

# The MIAPA ontology: An annotation ontology for validating minimum metadata reporting for phylogenetic analyses

Hilmar Lapp(1), Joachim Baran(2), Enrico Pontelli(3), Arlin Stoltzfus(4), Ramona Walls(5)

(1) National Evolutionary Synthesis Center (NESCent), Durham, NC, USA

(2) Ontario Institute for Cancer Research, Toronto, ON, Canada

(3) Department of Computer Science, New Mexico State University, Las Cruces, NM, USA

(4) Biosystems and Biomaterials Division, National Institute of Standards and Technology, and Institute for Bioscience and Biotechnology Research, Rockville, MD, USA

(5) The iPlant Collaborative, University of Arizona, Tucson, AZ, USA

Phylogenetic trees are published at a rapidly increasing rate, with ever wider taxonomic coverage and deeper taxonomic sampling. Yet, reusing this knowledge for a myriad of biological fields is difficult due to a number of challenges, including effective online sharing, discovery, and machine-based consumption of phylogenies. Aside from the problem of digital access, a critical part of deciding to reuse a published phylogeny instead of recreating it is evaluating a phylogenetic tree's fitness for purpose. Such an evaluation requires certain metadata, such as the type of alignment from which the tree was constructed, the methods used to infer alignment and tree, whether branch lengths are present, and whether tip labels were reconciled with a taxonomy or identifier system. Standardizing these metadata through community consensus motivated the creation of the Minimum Information About a Phylogenetic Analysis (MIAPA) metadata reporting standard, first articulated in 2006 by Leebens-Mack et al. However, a formal codification of which metadata are relevant, which are required or recommended, and how their values are to be provided is lacking, yet is required by tools that enable users to reliably extract MIAPA-relevant information from a phylogenetic tree under consideration. To address this gap, we created the MIAPA ontology, an application ontology designed for annotating phylogenetic data according to the recommendations made by an informal draft consensus checklist created in 2011. The ontology is developed in OWL, and reuses concepts from several existing ontologies, including the Comparative Data Analysis Ontology (CDAO), the Information Artifact Ontology (IAO), the Software Ontology (SWO), and the W3C Provenance Ontology (PROV). We demonstrate use of the ontology for data annotation and integration using previously published trees converted from Newick or NEXUS format to a Resource Description Framework (RDF) representation. Developing the ontology and applying it revealed some of the challenges in translating informal, sometimes ambiguous checklist recommendations into a formal standard. We also consider the potential of formalizing a minimum metadata reporting standard as an ontology for enabling data sharing platforms to assess, rate, and thereby incentivize metadata quality and richness of their data contents for the benefit of their users.

Availability: <http://github.com/miapa/miapa> under Creative Commons Zero public domain waiver.

**Acknowledgements:** The MIAPA ontology is a product of the Hackathons, Interoperability, Phylogenies (HIP) Working Group supported by the National Evolutionary Synthesis Center (NESCent; NSF #EF-0905606).