RVO - The Research Variable Ontology

Madhushi Bandara(✉), Ali Behnaz, and Fethi A. Rabhi

University of New South Wales, Sydney, Australia
{k.bandara,a.behnaz,f.rabhi}@unsw.edu.au

Abstract. Enterprises today are presented with a plethora of data, tools and analytics techniques, but lack systems which help analysts to navigate these resources and identify best fitting solutions for their analytics problems. To support enterprise-level data analytics research, this paper presents Research Variable Ontology (RVO), an ontology designed to catalogue and explore essential data analytics design elements such as variables, analytics models and available data sources. RVO is specialised to support researchers with exploratory and predictive analytics problems, popularly practiced in economics and social science domains. We present the RVO design process, its schema, how it links and extends existing ontologies to provide a holistic view of analytics related knowledge and how data analysts at the enterprise level can use it. Capabilities of RVO are illustrated through a case study on House Price Prediction.

Keywords: Data analytics · Semantic modeling · Research variables

1 Introduction

Motivation: Designing correct analytics solutions which meet the respective business objectives is challenging [15]. It involves different decision making such as selecting suitable tools, algorithms, datasets and deciding how to generate results and report them accurately. A data analytics researcher in an organisation spends lots of his/her time understanding the domain, analytics problem at hand and the existing related knowledge. Many new analytics research projects start with a literature survey to investigate and find a suitable approach, explore available data, and run numerous trial-and-error experiments. The iterative process includes cleaning and pre-processing data, identifying suitable variables to input into the right model, and evaluating the output of the model. Yet the outcome of the project is limited to an analytics model, usually in the form of a spreadsheet or a software code, and all the experience and knowledge accumulated by the researchers are not recorded or made available for future use. As no data analytics solution fits all problems [11], a solution cannot constantly perform well over a long period without accommodating changes in data and business goals. In order to maintain the expected performance, designed analytics solutions require frequent interventions and modifications from researchers.

While there are numerous papers on designing data analytics models, surprisingly only limited number of work is focused on providing adequate analytics

P. Hitzler et al. (Eds.): ESWC 2019, LNCS 11503, pp. 412–426, 2019.
https://doi.org/10.1007/978-3-030-21348-0_27

infrastructures such as knowledge repositories or design engines to assist data analytics solution design and management [5]. Existing platforms that use meta-learning [4] or SOA and workflow-based platforms (e.g. WINGS [8], ADAGE [6]) to deliver analytics services, lack a sound information model that can capture the semantics of the analytics models [5]. Hence, they have limited ability to accumulate expert knowledge and reuse it for efficient solution design. Benefits of an enterprise-wide knowledge repository that can accumulate and link analytics related expert knowledge and resources together are multifold. It can store and link domain knowledge, analytics related facts and findings and available resources (i.e. data providers, execution services etc.) together. Data analysts can interrogate the knowledge repository to learn and get recommendations from accumulated knowledge. This reduces the time and resource spent on initial experiments and literature surveys.

As ontologies provide flexibility and extensibility to knowledge as well as the ability to integrate existing linked data, we believe capturing this knowledge through semantic models will significantly help researchers.

Limitations of Existing Resources: To identify how semantic technology is used to aid data analytics researchers, we conducted a systematic literature survey [2]. Based on the survey, we concluded that researchers could benefit from four spheres of knowledge related to an analytics problem: domain, analytics, service and intent. A majority of identified studies only used ontologies to support isolated activities such as data integration or model selection. We observed that the literature was able to cover data mining and knowledge discovery process to a certain extent (e.g., OntoDM by Panov et al. [14]). Yet there was no semantic model that can capture multiple aspects of the solution design, from variables, data, analytics models and data sources, with the ability to answer questions raised by analysts, particularly when conducting empirical analytics that try to prove a hypothesis expressed through variables via building a model [16].

RVO and Advances to State of the Art: RVO proposes a schema that can be used to record empirical data analytics research details. It can also serve as a knowledge-base to support the knowledge exploration phase in a new analytics project with the purpose of learning and recommending the choice of variables. Therefore, RVO is designed around the research variables, which form the basis of the hypothesis that analysts test through building a model [16]. Main tasks of such empirical analytics process [16] are shown in Fig. 1.

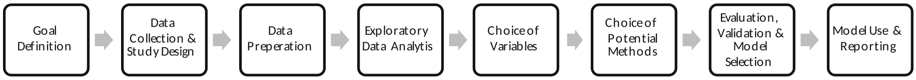

Fig. 1. Steps of building an empirical analytics model [16].

RVO can help answer an array of questions (a few examples are listed in Table 1) raised by data analysts when conducting a study. It can also provide recommendations and alternatives to assist in analysts decision making. 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 integrates with existing data models and vocabularies (such as DBpedia[1], RDF Data Cube vocabulary[2], FaBIO[3]) to facilitate efficient reuse in real-world applications by using semantic technologies and open standards (RDF, OWL, SPARQL). The main strength of RVO is its linking with other analytics-related and domain-specific ontologies.

The objectives of RVO are to:

– Assist data analytics process stages such as variable selection, data source selection, dataset selection and evaluation.
– 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 a particular model, what the relationships are between variables and datasets, and what are the relationships between variables and analytics models.
– Establish a common terminology for the organisations to represent classes and properties in the empirical analytics domain, relating to existing taxonomies, ontologies and data standards.
– Integrate existing ontologies to enrich the knowledge base in an organisation when conducting analytics. These ontologies can be domain-specific ontologies, ontologies representing analytical models, people, standards, etc.

Comparison to Other Existing Resources: Through our survey [2] we observed that domain ontologies such as DBpedia, SSN or Gene Ontology are used to express domain knowledge used for analytics. There are also ontologies to capture metadata about datasets or describe data such as RDF Data Cube vocabulary, OntoDT[4], and many researchers use domain ontologies to annotate datasets or streams to support data integration.

Besides, semantic models such as OntoDM [14] and OntoKDD [6] are designed to capture particular aspects of data analytics. For example, OntoDM has a suite of ontologies to capture the CRISP-DM process for data mining (OntoDM-KDD) and data mining entities (OntoDM-core) such as data mining tasks and algorithms. Its main limitation is in supporting empirical analytics approaches which start from hypothesis and explore variables and measures. MAMO [10], the MAthematical Modelling Ontology, provides a classification of the different types of mathematical models and their variables. However, it has limited ability to capture relationships between models, variables and expert knowledge. Ontologies such as FOAF[5] or FaBIO are useful to capture traceability or origin of the knowledge, which are both critical for researchers.

1 https://wiki.dbpedia.org.
2 http://www.w3.org/TR/vocab-data-cube/.
3 http://purl.org/spar/fabio.
4 http://www.ontodm.com/doku.php?id=ontodt.
5 http://xmlns.com/foaf/spec/.

All these ontologies represent some aspects of data analysis research; however, literature considers them in isolation. RVO is aimed to integrate these different knowledge spheres that represent domain concepts, data definitions, variables, analytic models, and their origins together to align with the thinking pattern of analytics researchers. It makes querying and exploring the knowledge intuitive and user-friendly. RVO provides a generic schema, not limited for a particular application domain or an analytics platform.

2 Relevance

Relevance to the Semantic Web community and Society: Researchers in organisations of any scale who conduct data analytics research or those who are attempting to design analytics systems, can significantly benefit from RVO. RVO can be used to accumulate and organise knowledge and learn from analytics experiments and systems. It supports linked open data practices, and resulting data can be published, shared with the wider analytics community and reused to gain insights into a problem at hand.

Relevance to Data Analytics Researchers: RVO is a schema which can be used by data analysts to record or explore knowledge related to research variables. An enterprise can adapt RVO and build analytics supporting systems around that as proposed by Bandara et al. [1]. Then researchers can easily reuse knowledge published as linked data by other data analysts using RVO or create their knowledge repository by conducting literature surveys or hiring data analytics professionals. Once the knowledge repository is in place, a minimal cost will incur for any background study in future data analytics projects. Enterprises can maintain the knowledge repository and update it when new knowledge (i.e. new models, data sources, facts) becomes available.

Impact in Supporting the Adoption of Semantic Web Technologies: RVO follows best practices in ontology design and publishing. Firstly it relies on W3C standards to make the data reusable for a variety of real-world applications. Open world assumption allows the analytics knowledge to be flexible and malleable to different organisational contexts. Moreover, it supports smooth integration with existing ontologies and the ability to be extended.

3 RVO - The Research Variable Ontology

3.1 Ontology Development Process

The ontology development process we followed aggregates ontology design principles from the NeOn methodology [17] and the methods proposed by Grüninger and Fox [9] and Ushold and King [18]. Also, we followed the FAIR principles[6] and the guidelines and best practices proposed by Noy and McGuinness [12].

[6] https://www.force11.org/group/fairgroup/fairprinciples.

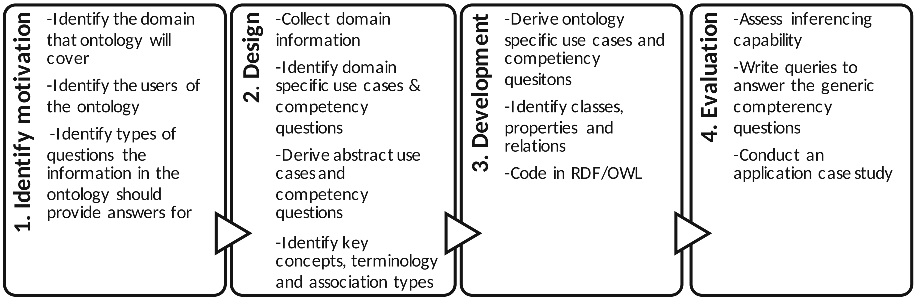

Fig. 2. Steps of the RVO development process.

The ontology development process includes four steps as illustrated in Fig. 2. We designed two domain-specific data analytics use cases in digital marketing [19] and commodity trading domains [3], to identify specific relations and terms which are required to model the domain. Use cases were designed based on data analytics literature in digital marketing and commodity trading domains as well as the authors' expertise and involvement in related projects. The use cases capture the thought process of analysts attempting an analytics problem and questions they seek answers for when designing a solution. Based on the use cases, we derived two sets of domain-specific competency questions in natural language. They formed a basis for a set of abstract competency questions that can satisfy requirement for studying and supporting data analytics solution design in general.

At the stage of identifying key concepts and association types, several meetings were held with analysts from economics and social science backgrounds to clarify and verify terminology and to identify gaps. We identified Variable, Measure, Model and Concept as four key elements for RVO. Around them, ontology specific use cases and competency questions were developed, and the classes, properties and relations were identified. A selected set of competency questions are shown in Table 1. Finally, the ontology was designed using the Protégé tool.

Ontology development was an iterative process where each version was evaluated over the generic and domain-specific competency questions and iteratively improved by evaluating its ability to answer the competency questions. After designing the ontology, we created two instance repositories for digital marketing and commodity trading and were able to answer all the generic and domain-specific competency questions via SPARQL queries. Finally, the RVO's ability to represent analytics research knowledge and support data analytics researchers in a large-scale project are illustrated through a case study, as discussed in Sect. 4.

3.2 Ontology

The components of the ontology are shown in Fig. 3. Main classes of the ontology are Variable, Measure, Model and LinkedVariables. LinkedVariables class captures a link between any two variables. Details of the classes are given in Table 2.

Table 1. Ontology specific competency questions for RVO.

Competency questions about variables
What are the variables linked to the variable of interest (V1)? What type of relationship exists between two variables of interest (V1 & V2)? What are the variables which have a causal impact on V1? What variables are assumed to be linked (hypothetically) with V1? What variables can be used to proxy V1? What is the origin/proof of relationship exists between V1 & V2? Given there is a causal link between V1 & V2, which research publications established that? Given there is a causal link between V1 & V2, which expert established that? Given there is a causal link between V1 & V2, which model was used to establish that? Which model has V1 as a dependent variable?
Competency questions about measures
What are the measures for V1? Which dataset contains measures for V1?
Competency questions about domain concept
What domain concept is represented by V1?
What are the properties of a domain concept related to V1?
Competency questions about model
What variables are involved in a model of interest (M1)? What dataset is used to train the model? What type of model is M1?

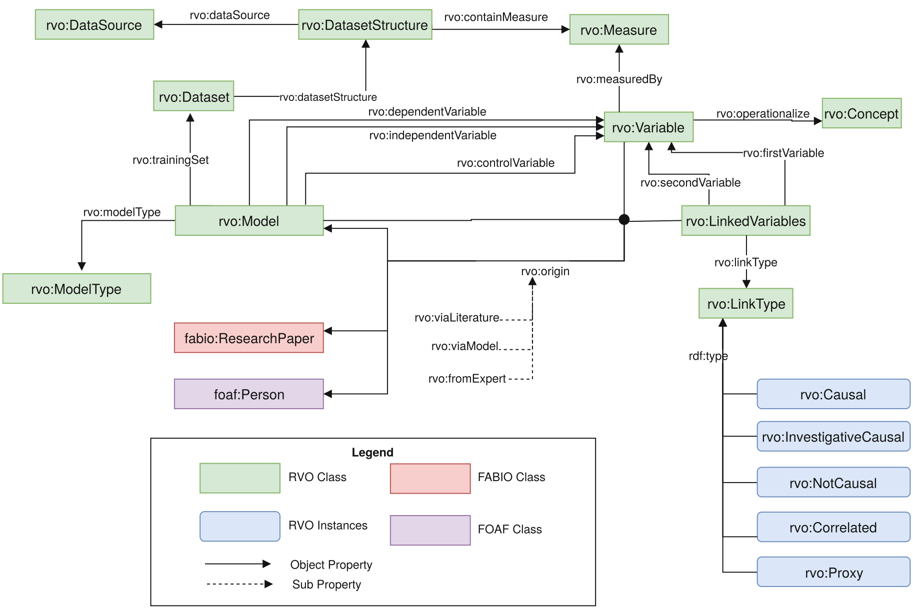

Fig. 3. Main classes and properties of RVO.

Table 2. Definitions of the classes in RVO.

Class	Definition
Concept	Concept is used to identify any domain concept that links a variable to a real-world entity. This is used to link domain ontologies to RVO and provide context to the variable
Variable	Variables exist as metrics to quantify concepts associated with a value and whose associated value may be changed
Measure	Measures are the metrics for observed values for a variable
LinkedVariables	LinkedVariables class describes a connection between two variables, the nature of this connection and its origin
LinkType	LinkType describes the type of link that exists between two or more variables. Identified LinkTypes are:- Causal: a variable causes another variable, Investigative Causal: there is a hypothesis that one variable cause another and the hypothesis needs to be tested, NonCausal: Proven not to have a causal relationship between two variables, Correlated: Two variables are dependence or associated, Proxy: One variable can be used in place of the other
DataSource	Datasource refers to a data generator or provider
Dataset	Dataset refers to a collection of observations for a set of measures, usually stored in a data file
DatasetStructure	DatasetStructure refers to the metadata for understanding a dataset such as what measures are captured, unit of measure
Model	The model is composed of a set of variables and a set of equations that establish relationships between the variables to describe particular phenomena or a system
ModelType	ModelType class is used to define types of model, (e.g.:- statistical model such as logistic regression, Bayesian model)

The most important association in RVO is defined by rvo:origin property. It facilitates traceability of the knowledge and contains three sub-properties. The origin for a certain Variable, Model or LinkedVariable could come from an expert or a research paper. Additionally, the origin of a LinkedVariable could be a Model as well.

3.3 Integrating Existing Ontologies

Domain Concepts: To fulfil the researchers' need to explore and understand the context of research, RVO facilitates linking existing domain ontologies to variables via rvo: operationalise property. For example, researches can define a variable for Gross Domestic Production and link it to DBpedia definition of GDP[7], enabling all the meta-data about GDP accessible via RVO.

Connecting to Origin: Related to rvo:origin property, we identified experts and published literature as two main types of origins. Hence we integrated two well-developed public ontologies: FOAF and FaBIO, that can represent knowledge about people and research papers respectively. Coloured rectangles in Fig. 3 illustrates the link between RVO and the concepts from FOAF and FaBIO.

Representing Data: RVO facilitates integration and reuse of existing ontologies that represent datasets and their structures. This capability is illustrated by integrating RVO with two prominent ontologies as shown in Table 3. This table shows that the RDF Data Cube vocabulary (presented by the prefix "qb") can be used to represent datasets and link rvo:Measure concept with RDF-Cube qb:MeasureProperty. This way our ontology can be easily integrated into existing linked datasets defined in RDF-Cube structure. OntoDT is a comprehensive ontology designed to capture knowledge about data types. rvo:DatasetStructure can be aligned with the aggregate datatype class that represents metadata about datasets.

Table 3. Aligning RVO with existing ontologies.

RVO concept	Aligning concept	Aligning ontology URI	Related by property
rvo:Dataset	qb:Dataset	http://purl.org/linked-data/cube	owl:sameAs
rvo:Dataset Structure	qb:DataStructure Definition	http://purl.org/linked-data/cube	owl:subClassOf
rvo:Dataset Structure	ontodt:OntoDT_378476 (aggregate datatype)	http://www.ontodm.com/OntoDT/	owl:subClassOf
rvo:Measure	qb:MeasureProperty	http://purl.org/linked-data/cube	owl:subClassOf
rvo:Model	ontodm:OntoDM_000228 (predictive model specification)	http://www.ontodm.com/OntoDM-core/	owl:subClassOf
rvo:Model	mamo:MAMO_0000037 (model)	http://identifiers.org/mamo/	owl:sameAs

Representing Models: There are many taxonomies and ontologies published by analysts and data mining experts. RVO encourages to reuse them to extend the knowledge about analytics models. RVO presents rvo:Model class as an anchor to link existing ontologies. For example, Table 3 illustrates how concepts from two prominent ontologies: OntoDM-Core and MAMO can be integrated to RVO so that data published through those can be integrated and queried through RVO.

[7] http://dbpedia.org/resource/Gdp.

3.4 How RVO Can Assist in the Analytics Process

Figure 4 presents different tasks associated with an example predictive analytics process instance to highlight how RVO can be used in multiple stages of analytics research to assist in domain understanding, planning and decision making. The classes and associations in RVO can help researchers to filter and select suitable variables, measures, compatible models and datasets from the instance repository at each stage of the analytics process. This can be done directly by answering questions as SPARQL queries or by building applications with dialogue-based front ends to support specific process flows with a SPARQL query engine in the back end. Bandara et al. [1] discuss further on how RVO can be used to express analytics requirements and use it for analytics process planning.

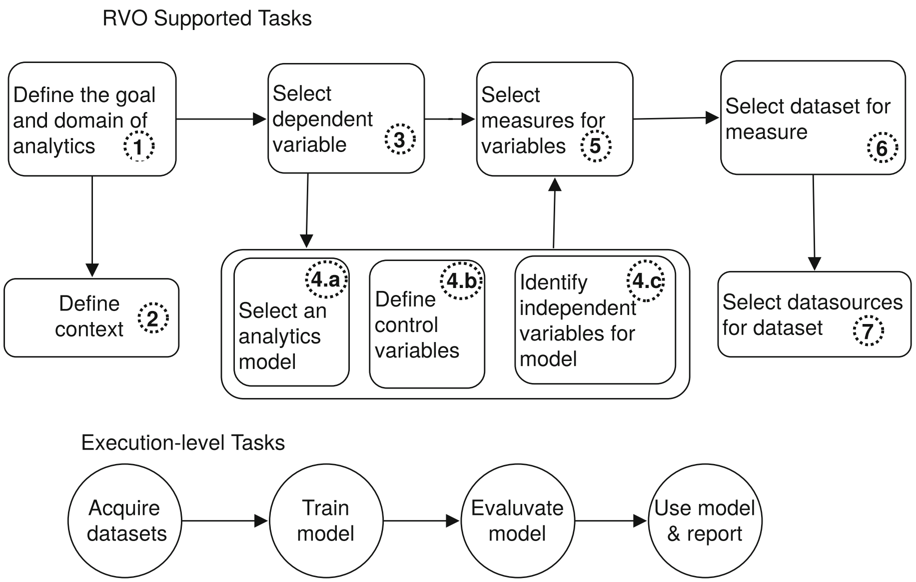

Fig. 4. Different tasks associated with an example predictive analytics process instance.

Please note that Fig. 4 is only an example process instance, and there can be other different processes with similar tasks arranged in a variety of combinations. For example, an analyst may select data first and identify model and variables aligning with that. Supporting the execution-level tasks is out of RVO's scope.

3.5 Reusability of RVO

To facilitate researcher communities to access and reuse RVO efficiently we provide the latest version of the RVO ontology to be download in RDF/XML format and turtle format. We also publish a SPARQL query interface and a REST API[8] to

[8] http://adage2.cse.unsw.edu.au/rvo/sparqlEnd.html.

explore the ontology and datasets associated with that related to a case study, so that users can navigate and understand how to use RVO to capture their analytics related knowledge. The home page of RVO website provides comprehensive documentation of the resource including schema diagram and example queries. RVO is modelled in RDF and is highly extensible. Interlinking with existing ontologies is important in reuse. We demonstrated in Sect. 3.3 how to link RVO with other ontologies. Users can replicate such integrations with other similar ontologies.

RVO follows best practices in ontology design. It uses RDF W3C standard to model and interlink the concepts. RVO adopts the open source and open data approach to make it available to a wide audience and facilitates ontology and data reuse. RVO provides differentiable URIs and implements a persistent strategy to maintain its URIs, ensuring that the same URIs are consistently reused for the same real-world objects. RVO follows FAIR principles[9] to make it findable, accessible, interoperable and reusable. The RVO description is available in human and machine readable formats at the RVO homepage and BioPortal. RVO reuses and extends into established ontologies to describe analytics and variables related information. It includes, and reuses existing vocabularies (e.g. FOAF, FaBIO).

It was a design decision to publish RVO with a small number of classes, making it more flexible and adaptable. A large taxonomy may overwhelm the users as we have observed in many existing ontologies. Further, it provides the ability to integrate existing taxonomies of models, variables etc. without any conflicts.

At the moment, RVO is used to capture variables related to a house price prediction research project, and we believe due to its unique nature of capturing and linking analytics variables, it would be used by third parties such as analytics researchers or system designers as discussed in Sect. 2. Details on how to use RVO to support data analysts can be found in our recent work [1,3].

3.6 Availability

RVO uses open standards and is publicly available under a persistent URI[10] under the MIT license[11]. The RVO homepage[12] is also published under a persistent URI, and provides information on how to use the resource. We have published software scripts to transform analytics knowledge in tabular format into RVO as open source software on GitHub[13] under the MIT License. RVO is also available in BioPortal[14] for the public to search and access.

The sustainability of RVO is ensured through three building blocks: (1) Publishing ontology as an open source resource, so that community can reuse the

9 https://www.force11.org/group/fairgroup/fairprinciples.
10 http://w3id.org/rv-ontology.
11 https://opensource.org/licenses/MIT.
12 http://w3id.org/rv-ontology/info.
13 https://github.com/madhushib/RVO.
14 http://bioportal.bioontology.org/ontologies/RVO/.

ontology to extract and represent new datasets or extend the ontology to include more classes and properties. (2) Integration of existing publicly available ontologies: the reference ontologies that are integrated with RVO are publicly available, and many of them are maintained by the community so that it is possible to maintain a fresh version of the ontology, in particular, to include new analytics knowledge. (3) Maintenance of RVO: the authors plan to perform regular RVO updates. The URIs of RVO resources will be maintained and remain stable across versions.

4 Case Study

4.1 Introduction

We conducted an application case study for RVO as a part of an ongoing data analytics solution design project, conducted by a team of researchers from the University of New South Wales, the University of Technology Sydney and the University of Wollongong in Australia. The research team comes from different backgrounds including urban planning, real estate, econometrics, statistics and software engineering. The research aim is to develop a framework that provides an accurate, long-term (20 years) forecast of real estate prices in Sydney residential property market [13]. The project outcome will contribute to the decision making of New South Wales state government, regarding complex urban transformation projects across Sydney. Based on the knowledge gathered by the research team via a literature survey and experimentations, we created an instance repository of RVO and a navigation and visualisation tool that can support analysts to explore and gain insights.

We used RVO to create an instance repository of concepts identified in the literature survey related to house price prediction, investigating more than 90 research papers. It contains knowledge about 188 variables linked to respective DBpedia concepts, 267 measures, 14 dataset types from 9 data sources, 11 house price prediction models of 9 model types and 49 research papers. Users can access and explore this knowledge through our publicly available query interface and REST API[15].

Figure 5 represents a snapshot of knowledge extracted from the work by Esteban and Altuzarra [7], represented in RVO and aligned with the analytics process presented in Fig. 4. The numbers are used to match the steps in the process with the concepts defined in the instance repository. Furthermore, each concept type according to the RVO is indicated in italic. This case study illustrates the unique capabilities of RVO as none of the related work identified through the literature survey [2] would be able to capture the links between variables, analytics models, measures and domain concepts.

[15] http://adage2.cse.unsw.edu.au/rvo/sparqlEnd.html.

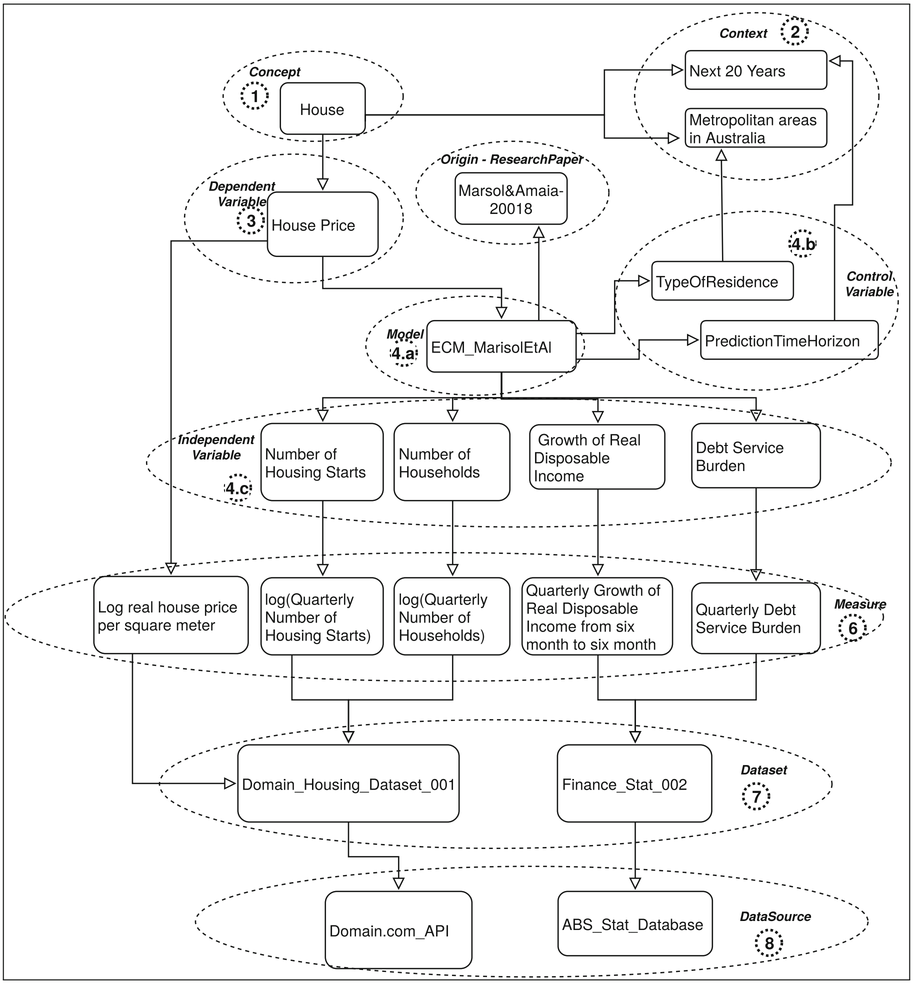

Fig. 5. Snapshot of information captured through RVO related to the house price prediction study.

4.2 Visualizing RVO

This section illustrates a tool for visualising RVO instances implemented as a web-based front end[16]. It is a navigation and visualisation tool linked to the RVO query processor. The current knowledge map demonstrates the ontological concepts and instances of RVO captured through the case study.

Figure 6 illustrates a screen-shot of the implementation for the navigation and visualisation tool. In this figure, the variables linked to "House Price" concept are shown. The top drop-down menu allows the user to select variables based

[16] http://adage2.cse.unsw.edu.au:3000/model-builder.

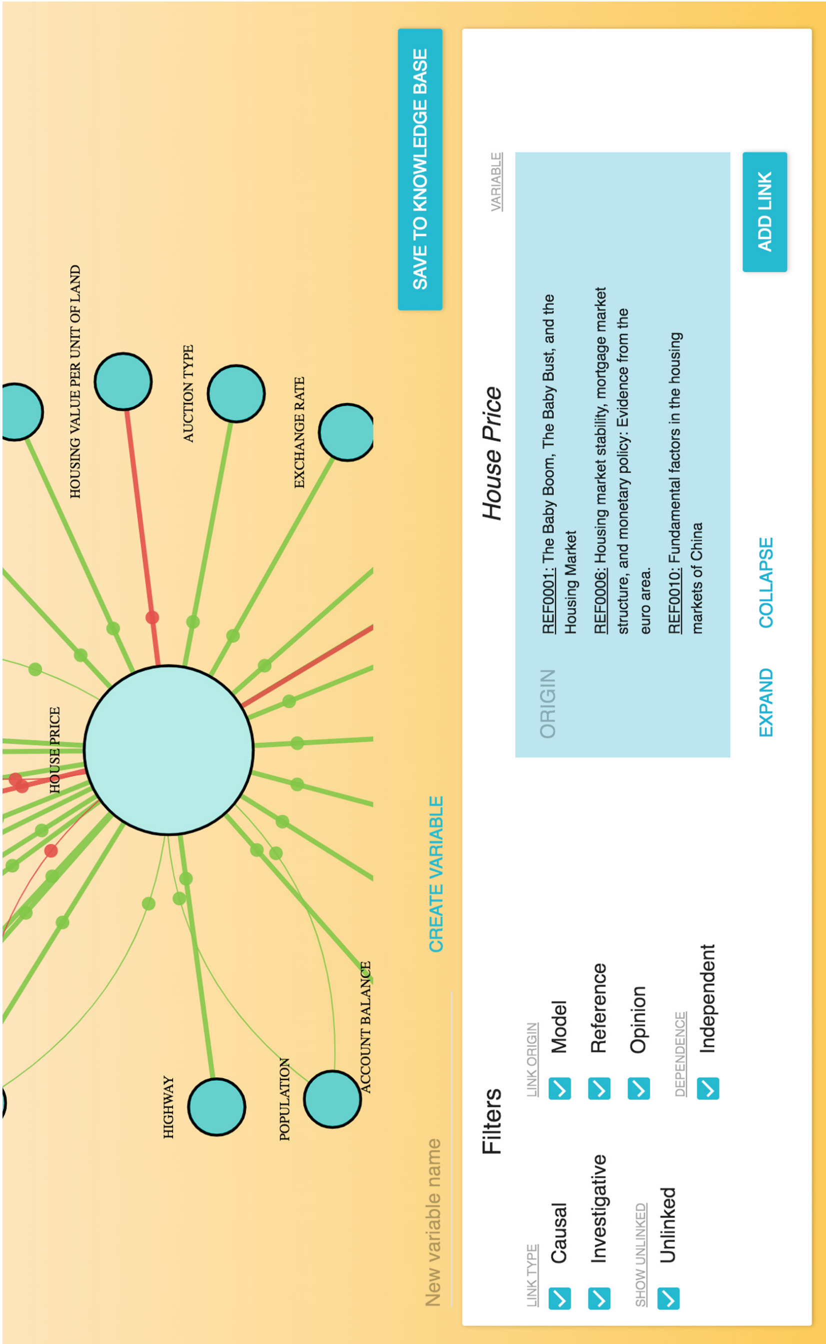

Fig. 6. Snapshot of RVO visualization & navigation tool.

on their categories. The bubble size is proportional to the number of variables connected to each variable. The directed links illustrate the connection from independent variables to dependent variables. The knowledge map interface creates an ecosystem where expert users can collaborate and contribute to forming of the knowledge base. Furthermore, the knowledge base can be applied to answer complex queries. This application can be used by research teams to explore variables, and identify those worth investigating further given their impact on house price. It saves significant time from researchers by abstracting and modelling the knowledge presented in those research papers.

5 Conclusion

This paper presents RVO- the Research Variable Ontology, designed to capture knowledge around empirical data analytics process and support researchers in different stages of the analytics process by providing recommendations and resources accessible via queries. We followed a formal process and best practices in designing the ontology and provided quality documentation and resources to reuse RVO. Further, guidelines are provided to integrate existing ontologies, and the case study demonstrates how this ontology can be used in data analytics projects.

In future, we plan to publish on how to integrate RVO with existing analytics related ontologies identified through the survey [2]. Furthermore, we plan to create more datasets around RVO, and extend its capability in capturing different knowledge spheres related to analytics. Based on that, authors plan to improve and publish new versions of RVO, and share new datasets for the use of the analytics and semantic web community.

Acknowledgment. We are grateful to Capsifi, especially Dr. Terry Roach, for sponsoring the research which led to this paper and Gilles Sainte-Marie for developing RVO web page.

References

1. Bandara, M., Behnaz, A., Rabhi, F.A., Demirors, O.: From requirements to data analytics process: an ontology-based approach. In: Daniel, F., Sheng, Q.Z., Motahari, H. (eds.) BPM 2018. LNBIP, vol. 342, pp. 543–552. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11641-5_43
2. Bandara, M., Rabhi, F.A.: Semantic modeling for engineering data analytic solutions. Semant. Web J. (2019, in press)
3. Behnaz, A., Natarajan, A., Rabhi, F.A., Peat, M.: A semantic-based analytics architecture and its application to commodity pricing. In: Feuerriegel, S., Neumann, D. (eds.) FinanceCom 2016. LNBIP, vol. 276, pp. 17–31. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52764-2_2
4. Brazdil, P., Carrier, C.G., Soares, C., Vilalta, R.: Metalearning: Applications to Data Mining. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-73263-1

5. Crankshaw, D., Gonzalez, J., Bailis, P.: Research for practice: prediction-serving systems. Commun. ACM **61**(8), 45–49 (2018). https://doi.org/10.1145/3190574
6. Diamantini, C., Potena, D., Storti, E.: KDDONTO: an ontology for discovery and composition of KDD algorithms. In: Third Generation Data Mining: Towards Service-Oriented Knowledge Discovery (SoKD 2009), pp. 13–24 (2009)
7. Esteban, M., Altuzarra, A.: A model of the spanish housing market. J. Post Keynes. Econ. **30**(3), 353–373 (2008)
8. Gil, Y., Kim, J., Ratnakar, V., Deelman, E.: Wings for Pegasus: a semantic approach to creating very large scientific workflows. In: OWLED (2006)
9. Grüninger, M., Fox, M.S.: Methodology for the design and evaluation of ontologies (1995)
10. Imbert, S.: Mathematical Modelling Ontology (2017). http://bioportal.bioontology.org/ontologies/MAMO. Accessed 30 Nov 2018
11. Magdon-Ismail, M.: No free lunch for noise prediction. Neural Comput. **12**(3), 547–564 (2000). https://doi.org/10.1162/089976600300015709
12. Noy, N.F., McGuinness, D.L., et al.: Ontology development 101: a guide to creating your first ontology (2001)
13. UrbanGrowth NSW: Funding cities community of practice. Review and development of a predictive housing price model for the Sydney housing market (2017)
14. Panov, P., Džeroski, S., Soldatova, L.: OntoDM: an ontology of data mining. In: IEEE International Conference on Data Mining Workshops, ICDMW 2008, pp. 752–760 (2008)
15. Rabhi, F., Bandara, M., Namvar, A., Demirors, O.: Big data analytics has little to do with analytics. In: Beheshti, A., Hashmi, M., Dong, H., Zhang, W.E. (eds.) ASSRI 2015/2017. LNBIP, vol. 234, pp. 3–17. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76587-7_1
16. Shmueli, G., Koppius, O.R.: Predictive analytics in information systems research. Mis Q. **35**, 553–572 (2011)
17. Suárez-Figueroa, M.C.: NeOn Methodology for building ontology networks: specification, scheduling and reuse. Ph.D. thesis, Informatica (2010)
18. Uschold, M., King, M.: Towards a methodology for building ontologies. Citeseer (1995)
19. Yang, S., Lin, S., Carlson, J.R., Ross Jr., W.T.: Brand engagement on social media: will firms' social media efforts influence search engine advertising effectiveness? J. Mark. Manag. **32**(5–6), 526–557 (2016)