Semantic Web 0 (0) 1
IOS Press

The Nova Scotia Disease Knowledge Graph

Enayat Rajabi [a,*], Rishi Midha [a], Jairo Francisco de Souza [b]
[a] *Shannon School of Business, Cape Breton University, Sydney, NS, Canada*
E-mails: enayat_rajabi@cbu.ca, cbu19ffj@cbu.ca
[b] *Department of Computer Science, Federal University of Juiz de Fora, Brezil*
E-mail: jairo.souza@ice.ufjf.br

Abstract. The majority of published datasets in open government data are statistical. They are widely published by different governments to be used by the public and data consumers. However, most datasets in open data portals are not provided in RDF format. Moreover, the datasets are isolated from one another, while conceptually connected. Through this paper, a knowledge graph is constructed for the disease-related datasets of a Canadian government data portal, Nova Scotia Open Data. We transformed all the disease-related datasets to RDF and enriched them by semantic rules and an external ontology. The study shows that integrating open statistical datasets from multiple sources using ontologies and interlinking them potentially leads to valuable data sources and generates a dense knowledge graph with cross-dimensional information. The ontology designed to develop the graph adheres to best practices and standards, allowing for expansion, modification and flexible re-use.

Keywords: Open Statistical Data, Nova Scotia, Knowledge Graph, Disease Ontology

1. Introduction

The open government data movement has led to open data portals that provide a single point of access for a province or country. Open government data increases government transparency, accountability, contributes to economic growth and improves administrative processes [1]. This data is published hoping that it can be used by different organizations, data consumers in the public and private sectors. A variety of published open data include multidimensional and statistical information such as census data, demographics, public health data (e.g., number of disease cases) or business data (number of unemployment), which can be used in public services and provides social value to citizens [2, 3]. In itself, the data can be restrictive and not powerful enough to draw meaningful inferences from. The datasets act as isolated pools of information that cannot be queried or linked. These sources are scattered in the government data portals, and users can access the information through specific searches in that data portal. The lack of meaning behind the statistical data makes it impossible to form a network and link this kind of data to infer, create and query knowledge [4]. Interconnectivity between isolated datasets in open data gives a machine a lot of information to work with, thereby strengthening its ability to deduce relations and infer meaning. By transforming the isolated datasets in an open data government into a set of linked datasets, a knowledge graph can be generated and meaningful information can be inferred and queried [5]. This study focuses on constructing a knowledge graph for disease-related datasets of Nova Scotia Open Data (NSOD), Canada's provincial Open Data portal. Overall, there are 11 provinces and territories in Canada with approximately 11,771 published datasets in different domains ranging from "Business and Economy" to "Health and Wellness" [5]. Most open data portals present their data using other formats, including CSV, JSON, and Excel. A few of them allow users to export data in RDF format, although they do not follow the Linked Data vocabulary five-star standards [6]. The datasets are also isolated within or among the data portals, while many of them can be conceptually linked. The majority of datasets in this domain include statistical reports. For example, statistical data regarding diseases in a province in different years, however, they should

*Corresponding author. E-mail: enayat_rajabi@cbu.ca.

be manually analyzed to answer some essential questions like: "Which viral diseases had the most number of cases in a province in 2017?". This study intends to answer such questions using the Semantic Web technologies such as ontologies, RDF multi-dimensional models, and deductive reasoning rules and generate a knowledge graph with semantic relationships. As there are several disease-related datasets in NSOD that are not connected and the type of diseases is not evident in the raw data, we link the instances of the datasets (metadata, dimensions, measures, and attributes) semantically on a schema-level following the W3C vocabularies and enrich them by a disease ontology. After creating the knowledge graph for disease datasets, a quality and refinement process was performed using a specific quality metric to measure the accuracy and precision of the created knowledge graph and its entities and classes based on refinement standards [7, 8].

The structure of this paper is as follows: Section 2 explains the background and the related studies in publishing datasets, particularly in the domain of multi-dimensional data. Section 3 describes the existing NSOD dataset. Section 4 presents the architecture, model, and ontologies used in this paper to construct the knowledge graph. Section 5 outlines the transformation challenges, followed by a conclusion in Section 6.

2. Background

To construct a knowledge graph for open statistical data, a multi-dimensional structure should be defined consists of measures (e.g., number of cases) and dimensions describing the measures (e.g., regions). There is a growing need for the statistical data to be published so that it can be linked to related datasets and combined with associated information (metadata included). The literature cites many examples where researchers and organizations alike have implemented the RDF Data Cube vocabulary. [9] describes the process of improving and enriching the quality of Barcelona's official open data platform employing multi-dimensional data, applying linked open data assessment process and using external repositories as a knowledge base. In another example, [10] described how the Czech Social Security Administration (CSSA) published their official pension statistics as linked open data (LOD). These LOD datasets were modelled using the Simple Knowledge Organization System (SKOS) vocabulary and the RDF Data Cube Vocabulary. The use of open statistical data in the medical industry, in the form of health reports, medical reports and other such data, has been used in the literature. An ontology in the health IT interventions domain, developed and published in the study [11], builds on existing health and medical ontologies. The study outlines an inductive-deductive approach to establish a glossary, define concepts such as classes and instances, and finally publish the ontology as linked open data. The PubMed knowledge graph [12] is another study in this domain created from the PubMed library. The study outlines the extraction of over 29 million records from the library to generate a graph that links bio-entities, authors, funding, affiliations and articles. Subsequent data validation yielded promising results, and the graph can create and transfer knowledge, profile authors and organizations and realize meaningful links between bio-entities. The study covers familiar territory in terms of knowledge graph and generation compared to the work done in this research study.

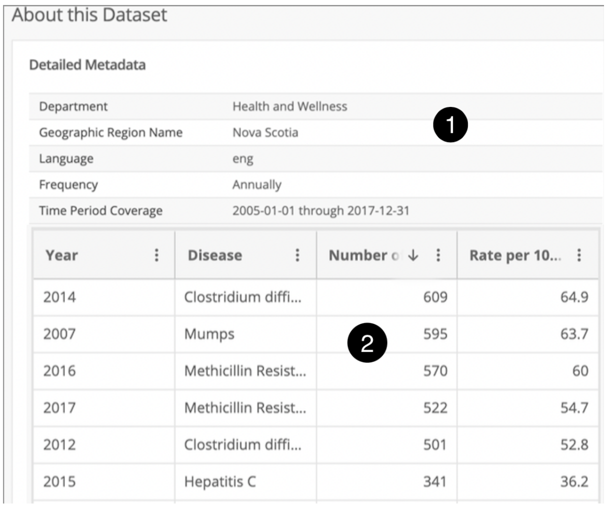
Similar various studies for publishing datasets in different domains [13, 14] based on the Semantic Web recommendations, this work publishes a public knowledge graph in the context of statistical government data to link different isolated published datasets. The use of Linked Data standards and patterns [6, 15] and strict adherence to well-established rules and protocols of the semantic Web prescribed by W3C ensure compatibility with past works as well.

3. Nova Scotia Open Data

Nova Scotia's government has an abundance of resources in terms of data and information. All this data has been collected and stored on the Nova Scotia Open Data (NSOD) web portal [1] in the form of datasets. The NSOD datasets are available thourgh Socrata API[2]. The main purpose of this web portal is to allow individuals, particularly Nova Scotians, to efficiently access the information, understand their government, support their businesses, gain new insights, and make discoveries. In this study, we retrieved the NSOD datasets using Socrata API using Python[3] programming language. We wrote a command-line tool to fetch the datasets and performed an exploratory analysis to understand the data. At the time of this research, there are 669

[1] https://data.novascotia.ca
[2] https://dev.socrata.com/
[3] https://www.python.org

datasets in 28 categories, of which 77.8% are archived datasets and 22.2% are currently active. The majority of the datasets have been created between April 2016 and June 2016, and they are gradually updated each year. In terms of language, the majority of datasets are in English. Around 79.7% of the datasets have Nova Scotia province defined as their region, while 20.3% datasets have missing values in region metadata. The top categories of datasets are “Environment and Energy" (58), “Health and Wellness" (52), “Population and Demographics" (48), “Business and Industry" (37) and “Education" (32). Overall, there are 21 disease-related datasets in the department of “Health and Wellness" category found by searching in the NSOD web portal (see Table 1). Each dataset has a metadata section and an observation section which includes the statistical observations. As Figure 1 shows, these datasets have the same number of attributes in both metadata and observation sections. There are 13 observations in each dataset, including statistical information about specific disease cases in Nova Scotia between 2005 and 2017.


About this Dataset

Detailed Metadata

Department	Health and Wellness
Geographic Region Name	Nova Scotia
Language	eng
Frequency	Annually
Time Period Coverage	2005-01-01 through 2017-12-31

1

Year	Disease	Number o	Rate per 10...
2014	Clostridium diffi...	609	64.9
2007	Mumps	595	63.7
2016	Methicillin Resist...	570	60
2017	Methicillin Resist...	522	54.7
2012	Clostridium diffi...	501	52.8
2015	Hepatitis C	341	36.2

2

Fig. 1. A disease dataset in the NSOD web portal (1: metadata, 2:observations)

4. Methodology

A knowledge graph construction process can be performed based on the following steps: a) Knowledge acquisition: Collecting semi-structured data from an API, b) Knowledge extraction: Extracting entities and their relationships, c) Knowledge fusion: Constructing an ontology, assigning entities and relationships to the knowledge graph based on the defined ontology, and interlinking entities to external ontologies and datasets, and d) Knowledge storage: Storing the created knowledge graph in a triple store. To generate a knowledge graph for the disease datasets of NSOD, we follow the W3C standards to transform the ingested datasets to RDF using a data model, a custom ontology, a set of semantic rules, and an interlinking process. The following subsections describe the steps in detail.

4.1. Data Model

An open government dataset can include statistical information that corresponds to a defined structure. The data dictionary or metadata of each NSOD dataset consists of information about that specific dataset, such as name, publisher, publication date, category, department, etc which can be transformed to RDF using DCMI [4], DCAT [5], and RDFS vocabularies.

The observation of an NSOD dataset includes a collection of dimensions, measures and attributes. The dimensions act as unique identifiers i.e. with values for each dimension and a record can be identified. The measures are the actual values observed across logical space, time, or any other dimension. The attributes quantify the measurement and helps in its interpretation. The dimension, measures, and attributes of a dataset together comprise its structure and are thus aptly stored in the Data Structure Definition (DSD). Figure ?? shows an example of observation in an NSOD dataset.

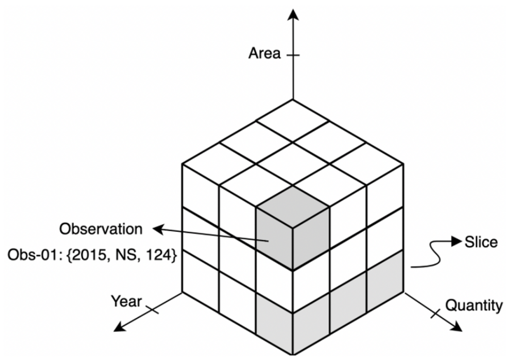


Fig. 2. An example of observation in an open statistical dataset [5]

[4]https://dublincore.org
[5]https://www.w3.org/TR/vocab-dcat-2/

Table 1
NSOD disease datasets

Dataset	URL
Giardiasis	https://data.novascotia.ca/Health-and-Wellness/Giardiasis/dxhy-v63n
HepatitisC	https://data.novascotia.ca/Health-and-Wellness/HepatitisC/w7bm-jav4
Gonorrhea	https://data.novascotia.ca/Health-and-Wellness/Gonorrhea/qsy8-u73r
Legionellosis	https://data.novascotia.ca/Health-and-Wellness/Legionellosis/fbx9-xzkq
Salmonellosis	https://data.novascotia.ca/Health-and-Wellness/Salmonellosis/5kjg-qidk
Typhoid	https://data.novascotia.ca/Health-and-Wellness/Typhoid/p42z-mdp9
Shigellosis	https://data.novascotia.ca/Health-and-Wellness/Shigellosis/s4zv-fef7
Pertussis	https://data.novascotia.ca/Health-and-Wellness/Pertussis/5itt-58m2
Cyclosporiasis	https://data.novascotia.ca/Health-and-Wellness/Cyclosporiasis/enig-mjwt
Cryptosporidiosis	https://data.novascotia.ca/Health-and-Wellness/Cryptosporidiosis/9ckr-t936
HepatitisA	https://data.novascotia.ca/Health-and-Wellness/HepatitisA/r9i7-qnuh
Campylobacteriosis	https://data.novascotia.ca/Health-and-Wellness/Campylobacteriosis/nn2f-jgv4
Rubella	https://data.novascotia.ca/Health-and-Wellness/Rubella/5mqu-g4g2
Botulism	https://data.novascotia.ca/Health-and-Wellness/Botulism/4vij-qbup
Tuberculosis	https://data.novascotia.ca/Health-and-Wellness/Tuberculosis/48h6-dubm
Mumps	https://data.novascotia.ca/Health-and-Wellness/Mumps/e2eh-63yz
Chlamydia	https://data.novascotia.ca/Health-and-Wellness/Chlamydia/sps3-eq7e
HumanImmunodeficiencyVirus	https://data.novascotia.ca/Health-and-Wellness/HumanImmunodeficiencyVirus/tsr3-8hh2
AcquiredImmuneDeficiencySyndrome	https://data.novascotia.ca/Health-and-Wellness/AcquiredImmuneDeficiencySyndrome/naay-xy8i
Tetanus	https://data.novascotia.ca/Health-and-Wellness/Tetanus/sr7h-uyxt
Malaria	https://data.novascotia.ca/Health-and-Wellness/Malaria/4p4c-dv5x

To model the multi-dimensional NSOD datasets, the RDF Data Cube vocabulary [6] is usually used, which is a W3C recommendation and has been widely used in different studies [5, 9, 16]. The RDF Cube allows publishers to integrate and slice across their datasets [17] and enables the representation of the statistical data in standard RDF format and publishes the data conforming to the principles of linked data [18].

4.2. Ontology

To the best of our knowledge, there were no existing ontologies that could be re-used based on the nature of the NSOD dataset. However, we re-used an existing data model for describing multi-dimensional data (RDF Cube vocabularies), an external disease ontology (DOID), and the best practice vocabularies such as SDMX to develop a custom ontology for disease-related datasets of NSOD. The datasets were coded as entities with distinct data structure definitions, slices and observations.

All the datasets in the ontology are all instances of class *DataSet* and the nomenclature used for datasets is *dataset-dataset_name*. Each dataset has one associated data structure definition (*DataStructureDefintion*), which defines the dataset's dimensions, measures, and attributes and is linked with *DataSet* by *structure* property. The dimensions, measures and attributes are linked with the data structure definition by properties *dimension*, *measure*, and *attribute* respectively. The dimensions can further be used to form logical slices of the cube structure of datasets. Instances of class qb:*Slice* and sub-class of *ObservationGroup* are used to group observations by fixing the value of one or more dimensions. Each slice is linked to the data structure definition using *sliceKey* property. Observations include the values for all the measures, called the multi-measure approach. The observations are attached to a dataset by the *observation* property and the respective slices by the *observationGroup* property. Figure 3 illustrates a sample observation based on the defined ontology. Table 2 also shows the prefixes used in the ontology.

[6]https://www.w3.org/TR/vocab-data-cube/

Table 2

Re-used vocabularies

Vocabulary	Prefix	Usage
RDF Cube	http://purl.org/linked-data/cube#	Multi-dimensional observations
Dublin Core	http://purl.org/dc/terms/	Metadata of datasets
DOID	http://purl.obolibrary.org/obo/doid#	The disease ontology
GeoNames	http://www.geonames.org/ontology#	Geographical information
Statistical Data and Metadata eXchange	http://purl.org/linked-data/sdmx/2009/code#	Dimensions and measures
Semnatic Web Rule Language	http://swrl.stanford.edu/ontologies/3.3/swrla.owl#	Semantic rules

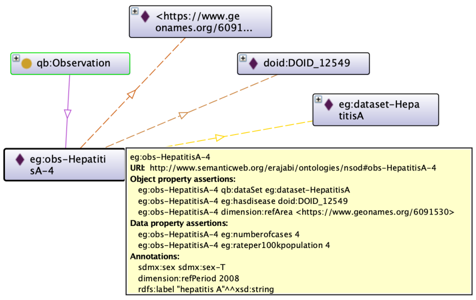


Fig. 3. A sample observation in the knowledge graph

4.3. Interlinking to External Ontology and Datasets

An external ontology (DOID[7]) has been used to enrich the knowledge graph with domain knowledge. We imported the DOID ontology into the knowledge graph. We linked the disease name and the super-classes of each disease to the disease ontology based on the similarity of the disease names. The interlinking of datasets through their parent class is carried out, which enriches the datasets to create a sound knowledge base. We also used Geonames [8] to represent regional dimension information instead of literals adds another possibility for knowledge inference and creation. This allows the addition of semantics to statistical data in case the other datasets are joined.

4.4. Rules

Complex formal semantics in a knowledge graph allows a reasoner to infer the relationship between data items in different datasets [19]. This step was carried out to add more meaning to data and form a dense knowledge graph and add another layer of complexity to the graph. This helps add another semantic layer and links the data together. The Semantic Web Rule Language (SWRL[9]), an example of a Rule Markup Language to standardize the publishing and sharing of inference rules, is used to incorporate and build semantic relationships. As a proof of concept, we designed a SWRL rule to infer the transitive relationship of diseases in a dataset using Protege[10] rule engine. This implies that if an observation includes a disease x which is a form of disease y (in the disease ontology), the graph will infer that observation x includes disease y implicitly.
The rule states that:

$$hasdisease(?x, ?y) \wedge doid{:}is\_a(?y, ?z) \implies hasdisease(?x, ?z)$$

Another semantic rule example is related to the observations with the highest number of cases for a particular disease. Based on the current number of cases in each disease in the Nova Scotia province, we considered 1,000 disease cases per 100,000 population is high in the Nova Scotia province. Those observations are defined in the following rule:

$$Observation(?obs) \wedge numberOfCases(?obs, ?n) \wedge swrlb{:}greaterThan(?n, 1000) \implies HighDiseaseCases(?obs)$$

The rule can be made highly specific by using constraints on threshold N (number of disease cases) serving as a cut-off to classify common diseases as well

[7] https://disease-ontology.org/
[8] https://www.geonames.org
[9] https://www.w3.org/Submission/SWRL/
[10] https://protege.stanford.edu

as other dimensions such as region, period, gender and disease.

4.5. Transformation Process

A knowledge graph can be constructed in a) top-down approach where the entities are added to the knowledge-base based on a predefined ontology, or b) bottom-up approaches where knowledge instances are extracted from knowledge base systems and then, the top-level ontologies are built based on the knowledge instances to create the whole knowledge graph [20]. In this study, we followed the top-down approach to construct a disease knowledge graph from NSOD disease datasets (see Figure 3). We gathered data and transformed it into RDF triples using the designed ontology and data model described in the previous sections. The ontology was then extensively processed to enrich data through internal and external linking and dimensional and logical relations. The structural metadata about the dimensions and measures of the NSOD datasets are different in general. We developed a configuration setting to specify the dimensions and measures of each dataset, in case other datasets with various dimensions and measures are added. This allows semi-automatic updating of the graph with input data and makes the datasets semantically and dimensionally connected to the external ontologies and the Linked Open Data cloud. For example, several disease datasets had *number of cases* property that could be used as one predicate (*eg:numberOfCases*) across the knowledge graph.

In the transformation process, the Dublin Core Metadata as the most widely used metadata schema, was used to describe the metadata elements of datasets such as published date, dataset title, subject or category, source, contributor, etc. The corresponding elements of each observation were also mapped to RDF triples based on the vocabularies mentioned in Table 3).

The defined rules are also translated into the constructor component to enable semantic reasoning over the knowledge graph. Finally, the datasets are added onto the graph as observations, ensuring that they conform to prescribed metadata, structure, and semantic web protocols. The graph was subjected to a quality and refinement check, and it is checked against well-received field works in terms of concept, schema, entity instances, and relations. This is followed by query retrieval to answer questions using SPARQL.

Table 3
Mapping vocabularies

Section	Element	Mapping voacbulary
Metadata	Dataset licence	dct:license
Metadata	Dataset language	dct:language
Metadata	Department	:department
Metadata	Dataset description	rdfs:comment
Metadata	Dataset keyword	dcat:keyword
Metadata	Dataset suject	dcat:theme
Observation	Year of observation	sdmx-dimension:refPeriod
Observation	Region of observation	sdmx-dimension:refArea
Observation	Number of cases for each disease	:numberOfCases
Observation	An observation belongs to a disease	:hasdisease
Observation	Case rate per 100,000 population	:rateper100kpopulation
Observation	Gender in observation	sdmx:sex
Observation	Geolocation of dataset	dct:spatial

4.6. Queries

We used the built-in SPARQL tab in Protege to pose a set of designed queries against the knowledge base through additional semantics, which cannot be explicitly expressed through linkage. The questions were designed with the help of Nova Scotia health stakeholders. In designing the question, we considered the semantic rules developed in Section 4.4 in the knowledge graph. For example, some of the disease datasets were the sub-classes of the infectious disease class in the disease ontology, and we can use this property to retrieve the results. The queries are outlined below.

Figure 5 shows two queries that we defined along with the sample results. In both queries, we leveraged the rules that we defined before.

Query 1: List of the viral infectious diseases along with their number of cases in Nova Scotia in different years.
In this query, we use *doid:is_a* relationship rule to identify all the disease classified as "viral infectious diseases".

Query 2: List of viral infectious diseases with high number of cases (more than 1,000 cases) in Nova Scotia in 2017.
In this question, we use the *HighDiseaseCases* class to infer the results based upon the rule defined in Section 4.4.

The results of queries were cross-checked and validated for accuracy and completeness. We also performed a knowledge graph refinement process to enhance the overall quality of the knowledge graph. It includes identifying and subsequently adding the miss-

[11] An online SPARQL editor was used to improve the readability of the SPARQL Queries.

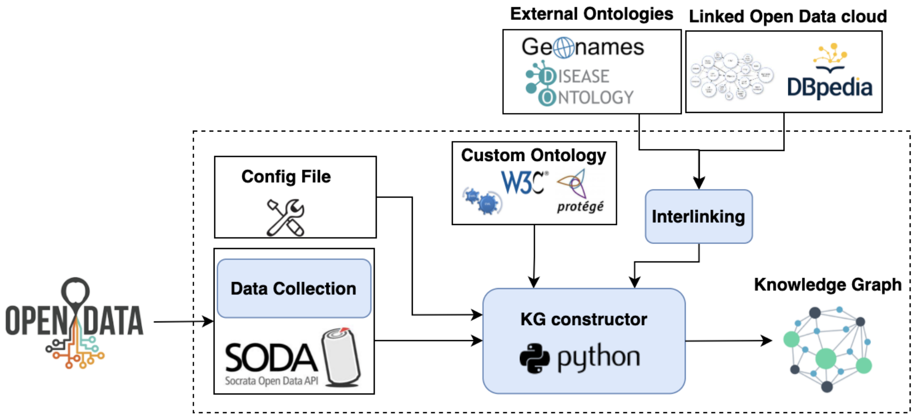


Fig. 4. An example of observation [5]

```
Query 1:

SELECT ?disease_label ?disease_parent ?numofcases ?year
{
    ?observation eg:hasdisease ?disease.
    ?observation rdfs:label ?disease_label.
    ?disease rdfs:label ?disease_parent.
    FILTER regex(?disease_name,
                    "viral infectious disease", "i")
    ?observation eg:numberofcases ?numofcases.
    ?observation dimension:refPeriod ?year.
}
Query 2:

SELECT  ?disease_label ?num ?period
{
    ?obs a eg:HighDiseaseCases;
    rdfs:label ?disease_label;
    eg:numberofcases ?num;
    dimension:refPeriod ?period
}
```

Fig. 5. Query [11]

ing knowledge as well as correcting erroneous information. The metrics to determine the quality of a knowledge graph have been theorized based on the various refinement techniques. To determine some of these metrics, the tool OntoMetrics[12] has been utilized. The results show that the knowledge graph quality checked passed all the tests (see Table 4).

4.7. Knowledge Graph

The final knowledge graph included 2,883 triples with 24 classes, 23 object properties, and two data properties. All 21 disease datasets were transformed to the knowledge graph successfully with the total of 252 observation. Each observation includes Gender (*sdmx:sex*), disease information (*eg:hasdisease*), observation year (*dimension:refPeriod*), disease label (*rdfs:label*), disease rate per 100k population of disease (*eg:rateper100kpopulation*), area of observation (*dimension:refArea*) and number of disease cases (*numberofcases*) properties. The knowledge graph is publicly available at Zenodo [13] under Creative Commons Universal Public Domain Dedication (CC0 1.0)[14] license.

Table 4

Quality Check Metrics With Values

Quality Check	Description	Metric	Value
Accuracy	The correctness and validity of the information presented, verified against a legitimate source.	Spelling Error Rate	0%
Domain-specificity	A horizontal or shallow ontology (high) covers more domains but not in-depth and a vertical or deep ontology (low) domain specific.	Inheritance Richness	77%
Consistency	The adherence to a structure i.e. precision.	Inconsistent Terms Ratio	0%
Informative	The information conveyed by ontology on the basis of relationships.	Relationship Richness	64%

5. Conclusion and Remarks

The study demonstrates that the integration of disease-related datasets of an open government data

[12]https://ontometrics.informatik.uni-rostock.de/ontologymetrics/

[13]https://doi.org/10.5281/zenodo.5517236

[14]https://creativecommons.org/publicdomain/zero/1.0/

portal. Ontologies and interlinking features of the Semantic Web were used to generate a knowledge graph to answer questions that the open statistical data portals in its current state cannot answer due to the lack of semantics and meaning.

Due to certain limitations identified below, there is a hindrance in the attainment of complete automation of constructing a knowledge graph. Although we developed a tool to retrieve open datasets from the NSOD portal, identification of disease datasets was carried out manually. That makes the knowledge graph construction process semi-automatic, where fetching data and creating knowledge graph are done separately, which will be addressed in the future downing the road of this study.

All the disease-related datasets found in the NSOD portal contain the same number of dimensions with the same data type. This might not be true for all the datasets though. One of the challenges in transformation is having a different number of dimensions with various data types. Lack of descriptive structural metadata that enlists the dimensions, measures, and attributes of each dataset explicitly is one of the biggest hurdle towards achieving complete automation. Alternatively, the lack of a vocabulary that supports properties that convey this information is another issue that prevents us from addressing it in a standardized manner.

Another problem is the lack of consistency across datasets in terms of structure and nomenclature. Structural consistency refers to strict rules defining the attributes in terms of atomicity and datatypes. For instance, it is best practice to express fields such as an address in a partitioned fashion comprising various elements such as postal code and street address to simplify information retrieval. The nomenclature for certain common dimensions and measures is not regularised, which further causes challenges in mapping.

Based on our exploratory analysis, there are different provincial open data portal across Canada that publish datasets with same structure and related topics. A Linked Data strategy, similar to what we described in this article, can be used to build an endpoint (e.g., in the Canada Open Data portal [15]) to link similar open statistical datasets across the country and facilitate query answering for data consumers and the open data community.

[15] https://open.canada.ca/

6. Acknowledgement

The work conducted in the study has been funded by the MITACS Research Training (IT21970) and NSERC (Natural Sciences and Engineering Research Council) Discovery Grant (RGPIN-2020-05869).

References

[1] R. P. Lourenço, "An analytvis of open government portals: A perspective of transparency for accountability," *Government information quarterly*, vol. 32, 2015.

[2] E. Kalampokis, D. Zeginis, and K. Tarabanis, "On modeling linked open statistical data," *Journal of Web Semantics*, 2019.

[3] A. Gregory and M. Vardigan, "The Web of Linked Data: Realizing the Potential for the Social Sciences," tech. rep., 2010.

[4] J. Marden, C. Li-Madeo, N. Whysel, and J. Edelstein, "Linked open data for cultural heritage: Evolution of an information technology," in *SIGDOC 2013 - Proceedings of the 31st ACM International Conference on Design of Communication*, 2013.

[5] E. Rajabi, "Towards linked open government data in Canada," *International Journal of Metadata, Semantics and Ontologies*, vol. 14, no. 3, pp. 209–217, 2021.

[6] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data: The story so far," in *Semantic services, interoperability and web applications: emerging concepts*, pp. 205–227, IGI global, 2011.

[7] B. Stvilia, "A model for ontology quality evaluation," *First Monday*, vol. 12, no. 12, 2007.

[8] D. Vrandečić, *Ontology Evaluation*, pp. 293–313. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[9] P. Escobar, G. Candela, J. Trujillo, M. Marco-Such, and J. Peral, "Adding value to Linked Open Data using a multidimensional model approach based on the RDF Data Cube vocabulary," *Computer Standards and Interfaces*, 2020.

[10] J. Klímek, J. Kučera, M. Nečaský, and D. Chlapek, "Publication and usage of official Czech pension statistics Linked Open Data," *Journal of Web Semantics*, 2018.

[11] V. Dornauer, M. Ghalandari, K. Höffner, F. Jahn, B. Schneider, A. Winter, and E. Ammenwerth, "Challenges and solutions while developing hito, a health it ontology," 07 2019.

[12] J. Xu, S. Kim, M. Song, M. Jeong, D. Kim, J. Kang, J. F. Rousseau, X. Li, W. Xu, V. I. Torvik, Y. Bu, C. Chen, I. A. Ebeid, D. Li, and Y. Ding, "Building a PubMed knowledge graph," *Scientific Data*, vol. 7, no. 1, p. 205, 2020.

[13] L. Lefort, A. Haller, K. Taylor, G. Squire, P. Taylor, D. Percival, and A. Woolf, "The acorn-sat linked climate dataset," *Semantic Web*, vol. 8, no. 6, pp. 959–967, 2017.

[14] K. Höffner, M. Martin, and J. Lehmann, "Linkedspending: Openspending becomes linked open data," *Semantic Web*, vol. 7, no. 1, pp. 95–104, 2016.

[15] L. Dodds and I. Davis, "Linked data patterns," *Online: http://patterns. dataincubator. org/book*, 2011.

[16] C. van Ooijen, B. Ubaldi, and B. Welby, "A data-driven public sector: Enabling the strategic use of data for productive, inclusive and trustworthy governance," OECD Working Papers on Public Governance 33, OECD Publishing, May 2019.

[17] C. Debruyne, D. Lewis, and D. O'Sullivan, "Generating executable mappings from RDF data cube data structure definitions," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[18] R. E. Cyganiak and D. E. Reynolds, "The RDF Data Cube Vocabulary," *W3C Recommendation*, 2014.

[19] A. Callahan, J. Cruz-Toledo, and M. Dumontier, "Ontology-Based Querying with Bio2RDF's Linked Open Data," *Journal of Biomedical Semantics*, 2013.

[20] Z. Zhao, S.-K. Han, and I.-M. So, "Architecture of knowledge graph construction techniques," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 19, pp. 1869–1883, 2018.

[21] M. Koho, E. Hyvönen, E. Heino, J. Tuominen, P. Leskinen, and E. Mäkelä, "Linked Death—Representing, Publishing, and Using Second World War Death Records as Linked Open Data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[22] M. Moshref and R. Al-Sayyad, "Developing Ontology Approach Using Software Tool to Improve Data Visualization (Case Study: Computer Network)," *International Journal of Modern Education and Computer Science*, 2019.

[23] M. R. Kamdar and M. A. Musen, "PhLeGrA: Graph analytics in pharmacology over the web of life sciences linked open data," in *26th International World Wide Web Conference, WWW 2017*, 2017.

[24] M. Alobaidi, K. M. Malik, and S. Sabra, "Linked open data-based framework for automatic biomedical ontology generation," *BMC bioinformatics*, 2018.

[25] M. Mohd Ali, R. Rai, J. N. Otte, and B. Smith, "A product life cycle ontology for additive manufacturing," *Computers in Industry*, 2019.

[26] E. Ammenwerth, V. Dornauer, M. Ghalandari, F. Jahn, N. De Keizer, and A. Winter, "An ontology for describing health IT interventions: Methodological considerations," in *Studies in Health Technology and Informatics*, 2019.

[27] E. Katis, H. Kondylakis, G. Agathangelos, and K. Vassilakis, "Developing an ontology for curriculum and syllabus," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[28] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, 2017.

[29] P. Jain, P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh, "Ontology alignment for linked open data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010.

[30] S. Staab, M. Erdmann, A. Maedche, and S. Decker, "An Extensible Approach for Modeling Ontologies in RDF(S)," in *Knowledge Media in Healthcare*, 2011.

[31] S. Decker, S. Melnik, F. Van Harmelen, D. Fensel, M. Klein, J. Broekstra, M. Erdmann, and I. Horrocks, "Semantic Web: The roles of XML and RDF," *IEEE Internet Computing*, 2000.

[32] I. Kollia, B. Glimm, and I. Horrocks, "SPARQL query answering over OWL ontologies," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[33] J. B. Lamy, "Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies," *Artificial Intelligence in Medicine*, 2017.

[34] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, "A survey of ontology learning techniques and applications," 2018.

[35] Z. Zhang, P. Yu, H. C. Chang, S. K. Lau, C. Tao, N. Wang, M. Yin, and C. Deng, "Developing an ontology for representing the domain knowledge specific to non-pharmacological treatment for agitation in dementia," *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, 2020.

[36] K. Höffner, J. Lehmann, and R. Usbeck, "CubeQA—question answering on RDF data cubes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.

[37] J. Zheng, M. R. Harris, A. M. Masci, Y. Lin, A. Hero, B. Smith, and Y. He, "The Ontology of Biological and Clinical Statistics (OBCS) for standardized and reproducible statistical analysis," *Journal of Biomedical Semantics*, 2016.

[38] M. Atzori, G. M. Mazzeo, and C. Zaniolo, "Querying RDF Data Cubes through Natural Language [Discussion Paper]," in *CEUR Workshop Proceedings*, 2018.

[39] Z. Syed, A. Pädia, T. Finin, L. Mathews, and A. Joshi, "UCO: A Unified Cybersecurity Ontology," in *AAAI Workshop - Technical Report*, 2016.

[40] B. Lantow, "Ontometrics: Putting metrics into use for ontology evaluation," in *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, IC3K 2016, (Setubal, PRT), p. 186–191, SCITEPRESS - Science and Technology Publications, Lda, 2016.

[41] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "Knowedu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31553–31563, 2018.

[42] A. Hofmann, S. Perchani, J. Portisch, S. Hertling, and H. Paulheim, "DBkWik: Towards knowledge graph creation from thousands of wikis," in *CEUR Workshop Proceedings*, 2017.