# Introduction

RVO, Research Variable Ontology, proposes a schema enterprise can use to record their data analytics experiments and use as a knowledgebase for learning and recommendation support. RVO is designed around the research variables, which form the basis of the hypothesis that analysts test through building a model. RVO can answer an array of questions raised by data analysts when they start designing a solution and provide recommendations and alternatives. All facts or expert knowledge recorded in RVO are traceable to its origin (i.e. a person, publication, validated model). RVO follows best practices in ontology design and reuse existing data models and vocabularies (such as DBPedia1, RDF-Cube2, FABIO3) to facilitate efficient reuse in real world applications by using semantic technologies and open standards (RDF, OWL, SPARQL).

The purpose of RVO is to:

* Establish a common terminology for the organization to represent classes and properties in the empirical analytics domain, relating to existing taxonomies, ontologies and data standards.
* Capture established (even if contradictory) complex analytics knowledge in a particular domain and its origin. The knowledge may include known relationships between different variables, how a variable is linked to in a particular model, what are the relationships between variable and data sets, what are the relationships between variables and analytics models.
* Integrate existing ontologies to enrich the knowledge base in an organization when conducting . These ontologies can be domain-specific ontologies, ontologies representing concepts, people, standards and etc. This integration provides the knowledge base across an organization.
* Assist data analytics process stages such as variable selection, data source selection, dataset selection and evaluation.

<<<<<<<<<Include the Image>>>>>>>>>>>>

The components of the ontology are shown in figure 2. Main classes of the ontology are Variable, Measure, Model, LinkedVariables and Origin. LinkedVariables class captures a link between any two variables. LinkType is defined to represent details of different links between variables as an extendable catalogue of link types (Causal, InvestigativeCausal, NoLink). Origin class is defined to represent details about the origin or reference to any concept. Main origin types in RVO are: ViaModel, ViaLiterature, and FromExperts.

One strength of this ontology is how it can be linked with any domain specific ontology via the operationalized relationship. We can take any concept (thing or property) defined in any third-party ontology as rvo:Concept and link it to a variable. In this way, the context of the variable is readily available through a domain specific ontology.

|  |  |
| --- | --- |
| Key Concepts | Definition |
| Concept | Concept or Thing refer to abstraction of any entity which may exist in an ontology. |
| Variable | Variables exist as metrics to quantify concepts. |
| Measure | Measures are the metrics that represent actual values for a variables. |
| Variable Link | A link describes a connection between two variables, the nature of this connection and how this connection was established. |
| Origin | Origin is where a concept (e.g. variable, Variable Link, Measure) is first defined or mentioned. RVO has three origins – Expert Person, Reference of a publication, Analytics Model |
| Link Type | Link Type describes the type of link that exists between two or more variables. Example of Link Types are: ·Causal: a variable causes another variable · Investigative Causal: there is a hypothesis that one variable cause another, the hypothesis needs to be tested · No link: there are instances that no link can be established between two variables. |
| Data Source | Data source references to the source from which a data is coming from. |
| Dataset | Dataset refers to the data file provided by a data source which contain the data for one or multiple measures. We recommend publishing dataset following RDF-Cube structure. |
| Dataset  Structure | Dataset structure refers to the meta data for a dataset such as what measures are captured, unit of measure. |

1.

<https://wiki.dbpedia.org/>

2. <https://www.w3.org/TR/vocab-data-cube/>

3. https://sparontologies.github.io/fabio/current/fabio.html