A Nano(publication) Approach Towards Big Data in Biodiversity

Mariya Dimitrova[‡,§], Teodor Georgiev[§], Lyubomir Penev[§,|]

‡ Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria
§ Pensoft Publishers, Sofia, Bulgaria
| Institute of Biodiversity & Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

Abstract

One of the major challenges in biodiversity informatics is the generation of machine-readable data that is interoperable between different biodiversity-related data infrastructures. Producers of such data have to comply with existing standards and to be resourceful enough to enable efficient data generation, management and availability. Conversely, nanopublications offer a decentralised approach (Kuhn et al. 2016) towards achieving data interoperability in a robust and standarized way. A nanopublication is a named RDF graph, which serves to communicate a single fact and its original source (provenance) through the use of identifiers and linked data (Groth et al. 2010). It is composed of three constituent graphs (assertion, provenance, and publication info), which are linked to one another in the nanopublication header (Kuhn et al. 2016). For instance, a nanopublication has been published to assert a species interaction in which a hairy woodpecker (*Picoides villosus*) ate a beetle (genus *Ips*), along with the license and related bibliographic citation[*1]. In biodiversity, nanopublications can be used to exchange information between infrastructures in a standardised way (Fig. 1) and to enable curation and correction of knowledge. They can be implemented within different workflows to formalise biodiversity knowledge in self-enclosed graphs. We have developed several nanopublication models[*2] for different biodiversity use cases: species occurrences, new species descriptions, biotic interactions, and links between taxonomy, sequences and institutions. Nanopublications can be generated by various means:
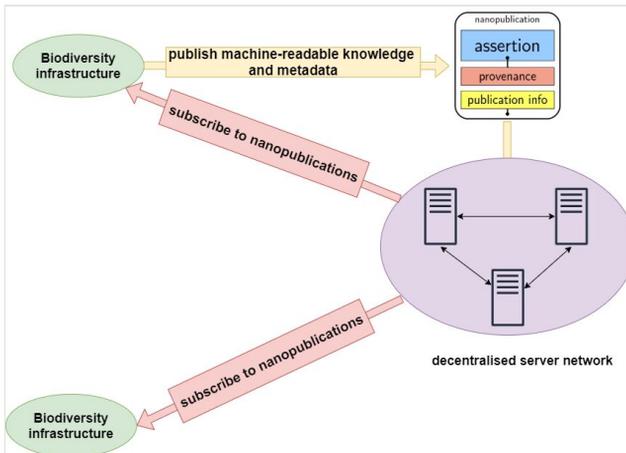
Figure 1.

Biodiversity infrastructures can communicate via nanopublications by publishing to and reading from the decentralised nanopublication server network.

1. semi-automatic extraction from the published literature with a consequent human curation and publication;
2. generation during the publication process by the authors via dedicated formalisation tool and published together with the article;
3. *de novo* generation of a nanopublication through decentralised networks such as Nanobench (Kuhn et al. 2021).

One of the possible uses of nanopublications in biodiversity is communicating new information in a standardised way so that it can be accessed and interpreted by multiple infrastructures that have a common agreement on how information is expressed through the use of particular ontologies, vocabularies and sets of identifiers. In addition, we envision nanopublications to be useful for curation or peer-review of published knowledge by enabling any researcher to publish a nanopublication containing a comment of an assertion made in a previously published nanopublication. With this talk, we aim to showcase several nanopublication formats for biodiversity and to discuss the possible applications of nanopublications in the biodiversity domain.

Keywords

nanopublications, RDF, provenance, decentralisation

Presenting author

Mariya Dimitrova

Presented at

TDWG 2021

Funding program

References

- Groth P, Gibson A, Velterop J (2010) The Anatomy of a Nanopublication. Information Services & Use 30 (1-2). https://doi.org/10.3233/ISU-2010-0613
- Kuhn T, Chichester C, Krauthammer M, Queralt-Rosinach N, Verborgh R, Giannakopoulos G, Ngonga Ngomo A, Viglianti R, Dumontier M (2016) Decentralized provenance-aware publishing with nanopublications. PeerJ Computer Science 2 https://doi.org/10.7717/peerj-cs.78
- Kuhn T, Taelman R, Emonet V, Antonatos H, Soiland-Reyes S, Dumontier M (2021) Semantic micro-contributions with decentralized nanopublication services. PeerJ Computer Science 7 https://doi.org/10.7717/peerj-cs.387

Endnotes

[*1]   http://purl.org/np/RAzquSkwsTAZm61nReG6MOjXEXUx8fNVfdWnAzyn6sOhU
[*2]   https://github.com/mdmtrv/nanopub-examples