



HAL
open science

Biodiversity Knowledge Graphs: Time to move up a gear!

Franck Michel, Antonia Ettorre, Catherine Faron, Julien Kaplan, Olivier Gargominy

► **To cite this version:**

Franck Michel, Antonia Ettorre, Catherine Faron, Julien Kaplan, Olivier Gargominy. Biodiversity Knowledge Graphs: Time to move up a gear!. TDWG 2021 annual virtual conference, Oct 2021, Virtual, France. pp.e73699, 10.3897/biss.5.73699 . hal-03373536

HAL Id: hal-03373536

<https://hal.science/hal-03373536>

Submitted on 11 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Conference Abstract

Biodiversity Knowledge Graphs: Time to move up a gear!

Franck Michel[‡], Antonia Ettore[‡], Catherine Faron[‡], Julien Kaplan[‡], Olivier Gargominy[§]

[‡] University Cote d'Azur, CNRS, Inria, Sophia Antipolis, France

[§] Muséum national d'Histoire naturelle, Paris, France

Corresponding author: Franck Michel (franck.michel@inria.fr)

Received: 30 Aug 2021 | Published: 31 Aug 2021

Citation: Michel F, Ettore A, Faron C, Kaplan J, Gargominy O (2021) Biodiversity Knowledge Graphs: Time to move up a gear! Biodiversity Information Science and Standards 5: e73699. <https://doi.org/10.3897/biss.5.73699>

Abstract

Harnessing worldwide biodiversity data requires integrating myriad pieces of information, often sparse and incomplete, into a global, coherent data space. To do so, projects like the [Global Biodiversity Information Facility](#), [Catalog of Life](#) and [Encyclopedia of Life](#) have set up platforms that gather, consolidate, and centralize billions of records from multiple data sources. This approach lowers the entry barrier for scientists willing to consume aggregated biodiversity data but tends to build silos that hamper cross-platform interoperability.

The Web of Data embodies a different approach underpinned by the Linked Open Data (LOD) principles (Heath and Bizer 2011). These principles bring about the building of a large, distributed, cross-domain knowledge graph (KG), wherein data description relies on vocabularies with shared, formal, machine-processable semantics. So far however, little biodiversity data have been published this way. Early efforts focused primarily on taxonomic registers, such as [NCBI](#), [VTO](#) and [AGROVOC](#). More recent efforts have started paving the way for the publication of more diverse biodiversity KGs (Page 2019, Penev et al. 2019, Michel et al. 2017).

Today, we believe that it is time for more biodiversity data producers to join in and start publishing connected KGs spanning a much broader set of domains, far beyond just taxonomic registers. In this talk, we wish to present an on-going endeavor in line with this vision. In a previous work, we published TAXREF-LD (Michel et al. 2017), a LOD

representation of the French taxonomic register developed and maintained by the French National Museum of Natural History. We modeled nomenclatural information as a thesaurus of scientific names, taxonomic information as an ontology of classes denoting taxa, and additional information such as ranks and vernacular names. Recently, we have extended the scope of TAXREF-LD to represent and interlink data as various as geographic locations, species interactions, development stages, trophic levels, as well as conservation, biogeographic, and legal status (regulations, protections, etc.).

We put a specific effort into working out a model that accurately accounts for the semantics of the data while respecting knowledge engineering practices. For instance, a common design shortcoming is to attach all information as properties of a taxon. This is a rightful choice for some properties like a scientific name or conservation status, but properties that actually pertain to biological individuals themselves, e.g. habitat and trophic level, should better be attached to class members. With the presentation of this work, we wish to advance the discussion about integration scenarios based on knowledge graphs with the different biodiversity data stakeholders.

Keywords

linked data, modeling, modelling

Presenting author

Franck Michel

Presented at

TDWG 2021

References

- Heath T, Bizer C (2011) Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology 1 (1): 1-136. <https://doi.org/10.2200/s00334ed1v01y201102wbe001>
- Michel F, Gargominy O, Terceire S, Faron-Zucker C (2017) A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF. 1933. Proceedings of the ISWC2017 workshop on Semantics for Biodiversity (S4BioDiv).
- Page R (2019) Ozymandias: a biodiversity knowledge graph. PeerJ 7 <https://doi.org/10.7717/peerj.6739>
- Penev L, Dimitrova M, Senderov V, Zhelezov G, Georgiev T, Stoev P, Simov K (2019) OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science. Publications 7 (38). <https://doi.org/10.3390/publications7020038>