VEO: an Ontology for CO2 Emissions from Vehicles

Conference Paper · September 2014

DOI: 10.13140/2.1.1701.5689

3 authors:

Bojan Najdenov
Ss. Cyril and Methodius University in Skopje

Milos Jovanovik
Ss. Cyril and Methodius University in Skopje

Dimitar Trajanov
Ss. Cyril and Methodius University in Skopje

VEO: an Ontology for CO_2 Emissions from Vehicles

Bojan Najdenov, Milos Jovanovik, Dimitar Trajanov

Faculty of Computer Science and Engineering
Ss. Cyril and Methodius University
Skopje, Republic of Macedonia

Abstract. The Linked Data best practices provide ways for easier data representation, while at the same time raise the quality of the information that comes with it. The idea behind these best practices is to interlink datasets from various sources which are distributed over different locations and publish the data in an open, machine-readable format so that it would be easier to retrieve and process it by software agents, thus providing opportunities that many new use-cases can be created, which otherwise would not be possible in isolated datasets. With this, the value of the data itself rises to a whole new level. Environmental care is one of the most important issues on a global level, which means that great effort and resources are being spent, to help researchers find new and innovative ways of preserving our world and also to raise awareness of the problem itself. CO_2 emissions from vehicles became a large problem in the past few decades, since the number of vehicles exponentially increases, and also people are becoming more mobile than ever, having to commute and travel on a regular basis. In this paper, we describe the process of transformation one-, two- and three-star data about CO_2 emissions from vehicles published by the European Environment Agency and various other sources, into five-star Linked Open Data. In addition to that, we developed the Vehicle Emissions Ontology (VEO) to be able to describe the transformed data. We also provide use-case scenarios to show the benefits of using the Linked Data and Open Data concepts in these fields, and provide a public SPARQL endpoint as an entry point for accessing and using the data.

Keywords: CO_2 emissions; Ontology; Vehicles; Open Data; Linked Data; RDF.

1 Introduction

Information and data play a significant role in our lives. Having the necessary information at your disposal in the right time is something that everyone can benefit from; businesses can grow and even lives can be saved. Working with big data has been a challenge in the past decade, mainly because the quantity of the information that the world has keeps expanding at an enormous rate. That implies that researchers are somehow becoming more interested to discover new ways of data storage, representation, visualization and querying.

The idea of the Open Data concept states that data should be published in an open, machine-readable way that is easy to be processed and worked with by software agents and machines, which gives additional weight and value to the data itself.

The Linked Data[1] concept on the other hand, as an idea which represents a concrete realization of the Semantic Web [1], imagines the Web as a network of data, not documents (web pages). The entity from the data represented as Linked Data has its own URI which represents its location as a resource on the Web, and can be used both as a global naming scheme and for access over the existing Web infrastructure. This, along with the use of RDF as a data representation format, allows for datasets to be interlinked with other data which are physically distributed. Having the data connected in this way, it can be queried with the SPARQL query language, which creates possibilities for developing new use-case scenarios, previously unavailable over isolated datasets. With this, the value and quality of the data itself is increased. The LOD Cloud[2] consists of data and their interconnections from various areas, published in the Linked Data format. There are many contributors in the Linked Open Data community and they are the ones responsible for the constant growth of the LOD Cloud [2].

Pollution, waste, emissions of different toxins, gasses, liquids, irrational use of energy, are one of the biggest issues that humanity has to deal with. There are many world agencies, initiatives, projects which aim to raise awareness of the huge problem that humanity has with the pollution of the environment and energy efficiency. Significant amount of money is being given to research that would help the process of solving this important problem. New standards and policies are being introduced; new air and water filters are implemented in factories, global transportation companies search for more green way of transport and construction companies are trying to build more environment-friendly homes [3].

2 Related Work

Having in mind how important ecology and environmental safety are on a global level and with that the datasets in these areas as well, we must say that there has not been much effort put into using the standards of the Semantic Web in ecology-related fields. However, there is a project which analyzes the challenges and opportunities of Open Data in ecology, pointing out the issues which arise due to the dispersed and heterogeneous nature of the data [4]. The authors suggest using the Linked Data standards as methods of interconnecting distributed data, arguing that this would create new possibilities for scientific studies.

Since the domain of our research is the emission from vehicles, we found that several projects which work with Open Data related to vehicles and their features exist. The Volkswagen Vehicles Ontology[3] is an ontology developed by the Volkswagen Company as a vehicles manufacturer, to serve as a vocabulary which describes the Volkswagen-specific features of vehicles. They define several classes and many prop-

[1] http://linkeddata.org/
[2] http://lod-cloud.net/
[3] http://www.volkswagen.co.uk/vocabularies/vvo/ns

erties such as marketing name, product code, fuel injection type, and luggage capacity with seats folded and so on. They also define a property which is of specific interest in our case: an emissions property, which quantitatively represents the CO_2 emissions which come from a vehicle, measured in g/km, a standard unit.

The Vehicle Sales Ontology[4] and the Car Options Ontology[5] are vocabularies which use and extend the Good Relations Ontology[6], mainly to describe properties of vehicles for commercial purposes and sales.

3 CO_2 Emissions Data from Vehicles

In this section, we will describe the datasets from the domain that we acquired, and will analyze their content, format, and type of information that is provided.

3.1 Open Data from the European Environment Agency

The European Environment Agency (EEA[7]) is a European Union agency and its major goal is to provide reliable information concerning the environment. EEA has 33 member countries and they represent a valuable information source for entities which evaluate and develop environmental policies. EEA's major clients are the European Union institutions, their member countries, as well as independent researchers. EEA empowers institutions and individuals to communicate and cooperate with them, making sure that the provided information is satisfying their needs.

The agency publishes data on their website and it can be accessed and downloaded in CSV and MDB formats. EEA provides access to various datasets containing different sort of information such as sea surface temperature, sulfur dioxide emissions, PM2.5 and PM10 values, projections and interpolated maps, etc. The dataset we are most interested in is related to the CO_2 emissions from passenger cars, so we used it as our data source.

3.2 Data from the American Bus Association

The American Bus Association (ABA[8]) is an organization which represents about 1000 tour companies from the United States and Canada, and its members operate charter, tour, regular tour, transit and other special services. Their mission lies into meeting the transportation and travel needs of the public.

Founded in 1926, ABA's main goals are to accomplish industry's policy, promote public awareness about the use of motorcoach travel, provide economic benefits for its members, as well as conduct research about the industry and the market. One of

[4] http://www.heppnetz.de/ontologies/vso/ns#
[5] http://www.volkswagen.co.uk/vocabularies/coo/ns
[6] http://www.heppnetz.de/projects/goodrelations/
[7] http://www.eea.europa.eu/
[8] http://www.buses.org

their research projects is related to the comparison of CO_2 emissions from different transportation modes [5], and contains information about energy use and CO_2 emissions from several transportation means. We extracted the data about busses and trains that they provided in a table from a PDF document and used it to generate a CSV file as a data source.

3.3 Aircraft Emissions Data

Another dataset that we acquired is used in a research about emissions that come from aircraft and focuses on different methodologies of measuring and modeling distribution of the emissions, as well as providing comparison between them and suggesting best practices [6]. The information we are most interested in is the CO_2 emissions for a specific aircraft type. Besides that data, additional information is given at the Reference Manual, IPCC Guidelines on National Greenhouse Gas Inventories [7].

Similar to the previous data sources, we gathered the CO_2 emissions data from a table in the published PDF document and transformed it into a CSV file.

4 Ontologies

In order to be able to transform the obtained one-star and three-star data into five-star Linked Open Data, according to the Data Star Rating System[9], we need to provide ontologies that would describe our entities. The most common and recommended practice is to reuse existing ontologies, their classes and properties and map data using them, but also create new or extend existing ontologies with properties which are not part of any ontology yet. To do so, we explored the domain and reused several ontologies which we describe in this section, but also developed our own ontology, the Vehicle Emissions Ontology (VEO), in order to be able to describe the properties from our datasets.

Table 1. Ontologies reused for the EEA Dataset.

Ontology	Prefix	URI
Geo Names	gn	http://www.geonames.org/ontology#
Good Relations	gr	http://purl.org/goodrelations/v1#
Volkswagen Vehicle Ontology	vvo	http://purl.org/vvo/ns#
Vehicle Sales Ontology	vso	http://purl.org/vso/ns#

4.1 Reused Ontologies

The dataset we obtained from EEA contains information about specific types of passenger vehicles, such as manufacturer name, commercial name, engine capacity, wheel base width, etc., but the information that we are specifically interested in are

[9] http://5stardata.info/

the CO_2 emissions per vehicle model. We reused several ontologies to be able to semantically annotate the data, as shown in Table 1. Table 2 shows the different properties we used from these ontologies. A single entry from the EEA dataset represents a single automobile and we annotate it with the Automobile class from the Vehicle Sales Ontology.

Table 2. Properties reused for the EEA Dataset.

Property	Description
gn:countryCode	A two letters country code in the ISO 3166 list.
gr:hasManufacturer	Represents the manufacturer of the vehicle.
gr:hasMakeAndModel	The name of the model of the vehicle.
vvo:marketingName	The commercial name of the vehicle.
vvo:emissions	The CO2 emissions from the vehicle measured in g/km.
vso:weight	The weight of the empty vehicle ("curb weight" for cars), i.e. with standard equipment.
vso:wheelbase	The distance between the centers of the front and rear wheels.
vso:fuelType	The type of fuel suitable for the engine or engines of the vehicle.
vso:enginePower	The power of the vehicle's engine in KWT (kilowatts).

Regarding the datasets we acquired about buses, trains and airplanes and their CO_2 emissions, we reuse the properties shown in Table 3. A single entity from the buses dataset is annotated with the vso:BusOrCoach class. Respectively, an entry from the trains dataset is mapped with the vso:Train class. Finally, an airplane entity is associated with the peo:AirplaneModel class from the Proton Extended Ontology[10].

Table 3. Properties reused for the Bus, Train and Airplane Datasets.

Property	Description
gr:hasMakeAndModel	The name of the model of the vehicle.
vvo:emissions	The CO2 emissions from the vehicle measured in g/km.

4.2 Vehicle Emissions Ontology

We developed the Vehicle Emissions Ontology (VEO) in order to be able to semantically annotate and describe the datasets that we acquired, i.e. to provide definition for the classes and properties which the other ontologies did not.

In the VEO ontology we introduce the class veo:Train, which is a subclass of the vso:Vehicle class, and represents a train entity in general, as a mean of transportation. The datatype properties we introduce in VEO are shown and described in Table 4.

Besides using the data from the datasets, we created two properties – veo:numberOfPassengers and veo:emissionsPerPassenger – used for extending the

[10] http://www.ontotext.com/proton-ontology

dataset with information about the number of passengers in a vehicle and a calculation about the average CO_2 emissions per passenger. We added this new data in order to provide better comparison methods in various use-case scenarios.

Table 4. Datatype Properties from the Vehicle Emissions Ontology.

Property	Description
veo:id	Local entity ID in the dataset.
veo:manufacturerHarmonised	Harmonized name of the manufacturer
veo:manufacturerNameAsInMSRegistry	Manufacturer name as in the Member State Registry of EEA.
veo:typeApprovalNumber	Type of approval number as in the dataset from EEA.
veo:type	Internal vehicle type from the EEA dataset.
veo:variant	Local variant id from the EEA dataset.
veo:version	Internal version number in the EEA dataset.
veo:categoryOfTheVehicleTypeApproved	Internal category of the vehicle approved in the EEA dataset.
veo:totalNewRegistrations	Number of total new registrations of a vehicle type.
veo:axleWidthSteeringAxle	Steering axle width of the vehicle measured in millimeters.
veo:axleWidthOtherAxle	Other axle width of the vehicle measured in millimeters.
veo:fuelMode	Fuel mode of the vehicle.
veo:engineCapacity	Engine capacity of the vehicle measured in cm3.
veo:electricEnergyConsumption	Electric energy consumption of the vehicle, measured in Wh/km.
veo:innovativeTechnology	Type of innovative technology implemented in the vehicle, related to energy savings and emissions reduction.
veo:emissionsReductionThroughInnovativeTechnologies	Reduction of the emissions due to the innovative technologies implemented in the vehicle, measured in g/km.
veo:manufacturerPooling	Name of the manufacturer pooling.

The Vehicle Emissions Ontology is published[11] and accessible according to the Linked Data principles. This provides an opportunity for the ontology to be further extended and reused by others.

[11] http://purl.org/net/veo#

5 Creating Linked Data

Our goal in this paper is to provide five-star Linked Data by transforming the one-, two- and three-star data that we acquired and also provide a public SPARQL endpoint for accessing the transformed data. In this section we will explain the transformation process. The stages of this process are shown in Fig. 1.

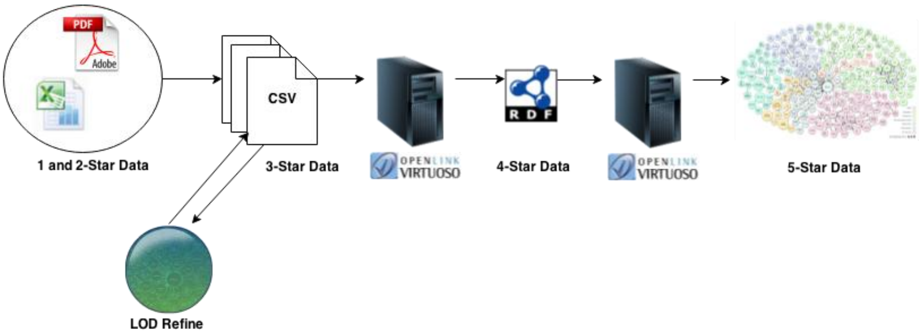

Fig. 1. Stages of the transformation process.

5.1 Extending the CSV data with DBpedia Entities

For the purpose of interlinking the vehicle entities from our datasets with resources from DBpedia[12] and the LOD Cloud, we used the LODRefine[13] tool. It allows reconciliation of data against DBpedia, i.e. detects and links DBpedia entities related to data in CSV files. We used it to add an extra column with URIs of DBpedia Automobile resources in our CSV datasets. With this, we created the basis for linking an Automobile resource from a CSV row with the LOD Cloud entity representing the same real-world object. We use the 'rdfs:seeAlso' property during the mapping, thus providing a relation between vehicle entities from our dataset and the LOD Cloud.

5.2 Mapping the Data from CSV to RDF

In order to transform and publish the data as Linked Data, we use the OpenLink Virtuoso Universal Server[14]. Virtuoso is a tool that provides views and management capabilities over different types of databases, but also has mechanisms that can facilitate the transformation process from CSV to RDF and enable a public SPARQL endpoint that can be used as a REST service.

As described in [8], we follow the same procedure for the transformation of the data. Once the CSV files are imported in Virtuoso and the tables are created, we trans-

[12] http://dbpedia.org/
[13] http://code.zemanta.com/sparkica/
[14] http://virtuoso.openlinksw.com/

form the relational databases into RDF views using the R2RML mapping language[15] that does the binding. Conceptually speaking, after that, each entry from the tables is represented as a resource that has its own URI, having all the properties that can be found in the columns of the table associated to it. Finally, we loaded all the triples we created from the four tables we had in the beginning – data about cars, buses, trains and airplanes – into a single RDF graph.

6 Use-Cases

6.1 Querying Data from our RDF Dataset

We introduce a scenario where we would like to get the information regarding CO_2 emissions for the cars manufactured by the 'Mazda' company, get their top 5 cars with most CO_2 emissions and display the details. The SPARQL query for this use-case is given below, and the results are shown in:

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix veo: <http://purl.org/net/veo#>
prefix vvo: <http://purl.org/vvo/ns#>

select distinct ?man ?cName ?em ?pass ?emPerPass
from <http://purl.org/net/veo/data#>
where { ?veoCar rdf:type vso:Automobile;
               veo:hasMSRegManufacturerName ?man;
               vvo:marketingName ?cName;
               vvo:emissions ?em;
               veo:numberOfPassengers ?pass;
               veo:emissionsPerPassenger ?emPerPass.
         filter ((lcase(?manufacturer))='mazda')
} order by desc (?emissions) limit 5
```

Table 5. Result from the SPARQL query.

Manufacturer	Model	CO_2 emissions (g/m)	Passengers	Emissions per Passenger
MAZDA	MAZDA CX-7	243	5	48,6
MAZDA	CX-7	243	5	48,6
MAZDA	MAZDA3/SP/2.3I/MPS	224	5	44,8
MAZDA	MAZDA3/SP/2.3I/MPS SPEZIAL	224	5	44,8
MAZDA	3 MPS	224	5	44,8

[15] http://www.w3.org/TR/r2rml/

6.2 Using Additional Data from DBpedia

In the following scenario we demonstrate the use of data and information from DBpedia, about the car Audi Q7 from our dataset. This type of queries is enabled by the relations we added in our dataset which point to the DBpedia dataset from the LOD Cloud. With the use of SPARQL federation, we can query the DBpedia SPARQL endpoint from our query, as well. The query is given below, and the results can be seen in Table 6.

```
prefix dbpedia-owl: <http://dbpedia.org/ontology/>
prefix dbpprop: <http://dbpedia.org/property/>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix vvo: <http://purl.org/vvo/ns#>

select ?abstract ?links ?photos where {
  graph <http://purl.org/net/veo/data#> {
    ?veoCar vvo:marketingName 'AUDI Q7'; rdfs:seeAlso ?DBcar.
  }
  service <http://dbpedia.org/sparql> {
    ?DBcar dbpedia-owl:abstract ?abstract;
           dbpedia-owl:wikiPageExternalLink ?links;
           dbpprop:hasPhotoCollection ?photos.
    filter langMatches( lang(?abstract), 'en')
  }
} limit 1
```

Table 6. Results from the query.

Abstract	External links	Photos
"The Audi Q7 is a full-size luxury crossover SUV unveiled in September 2005 at the Frankfurt Motor Show. Production of the Q7 began in autumn of 2005 in Bratislava, Slovakia…"	http://www.audi.co.uk/audi/uk/en2/new_cars/q7.html	http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/Audi_Q7

7 Conclusion

The Linked Data concept provides innovative ways of publishing and interlinking data from different remote sources and with it, provides a whole new level of usage possibilities, by both businesses and independent developers, for developing innovative applications and services. The opportunities that lie with the creation of new use-case scenarios are a field whose potential is becoming increasingly recognized [9].

CO_2 emissions on the other hand, as well as environmental care are one of the most important global issues. By interlinking the data about CO_2 emissions from vehicles, with data about the vehicles themselves from DBpedia, we show a new way of thinking and using relevant data to both car manufacturing companies and the end users as well. If manufacturers were to follow this standard of publishing their data as we are suggesting, it would help customers get more detailed info about vehicles that they are about to buy, and also help them choose a more environmentally friendly way of transportation.

In this paper we gave an overview of the transformation process of one-, two- and three-star data into five-star Linked Open Data. With this, we published data about CO_2 emissions from various vehicles and interlinked it with DBpedia, and also provided sample use-cases in order to demonstrate the benefits of using the Linked Data principles in this field. We hope that this effort serves as a motivation to car manufacturers to publish five-star Linked Open Data and use these ontologies to annotate data related to CO_2 emissions.

Acknowledgment. The work in this paper was partially financed by the Faculty of Computer Science and Engineering, at the Ss. Cyril and Methodius University in Skopje, as part of the research project "Semantic Sky 2.0: Enterprise Knowledge Management".

References

1. T. Berners-Lee, N. Shadbolt, "There's gold to be mined from all our data", The Times, 2012.
2. C. Bizer, T. Heath, T. Berners-Lee, "Linked Data - The Story So Far" International Journal on Semantic Web and Information Systems (IJSWIS), 2009, pp: 1-22.
3. Gradus, Raymond, Sjak Smulders, "The trade-off between environmental care and long-term growth—pollution in three prototype growth models", Journal of Economics 58.1 (1993): 25-51.
4. Reichman, O. J., Matthew B. Jones, Mark P. Schildhauer, "Challenges and opportunities of open data in ecology", Science (Washington) 331.6018 (2011): 703-705.
5. M.J. Bradley & Associates, "Comparison of energy use & CO2 emissions from different transportation modes", American Bus Association Foundation, May 2007.
6. Rypdal Kristin, "Good practice guidance and uncertainty management in national greenhouse gas inventories", (2000).
7. Change, Intergovernmental Panel On Climate, "2006 IPCC Guidelines for National Greenhouse Gas Inventories", Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan (2006).
8. Jovanovik, M., Najdenov, B., Trajanov, D.: "Linked Open Drug Data from the Health Insurance Fund of Macedonia", 10th Conference for Informatics and Information Technology (CIIT), 2013.
9. Kundra, V.: "Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect". Joan Shorenstein Center on the Press, Politics and Public Policy.