

## Conference Abstract

# Strategies for Assembling the Biodiversity Knowledge Graph

Roderic Page ‡

‡ University of Glasgow, Glasgow, United Kingdom

Corresponding author: Roderic Page ([roderic.page@glasgow.ac.uk](mailto:roderic.page@glasgow.ac.uk))

Received: 29 Sep 2020 | Published: 30 Sep 2020

Citation: Page R (2020) Strategies for Assembling the Biodiversity Knowledge Graph. Biodiversity Information Science and Standards 4: e59126. <https://doi.org/10.3897/biss.4.59126>

## Abstract

This talk explores different strategies for assembling the “biodiversity knowledge graph” (Page 2016). The first is a centralised, crowd-sourced approach using Wikidata as the foundation. Wikidata is becoming increasingly attractive as a knowledge graph for the life sciences (Waagmeester et al. 2020), and I will discuss some of its strengths and limitations, particularly as a source of bibliographic and taxonomic information. For example, Wikidata’s handling of taxonomy is somewhat problematic given the lack of clear separation of taxa and their names.

A second approach is to build biodiversity knowledge graphs from scratch, such as [OpenBioDiv](#) (Penev et al. 2019) and my own [Ozymandias](#) (Page 2019). These approaches use either generalised vocabularies such as [schema.org](#), or domain specific ones such as [TaxPub](#) (Catapano 2010) and the Semantic Publishing and Referencing Ontologies ([SPAR](#)) (Peroni and Shotton 2018), and to date tend to have restricted focus, whether geographic (e.g., Australian animals in [Ozymandias](#)) or temporal (recent taxonomic literature, [OpenBioDiv](#)). A growing number of data sources are now using [schema.org](#) to describe their data, including [ORCID](#) and [Zenodo](#), and efforts to extend [schema.org](#) into biology ([Bioschemas](#)) suggest we may soon be able to build comprehensive knowledge graphs using just [schema.org](#) and its derivatives.

A third approach is not to build an entire knowledge graph, but instead focus on constructing small pieces of the graph tightly linked to supporting evidence, for example via annotations. Annotations are increasingly used to mark up both the biomedical literature

(e.g., Kim et al. 2015, Venkatesan et al. 2017) and the biodiversity literature (Batista-Navarro et al. 2017). One could argue that taxonomic databases are essentially lists of annotations (“this name appears in this publication on this page”), which suggests we could link literature projects such as the Biodiversity Heritage Library ([BHL](#)) to taxonomic databases via annotations. Given that the [International Image Interoperability Framework](#) (IIIF) provides a framework for treating publications themselves as a set of annotations (e.g., page images) upon which other annotations can be added (Zundert 2018), this suggests ways that knowledge graphs could lead directly to visualising the links between taxonomy and the taxonomic literature.

All three approaches will be discussed, accompanied by working examples.

## Keywords

Biodiversity Heritage Library, OpenBioDiv, Ozymandias, schema.org, annotations, biodiversity literature, taxonomic literature, Wikidata

## Presenting author

Roderic Page

## Presented at

TDWG 2020

## References

- Batista-Navarro R, Zerva C, Nguyen NH, Ananiadou S (2017) A Text Mining-Based Framework for Constructing an RDF-Compliant Biodiversity Knowledge Repository. Information Management and Big Data. [ISBN 978-3-319-55209-5]. [https://doi.org/10.1007/978-3-319-55209-5\\_3](https://doi.org/10.1007/978-3-319-55209-5_3)
- Catapano T (2010) TaxPub: An Extension of the NLM/NCBI Journal Publishing DTD for Taxonomic Descriptions. National Center for Biotechnology Information (US) URL: <https://www.ncbi.nlm.nih.gov/books/NBK47081/>
- Kim J, Cohen KB, Kim J (2015) PubAnnotation-query: a search tool for corpora with multi-layers of annotation. BMC Proceedings 9 (Suppl 5). <https://doi.org/10.1186/1753-6561-9-S5-A3>
- Page RDM (2016) Towards a biodiversity knowledge graph. Research Ideas and Outcomes 2 <https://doi.org/10.3897/rio.2.e8767>
- Page RM (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 (2). <https://doi.org/10.3390/publications7020038>
- Peroni S, Shotton D (2018) The SPAR Ontologies. *The Semantic Web – ISWC 2018*. [ISBN 978-3-030-00668-6]. [https://doi.org/10.1007/978-3-030-00668-6\\_8](https://doi.org/10.1007/978-3-030-00668-6_8)
- Venkatesan A, Kim J, Talo F, Ide-Smith M, Gobeill J, Carter J, Batista-Navarro R, Ananiadou S, Ruch P, McEntyre J (2017) SciLite: a platform for displaying text-mined annotations as a means to link research articles with biological data. *Wellcome Open Research* 1 <https://doi.org/10.12688/wellcomeopenres.10210.2>
- Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good B, Griffith M, Griffith O, Hanspers K, Hermjakob H, Hudson T, Hybiske K, Keating S, Manske M, Mayers M, Mietchen D, Mitraka E, Pico A, Putman T, Riutta A, Queralt-Rosinach N, Schriml L, Shafee T, Slenter D, Stephan R, Thornton K, Tsueng G, Tu R, Ul-Hasan S, Willighagen E, Wu C, Su A (2020) Science Forum: Wikidata as a knowledge graph for the life sciences. *eLife* <https://doi.org/10.7554/eLife.52614>
- Zundert Jv (2018) On Not Writing a Review about Mirador: Mirador, IIIF, and the Epistemological Gains of Distributed Digital Scholarly Resources. *Digital Medievalist* 11 (1). <https://doi.org/10.16995/dm.78>