RDF Knowledge Graph in Astronomy

our practical experience

Volodymyr Savchenko

March 22, 2021

Context: In the last decade, it became possible to detect new kinds of signals from astrophysical objects: new telescopes yielded first unambiguous observations of cosmic high-energy neutrino and gravitation wave emission. Our team made key contributions to this domain. These efforts stressed particular methodological challenges: telescopes catching different kinds of signals emerged from traditionally diverse domains, complicating any efforts to combine observations. Furthermore, turns out that (at least for now) the most prolific multi-messenger sources are short-lived (transient), require immediate action, and rely heavily on automation.

This context stressed the need for quick, clear, and meaningful automated communication. Several initiatives undertook efforts to adopt and develop semantic annotations and metadata to tackle this challenge.

Methodology: To aggregate and organise the diverse multi-messenger and multi-discipline data and workflows we have gradually adopted a triple store Knowledge Graph. It combines inputs from various external sources, observational data incoming from operations of the INTEGRAL Science Data Centre, and integrates user contributions. It is accessible with SPARQL, but often experienced by users seamlessly through collection of tools and front-ends.

In addition to aggregating descriptions of data and scientific data analysis workflows, it stores descriptions of workflows which transform the KG itself, as well as produce various external actions (sending various events, suggesting new observations, etc).

Experience: Our adoption of a KG was driven by particular use case and as features were introduced they were routinely tested in production environment, which highlighted some of the challenges encountered, most related to performance, verbosity, and veracity of RDF store. I will explain how we overcome these challenges in our specific case (including trying to rely on some extensions, such as RDF-star, Atomic Data, merging a variety of databases, etc), and what seems to be difficult to address. I will be especially interested to hear feedback and advice on how to further advance this technology (if it appears suitable), and what challenges we may encounter in the future.