Towards Knowledge Graphs of Research Software metadata

Daniel Garijo[1], Yolanda Gil[1], Mr. Maximiliano Osorio[1], Varun Ratnakar[1], and Deborah Khider[1]

1 Information Sciences Institute, USC

Research software is a key asset for understanding, reusing and reproducing results in computational sciences. An increasing amount of software is stored in code repositories, which usually contain human readable instructions indicating how to use it and set it up. However, developers and researchers often need to spend a significant amount of time to understand how to invoke a software component, prepare data in the required format, and use it in combination with other software. In addition, this time investment makes it challenging to discover and compare software with similar functionality. In this talk I will describe our efforts to address these issues by creating and using Open Knowledge Graphs that describe research software in a machine readable manner. Our work includes: 1) an ontology that extends schema.org and codemeta, designed to describe software and the specific data formats it uses; 2) an approach to publish software metadata as an open knowledge graph, linked to other Web of Data objects; and 3) a framework for automatically extracting metadata from software repositories; and 4) a framework to curate, query, explore and compare research software metadata in a collaborative manner. The talk will illustrate our approach with real-world examples, including a domain application for inspecting and discovering hydrology, agriculture, and economic software models; and the results of our framework when enriching the research software entries in Zenodo.org.