S Svenningsen<sup>?, †</sup>

<sup>‡</sup> Global Biodiversity Information Facility - Secretariat, Copenhagen Ø, Denmark

<sup>§</sup> KTH Royal Institute of Technology, Stockholm, Sweden

| CSIRO, Hobart, Australia

¶ NTNU, Trondheim, Norway

# NINA, Trondheim, Norway

□ Global Biodiversity Information Facility, Secretariat, Copenhagen, Denmark

« CSIRO, Canberra, Australia

» University of Tartu, Tartu, Estonia

^ University of Gothenburg, Göteborg, Sweden

v Gothenburg Global Biodiversity Centre, Gothenburg, Sweden

‡ Stockholm University, Stockholm, Sweden

? University of Copenhagen, Copenhagen, Denmark

† Natural History Museum of Denmark, Copenhagen, Denmark

Corresponding author: Dmitry Schigel ([dschigel@gbif.org](mailto:dschigel@gbif.org))

Received: 01 Oct 2020 | Published: 09 Oct 2020

Citation: Schigel D, Andersson AF, Bissett A, Finstad AG, Fossøy F, Grosjean M, Hope M, Kõljalg U, Lundin D, Nilsson RH, Prager M, Jeppesen TS, Svenningsen CS (2020) Mapping and Publishing Sequence-Derived Data through Biodiversity Data Platforms. Biodiversity Information Science and Standards 4: e59212.

<https://doi.org/10.3897/biss.4.59212>

## Abstract

Most users will foresee the use of genetic sequences in the context of molecular ecology or phylogenetic research, however, a sequence with coordinates and a timestamp is a valuable biodiversity occurrence that is useful in a much broader context than its original purpose. To uncover this potential, sequence-derived data need to become findable, accessible, interoperable, and reusable through generalist biodiversity data platforms. Stimulated by the [Biodiversity Next](#) discussions in 2019, we have worked for about 10 months to put together practical data mapping and data publishing experiences in Norway, Australia, Sweden, and Denmark, as well as in the [UNITE](#) and the [GBIF](#) (Global

Biodiversity Information Facility) networks. The [resulting guide](#) was put together to provide practical instruction for mapping sequence-derived data.

Biodiversity data communities remain dominated by the macroscopic, easily detectable, morphologically identifiable species. This is not only true for citizen science and other forms of biodiversity popularization, but is also visible in the university and museum department structures, financial resource allocations, biodiversity legislation, and policy design. Recent decades of molecular advances have increased the power of genetic methods for detecting, describing, and documenting global biodiversity. We have yet to see the wide shift of data generating efforts from the traditional taxonomic foci of biodiversity assessments to the more balanced and inclusive systems focusing on all functionally important taxa and environments. These include soil, limnic and marine environments, decomposing plants and deadwood, and all life therein. Environmental DNA data enable recording of present and past presence of micro- and macroscopic organisms with minimal effort and by non-invasive methods. The apparent ease of these methods requires a cautious approach to the resulting data and their interpretation.

It remains important to define and agree on the organism recording and reporting routines for genetic data. DNA data represent a major addition to the many ways in which GBIF and other biodiversity data platforms index the living world. Our guide is resting on the shoulders of those who have been developing and improving [MixS](#) (Minimum Information about any (x) Sequence), [GGBN](#) (Global Genome Biodiversity Network) and other data standards. The added value of publishing sequence-derived data through non-genetic biodiversity discovery platforms relates to spatio-temporal occurrences and sequence-based names. Reporting sequence-derived occurrences in an open and reproducible way has a wide range of benefits: notably, it increases citability, highlights the taxa concerned in the context of biological conservation, and contributes to taxonomic and ecological knowledge.

## Keywords

genetic sequence, DNA, data standard, data mapping, barcoding, metabarcoding, eDNA

## Presenting author

Dmitry Schigel

## Presented at

TDWG 2020