

NGSOnto – keeping track of the NGS pipeline process using an ontological approach and its application on molecular epidemiology

Mickael Silva, Catia Vaz, Alexandre Francisco, Mário Ramirez, Francisco M Couto, João André Carriço

Nowadays, microbial typing methods are fundamental to clinical microbiology, being tools for outbreak detection and tracking, and for providing valuable information in the study of the evolution of virulence factors such as antimicrobial resistance mechanisms. With the advances in sequencing technologies, namely in the Whole Genome Sequencing techniques, it has become imperative to connect the information present in the former sequence based microbial typing methods and the new high sequence throughput data. Another critical issue is the lack of interoperability and querying facilities through web services, on the currently available microbial typing databases, which are spread out through several institutes worldwide. Considering these issues and the previous development of a microbial typing identification ontology (TypOn), a Next Generation Sequencing (NGS) ontology was developed. This ontology was built in order to describe a NGS pipeline process, being interoperable with the TypOn ontology as well as other ontologies. The Web Ontology Language (OWL) was used to build the ontology, following a Basic Formal Ontology 2.0 structure and using RDF format data to save information. The data mapped on the ontology may be later queried using a SPARQL endpoint made available or through the RESTful webservice that is under development. The NGSOnto, aims at capturing the workflow of all the processes involved, in order to ensure the reproducibility of the entire process, focusing particularly in the data analysis steps, downstream of the raw sequence data. In this way it will allow the validation of the use of NGS as bona fide source of data for microbial typing, by providing the basis to determine if the two results are comparable.