Catalog for ontologies towards semantic integration across environmental health science research data

Resham Kulkarni, Mike Conway, Deep Patel, Maria Shatz, Stephanie Holmgren, Charles Schmitt

Office of Data Science, National Institute for Environmental Sciences, Research Triangle Park, NC, USA

Abstract

Towards the aim to standardize metadata, controlled vocabularies and ontologies used by different data repositories and systems at the National Institute for Environmental Health Sciences (NIEHS), we recently implemented a metadata catalog "MetaCat". This catalog serves as a central repository for metadata terms, controlled vocabularies and ontologies, based on NIEHS use cases. The ontologies of interest span the biomedical, clinical and environmental domains. We are in the process of integrating MetaCat with NIEHS Data Commons.

Keywords:

Metadata Catalog, Metadata Standardization, Environmental Health Sciences

Introduction

To make the research data across the National Institute for Environmental Health Sciences (NIEHS) FAIR (Findable, Accessible, Interoperable, Reusable), we aim to standardize metadata, controlled vocabularies and ontologies used by the different data repositories and systems. Semantic integration enabled by this standardization will facilitate data management and downstream data integration. Principles developed through this effort can be extended to all environmental health science (EHS) data.

Results

We recently implemented a metadata catalog "MetaCat" to serve as a central repository for metadata terms, controlled vocabularies and ontologies, based on varied use cases at NIEHS. One of the use cases is to have standard metadata terms for NIEHS Data Commons (DC) data curation. A second use case from DC is to use standard terms for simple and advanced search e.g. be able to broaden the search based on the hierarchy of annotations within MetaCat. Several additional use cases range from using the ontologies to mine information from publications to making the ontologies available to the extramural EHS community. MetaCat enables these use cases by serving as a central resource, enabling compilation of a new ontologies from existing, building a new ontology from scratch, and mapping between ontologies. The ontologies of interest span the biomedical, clinical and environmental domains. We are developing a MetaCat Ontology to serve as a governance document to track the ontologies we are modelling.

Currently, we are in the process of integrating MetaCat with NIEHS DC. We are also evaluating approaches to use existing mapping or create mapping between ontology/metadata terms to further enhance the semantic integration.