





SALON ontology for the formal description of Sequence Alignments

Antonio Benítez-Hidalgo^{1,2,3} , José F. Aldana-Montes^{1,2,3} , Ismael Navas-Delgado^{1,2,3} , and María del Mar Roldán-García^{1,2,3} 

¹ Departamento de Lenguajes y Ciencias de la Computación, University of Málaga, Málaga, Spain

antobio.b@uma.es

² University of Málaga, ITIS Software, Ada Byron Research Building, Málaga, Spain

³ Instituto de Investigación Biomédica de Málaga – IBIMA, Málaga, Spain

Keywords: Sequence alignment, Ontology, Linked data, Data integration, Semantic web

Published in: BMC Bioinformatics 24, 69, 2023

Impact Factor: JCR 3.327 - Q2 - Position: 20/57 - Area: MATHEMATICAL & COMPUTATIONAL BIOLOGY

DOI: <https://doi.org/10.1186/s12859-023-05190-7>

Abstract. *Background.* Information provided by high-throughput sequencing platforms allows the collection of content-rich data about biological sequences and their context. Sequence alignment is a bioinformatics approach to identifying regions of similarity in DNA, RNA, or protein sequences. However, there is no consensus about the specific common terminology and representation for sequence alignments. Thus, automatically linking the wide existing knowledge about the sequences with the alignments is challenging. *Results.* The Sequence Alignment Ontology (SALON) defines a helpful vocabulary for representing and semantically annotating pairwise and multiple sequence alignments. SALON is an OWL 2 ontology that supports automated reasoning for alignments validation and retrieving complementary information from public databases under the Open Linked Data approach. This will reduce the effort needed by scientists to interpret the sequence alignment results. *Conclusions.* SALON defines a full range of controlled terminology in the domain of sequence alignments. It can be used as a mediated schema to integrate data from different sources and validate acquired knowledge.

