PROCEEDINGS OF TDWG
Biodiversity Information Science and Standards 1: e17010
doi: 10.3897/tdwgproceedings.1.17010
OPEN ACCESS

Conference Abstract

Associating Occurrences with Genes, Phenotypes, and Environments through the Distributed System of Scientific Collections (DiSSCo)

Dimitrios Koureas[‡], Wouter Addink[§]

‡ Natural History Museum, London, United Kingdom
§ Naturalis Biodiversity Center, Leiden, Netherlands

Corresponding author: Dimitrios Koureas (d.koureas@nhm.ac.uk)

Received: 20 Jul 2017 | Published: 25 Jul 2017

Citation: Koureas D, Addink W (2017) Associating Occurrences with Genes, Phenotypes, and Environments through the Distributed System of Scientific Collections (DiSSCo). Proceedings of TDWG 1: e17010. https://doi.org/10.3897/tdwgproceedings.1.17010

Abstract

Over the last few decades, the research practice in natural sciences has changed dramatically. Remote sensing, rapid identification and molecular approaches allow us to efficiently monitor the changing world around us and understand the cause of those changes. Advances of digital, genomic and information technologies enable natural science collections to provide novel discoveries and ask for new collection types and attributes, while fostering the development of innovative approaches to face the urgent societal challenges.

Natural Science Collections (three billion specimens globally) represent an unparalleled scientific asset. They constitute a unique source of diverse data classes, including genomic, chemical, morphological and geo-spatial information. Despite existing successful examples of infrastructures, aggregating and publishing specific data classes (such as the Global Biodiversity Information Facility, GenBank or the Encyclopedia of Life - TraitBank), the landscape remains fragmented with limited capacity to bring together this information in a systematic and robust manner.

The Distributed System of Scientific Collections (DiSSCo) represents a pan-European initiative, and the largest ever agreement of natural science museums, to jointly address the fragmentation of European collections. DiSSCo is set to unify European natural science collections into a coherent new research infrastructure, able to provide bio- and geo-diversity data at the scale, form and precision required by a multi-disciplinary user base. At the heart of the technical implementation of DiSSCo, is the development of a cloud-based non-relational data store that links occurrence, genomic, chemical and trait data classes, by robustly and unambiguously anchoring each data object back to the physical object. By harmonising digitisation, curation and publication processes and workflows, across all its nodes, DiSSCo can populate and serve a knowledge graph for European natural science collections.

In this paper we will introduce the vision, mission and objectives of DiSSCo, discuss the technical approach and touch upon the socio-cultural and governance aspects supporting this large-scale European endeavour.

DiSSCo is applying for inclusion in the 2018 European roadmap for research infrastructures, through an evaluation process organised by the European Strategy Forum on Research Infrastructures (ESFRI). This process is politically and/or financially supported by 12 European countries and an expanding network of 95 natural science museums in 20 countries.

Keywords

Research Infrastructures, Natural Science Collections, knowledge graph, ESFRI, Linked Data

Presenting author

Dimitrios Koureas