Connecting the Dots: Joint development of best practices between infrastructures in support of bidirectional data linking

Wouter Addink[‡,§], Niki Kyriakopoulou[§]

‡ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
§ Naturalis Biodiversity Center, Leiden, Netherlands

Corresponding author: Wouter Addink (wouter.addink@naturalis.nl)

Abstract

Working together is key in terms of knowledge exchange and in the Biodiversity Community Integrated Knowledge Library project (BiCIKL), infrastructures involved in the biodiversity data landscape are working together to connect data from their different but related data domains. This will be key to connecting the dots towards the development of global collections and global specimen networks in line with e.g., the Extended Specimens Network (Lendemer et al. 2019), which describes a strategy to enhance US biodiversity collections and promote research and education by enabling seamless data integration, attribution and use tracking. It will enable and improve linkages in the fragmented landscape of literature, specimens, collections, species and sequence data. It is required to reliably connect collection level data and specimen level data with derived and related data served through these infrastructures.

BiCIKL is a project funded by the European Commission that will develop bidirectional data linkages through persistant identifiers between the data so that they become part of a big FAIR (Findable, Accessible, Interoperable, Reusable) data pool available to researchers, public authorities and business to foster innovation in the digital economy and to contribute in solving global environmental challenges. This will enable end-users to seamlessly access a larger pool of enriched data and knowledge, thereby bypassing the traditional barrier of having to search multiple databases and then link up different data types, or to manually extract data from literature.

Two elements are key for development of the data linkages: 1) the global adoption of agreed upon persistent identifiers for digital specimen objects, taxonomic treatments, collections, persons, sequences and taxa across stakeholders, and 2) best practices for the generation, management and curation of the linked data. We will present the best practices that have been jointly developed by the infrastructures in the BiCIKL project for findability, re-use and accessibility. The presentation will focus on the best practices that are essential for reliably linking specimen- and collection-level data with other types of

data. Implementing these practices will be an important contribution towards connecting the dots in the biodiversity data landscape and will enrich the information at the data provider level.

Keywords

specimens, global specimens network, BiCIKL, persistent identifiers, linking data

Presenting author

Wouter Addink

Presented at

TDWG 2022

Acknowledgements

Funding program

Conflicts of interest

References

- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2019) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. Bioscience 70 (1): 23-30. https://doi.org/10.1093/biosci/biz140