

Linking Data in the Insurance Sector: A Case Study

Katerina Kalou and Dimitrios Koutsomitropoulos

HPCLab, Computer Engineering and Informatics Dpt., University of Patras, Building B,
26500 Patras-Rio, Greece

{kaloukat, kotsomit}@hpclab.ceid.upatras.gr

Abstract. Enterprise data model (EDM) has been designated as a well-established approach in order to define a common way for communication and interoperability within an organization and across the industry. In recent years, business areas can also profit by the concept of Big Data, thus achieving effective management of massive amount of data. In this paper, we propose an example of how a data model can be enriched with the power of semantic web technologies, by converting it into an ontology. In this way, the data within the enterprise can be efficiently integrated into the Linked Open Data (LOD) cloud. The ultimate aim of this attempt is to set the infrastructure for exploring and linking Big Data in order to derive new value by exploiting Linked Data. We have selected to work on the insurance sector and apply the aforementioned ideas to the Property and Casualty data model, proposed by OMG.

Keywords: insurance, property, casualty, OMG, ontology, linked data, big data.

1 Introduction

An enterprise data model (EDM) captures the semantics of an organization for the purpose of communicating with the business community therefore, it can be considered a kind of ontology [3]. The components of an ontology, such as rules, classes, properties and individuals, can effectively reflect entities and activities described by the enterprise data model. Both an ontology and a data model can contribute to the meaning of data and to facilitate interoperability, integration, and adaptability of the informational systems that adopt them. Even though an EDM encapsulates the notion of ontology and shares common entities, it lacks the power of fine-grained expressiveness, reasoning and knowledge-based inference capabilities.

In recent years, Big Data [9] is used to describe all the new technologies that allow the manipulation of massive amounts of data in order to develop existing business areas or to drive new ones by improving on insight, decision-making and detecting sources of profit. Big Data can be thoroughly described by the following characteristics: Variety (the number of types of data), Velocity (the speed of data generation and processing), Volume (the amount of data) and Value (the knowledge).

In parallel of Big Data evolution, Linked Data [2], one of the most widely-used Semantic technologies, constitutes a means to evolve the Web into a global data space. Linked Data was introduced as a set of best practices for publishing and inter-linking structured data on the Web by using RDF as the data model and HTTP as the

protocol. Even though Linked Data gains its success as a technology by fostering data publication in the WWW, it can also be utilized as an architectural style for integrating either applications or data within the Enterprise.

In the case of Big Data, the Linked Data principles can contribute effectively to the aspect of Variety [10] by simplifying data management. For example, data coming from unstructured databases can be linked to data in traditional OLTP and OLAP stores without changing existing schemas.

The interlinking of data following different models (OLAP/OLTP databases, documents, NoSQL etc.) allows more data to be linked and then queried more effectively. So, the context of data expands by achieving more insight as well as better data analysis. The usage of Linked Data in combination with Big Data supports and boosts the fourth V, the Value, too [10].

In this work, we make an attempt to gather all the aforementioned aspects, which can leverage the growth of an industry, into one unified approach. Beginning from an EDM, we move to a more semantically enhanced model, an ontology, thus developing the foundation for more efficient consumption of Big Data.

More precisely, we exploit the existing data model for Property and Casualty [6], inspired and proposed by the Object Management Group, in order to build an ontology. This ontology meets the needs of and could also be used as the basis for publishing Linked Data in the industry of insurance. The interlinking of data ensures more interoperability as well as sets the infrastructure for adopting Big Data in the Insurance sector.

The usage of an ontology to express the P&C Model as well as any other EDM in the industry may have the following benefits:

- Enhance the existing expressive power of the EDM
- Achieve interoperability and extendibility of higher level
- Support more efficiently the possibilities of scalability
- Accomplish knowledge inference and discovery by exploiting the reasoning power
- Bridge the gap between any other subsystems (different kinds of data storage, databases, documents etc.) in the industry without negating the benefits of conventional stores or NoSQL databases

The rest of this paper is organized as follows. In Section 2, we start by providing some broad definitions and discussing the concepts of OMG's Property and Casualty Model. Furthermore, in Section 3, we explain in detail the entire process of converting the terms of data model into ontology components. Next, Section 4 outlines an indicative application scenario in order to illustrate the ontology. Finally, Section 5 summarizes our conclusions.

2 OMG's Property and Casualty Model

The Object Management Group (OMG) [13] has developed a property-casualty insurance enterprise data model with the contribution of representatives from many carriers

and using experts in each subject area. The P&C data model [6] addresses the data management needs of the Property and Casualty (P&C) insurance community. OMG's Model Driven Architecture principles and related standards have been utilized in order to construct the data model. Also, existing P&C industry standards (e.g., IBM's IAA) have been picked out as a source for the proposed P&C Business Glossary and associated models.

There are plenty of benefits for the organizations that decide to incorporate this data model into their procedures. P&C data model can provide significant value during a wide variety of warehousing, business intelligence, and data migration projects. In addition, these models can be leveraged to improve the use of data and in support of data governance within the enterprise.

Among the components of the P&C data model, there are the entities, attributes and relationships. Due to lack of space, only a snippet of all these is presented in this work. The main objects of the data model are the *Entities*. An Entity represents a person, organization, place, thing, or concept of interest to the enterprise. Next, an Entity can express a very broad and varied set of instance data, or it can be very specific. This is referred to as *levels of abstraction*. Some entities that have been used for the application scenario described in Section 4 are listed below.

- **Insurable Object:** An Insurable Object is an item which may be included or excluded from an insurance coverage or policy, either as the object for which possible damages of the object or loss of the object is insured, or as the object for which damages caused by the object are insured. Examples: residence, vehicle, class of employees.
- **Vehicle:** A Vehicle is an Insured Object that is a conveyance for transporting people and/or goods.
- **Person:** Person can be a human being, either alive or dead.
- **Organization:** An Organization is a Party that is a business concern or a group of individuals that are systematically bound by a common purpose. Organizations may be legal entities in their own right. Examples: commercial organizations such as limited companies, publicly quoted multinationals, subsidiaries.
- **Agreement:** Agreement is language that defines the terms and conditions of a legally binding contract among the identified parties, ordinarily leading to a contract. Examples; policy, reinsurance agreement, staff agreement.
- **Claim:** Claim is a request for indemnification by an insurance company for loss incurred from an insured peril or hazard.
- **Claim Amount:** Claim Amount is the money being paid or collected for settling a claim and paying the claimants, reinsurers, other insurers, and other interested parties. Claim amounts are classified by various attributes.
- **Claim Folder:** A Claim Folder is the physical file in which all claim documents are maintained.
- **Claim Document:** Claim Document is written information such as agreements, financial statements, offers, proposals, etc., to provide backup and depth to agreed-upon or discussed claim matters.

- **Claim Offer:** A Claim Offer is a proposal made by an insurer to a claimant or third party in order to settle a claim.
- **Party Role:** Party Role defines how a Party relates to another entity. Role is the definition of what a Person, Organization, or Grouping DOES whereas Relationship is about what a party IS.
- **Policy Coverage Detail:** Policy Coverage Detail defines the coverages included in an insurance policy (refer to Coverage definition). It is a key link in the data model among Policy, Claim, and Reinsurance.
- **Assessment:** Assessment is a formal evaluation of a particular person, organization, or thing that enables business decisions to be made. Examples: Credit Score, Underwriting Evaluation, Medical Condition.
- **Assessment Result:** An Assessment Result is the outcome of the assessment. There may be multiple results for one assessment.
- **Fraud Assessment:** Fraud Assessment is a subtype of Assessment Result that identifies whether fraud has occurred or is occurring, the extent and impact of the fraud, and to determine who is responsible for the fraud.

A *Relationship* should be a verb or a verb phrase that is always established between two entities, the 'Parent Entity' and the 'Child Entity'. The relationship is used in order to form a sentence between its two entities. Moreover, the type of relationship can be 'Identifying', 'Non-identifying' or 'Subtype'. Each relationship can be also characterized by the notion of cardinality. For example, the 'Subtype' relationship has always 'Zero-to-one' cardinality. Table 1 presents the majority of the relationships between the entities that are described above and utilized in the usage scenario of Section 4.

Table 1. P&C relationships definitions for the entities of Section 4

Parent Entity Name	Parent to Child Phrase	Relationship Cardinality	Child Entity Name
Insurable Object	is covered as defined in	One-to-Zero-One-or-More	Policy Coverage Detail
Insurable Object	involved in	Zero-or-One-to-Zero-One-or-More	Claim
Assessment	results in	Zero-or-One-to-Zero-One-or-More	Assessment Result
Claim	documented in	One-to-Zero-One-or-More	Claim Folder
Claim	settlement results in	One-to-Zero-One-or-More	Claim Amount
Claim Folder	contains	One-to-Zero-One-or-More	Claim Offer
Claim Folder	contains	One-to-Zero-One-or-More	Claim Folder Document
Claim Offer	results in	Zero-or-One-to-Zero-One-or-More	Claim Amount

Attributes are usually defined within an entity and is considered as a property or descriptor of this entity. An attribute is meaningless by itself. For example, date of birth comes from the context of the entity to which it is assigned, for example, date of birth of an employee. Every attribute in a data model is connected to a domain that provides for consistent names, data types, lengths, value sets, and validity rules. The main elements in order to define an attribute for an entity are the Attribute Name, Entity Name and Data Type. Table 2 presents some of the attributes for the entities *Insurable Object* and *Vehicle*.

Table 2. Attributes for the entities *Insurable Object* and *Vehicle*

Entity Name	Attribute Name	Data Type
Insurable Object	Insurable Object Identifier	INTEGER
	Insurable Object Type Code	VARCHAR(20)
	Geographic Location Identifier	INTEGER
Vehicle	Insurable Object Identifier	INTEGER
	Vehicle Model Year	NUMBER(4)
	Vehicle Model Name	VARCHAR(40)
	Vehicle Driving Wheel Quantity	INTEGER
	Vehicle Make Name	VARCHAR(40)
	Vehicle Identification Number	VARCHAR(100)

Figure 1 illustrates how *Insurable Object*, *Vehicle*, *Claim* and *Claim Folder* entities can be defined in terms of attribute and can be linked to each other via relationships in the context of P&C data model.

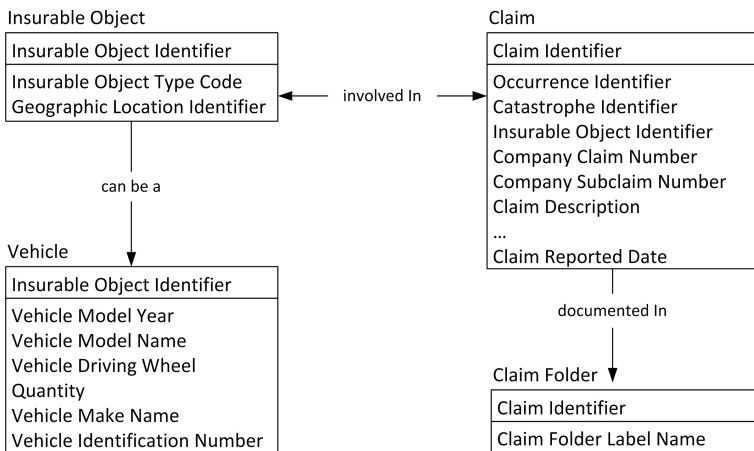


Fig. 1. A snippet of the P&C data model

3 P&C Model as Ontology

In this section, we describe how the Property and Casualty model can be expressed in terms of an ontology using the Web Ontology Language (OWL) [1]. Having the conceptual data model as a reference, we present how data entities representing most of P&C insurance business processes can be converted into ontology components. Note that the reverse is also feasible, i.e. the logical models can be generated from OWL ontologies. According to InfoSphere Data Architect guidelines [11], we can simply correlate logical data model's data types to OWL data types, logical data model's entities to ontology elements as well as logical data model's relationships to OWL object properties.

In the attempt to design our P&C ontology, we first convert all entities (see Section 2), included in the P&C Data Model, to ontology classes. In detail, the convention adopts the restrictions that are presented in Table 3.

Table 3. Logical data model to OWL object mappings

Logical Data model	OWL ontology elements
Entity	owl:Class
Entity - Name	rdf:about
Entity - Label	rdfs:label
Entity - Namespace	Namespace of owl:Class
Entity - Definition	rdfs:comment
Entity - Primary Key	owl:hasKey
Relationship	owl:ObjectProperty
Relationship – Name	rdf:about
Relationship – Label	rdfs:label
Relationship – Owner	rdfs:domain
Relationship – Namespace	Namespace of owl:ObjectProperty
Relationship – Child Entity	rdfs:domain
Relationship – Parent Entity	rdfs:range
Relationship – Annotation	rdfs:seeAlso rdfs:DefinedBy owl:FunctionalProperty owl:InverseFunctionalProperty owl:AnnotationProperty
Attribute	owl:DatatypeProperty
Attribute – Name	rdf:about
Attribute – Label	rdfs:label
Attribute – Namespace	rdfs:domain
Attribute – Data Type	rdfs:range
Attribute – Length/Precision	xsd:maxLength xsd:length xsd:totalDigits
Attribute – Scale	xsd:fractionDigits
Attribute – Primary Key	owl:hasKey
Attribute – Documentation	rdfs:comment
Attribute – Annotation	rdfs:seeAlso rdfs:isDefinedBy owl:AnnotationProperty

All 'Subtype' relationships (See section 2) have been utilized to form a class hierarchy. So, the 'Parent Entities' are translated into superclasses, while the 'Child Entities' into subclasses. Table 4 provides a listing of a part of 'Subtype' relationships included in the P&C Data Model which relates the entity 'Insurable Object' to any entity that can be characterized as 'Vehicle'.

Table 4. Logical data model 'can-be-a' relationship

Parent Entity Name	Parent to Child Phrase	Relationship Cardinality	Child Entity Name
Insurable Object			Vehicle
Vehicle			Automobile
Vehicle			Van
Vehicle			Trailer
Vehicle			Truck
Vehicle	can be a	Zero-to-One	Boat
Vehicle			Bus
Vehicle			RecreationVehicle
Vehicle			Watercraft
Vehicle			Motorcycle
Vehicle			ConstructionVehicle

Figure 2 depicts how the aforementioned relationships can be converted into components of the P&C ontology.

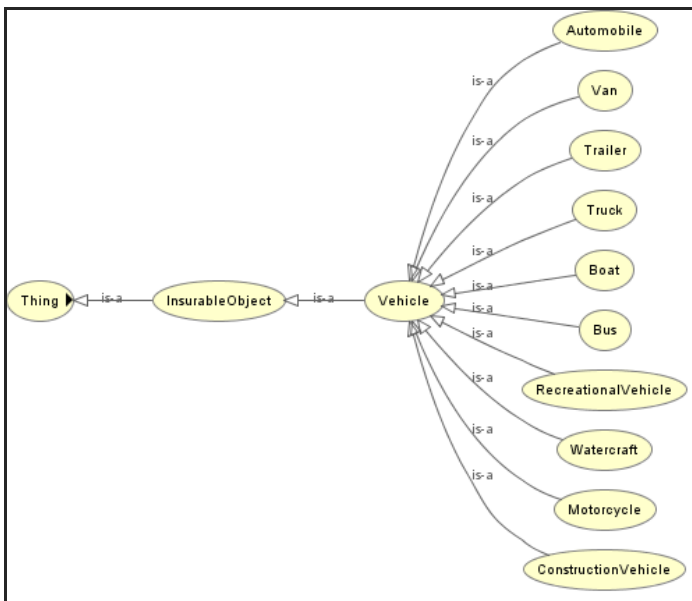


Fig. 2. The mapping of 'Insurable Object' entities in the P&C ontology

All the other types of relationships are represented as object properties in the proposed ontology. Table 3 gathers all the restrictions that must be followed for an effective conversion of these relationships. Two sample relationships, expressed in OWL, are listed below using RDF/XML syntax.

```
<owl:ObjectProperty
rdf:about="http://www.levantes.ceid.upatras.gr/OMG_Insurance#resultsIn">
  <rdfs:range rdf:resource="http://
www.levantes.ceid.upatras.gr/OMG_Insurance #AssesmentRe-
sult"/>
  <rdfs:domain rdf:resource="http://
www.levantes.ceid.upatras.gr/OMG_Insurance #Assessment"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="http://
www.levantes.ceid.upatras.gr/OMG_Insurance #documentedIn">
  <rdfs:domain rdf:resource="http://
www.levantes.ceid.upatras.gr/OMG_Insurance#Claim"/>
  <rdfs:range rdf:resource="http://
www.levantes.ceid.upatras.gr
/OMG_Insurance#ClaimFolder"/>
</owl:ObjectProperty>
```

Each entity type will have one or more data attributes. Each attribute of the P&C model has been mapped to a datatype property in the ontology, according to the restrictions in Table 3.

4 Usage Scenario for Vehicle

A usage scenario (in natural language), relying on the OMG model vocabulary and the implemented ontology, is cited in this subsection. The scenario deals with the case of vehicle insurance and examines the procedure of a fraud assessment. All the nouns, which are in italics, reflect instances of ontology classes. Figure 3 illustrates more thoroughly the usage scenario. Since this is only a snapshot of the model's full range, further refinements are possible.

A vehicle (*Vehicle_1*) has an insurance (*Agreement_1*). This vehicle has been insured for a time interval A with B amount and X coverage (*PolicyCoverageDetail_1*). The *Person_1*, having the role of *Insurer* and being employed by the Organization, *Orgabnization_1* has insured the *Vehicle_1* via the *Agreement_1*. The *Person_2* is owner of the *Vehicle_1*. At the same time, the *Person_2* has also another role, the role of *Driver*.

An accident with significant damages occurred. *Person_2*, the owner of *Vehicle_1*, is involved in a claim (*Claim_1*). The *Claim_1* is documented in the folder (*Claim-Folder_1*) which contains two documents (*ClaimDocument_1* and *ClaimDocu-ment_2*) as well as an offer (*ClaimOffer_1*). The *ClaimOffer_1* results in a specific amount, the *ClaimAmount_1*. An assessment (*Assesment_1*) has been done for the *Claim_1* and results in the *FraudAssesment_1*.

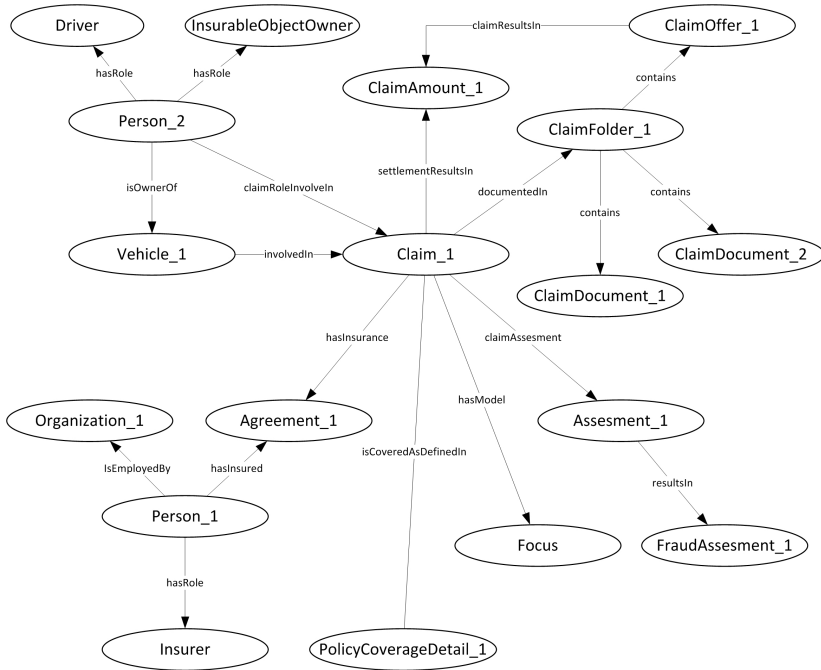


Fig. 3. Usage scenario for Insurance of Vehicle and Fraud Detection

Based on these, a series of business requirements can be represented in the form of logical rules or *axioms* that can be used to automatically infer possible added-value knowledge [5]. For example, a rule can be imposed stating that if a *Driver* has been *involved* in more than e.g. 5 insurance *claims* within a certain time interval, then he is susceptible to *FraudAssesment* to be performed on him.

5 Conclusion and Future Work

In this work, we have shown the feasibility and benefits of adopting a semantic approach towards Big Data manipulation in the insurance sector. To our knowledge, this is the first attempt to express the OMG P&C Model in the form of an OWL 2 ontology, so that reasoning-based conclusions can be enabled. As such, the added-value of adopting (even part of) the model at the insurance sector is leveraged, as it can be used towards automating standard legacy processes, such as fraud detection. Instead of manual resolution, an OWL 2 ontology can be used to highlight possible cases that need further investigation.

On the other hand, the specification of a relaxed ontology schema can serve as a scaffold for implementing Linked Data publishing and dissemination of insurance agreement and claim information. In this manner, the information flow between insurance companies, agents and insurers can be streamlined, by automatically transmitting messages through compatible services.

It is expected that contemporary insurance companies would have to deal with a great deal of such information, such resulting in an increased volume and velocity of data. It is our belief that Semantic Web technologies can stand up to these requirements, by establishing a scalable infrastructure around a semantic repository [12], [8] that secures performance without sacrificing the added-value inference capabilities inherently are equipped with.

References

1. Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., Mc Guinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL Web Ontology Language Reference, W3C Recommendation (2004), <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.* 5(3), 1–22 (2009)
3. Hay, C.H.: Data Modeling, RDF, & OWL - Part One: An Introduction To Ontologies (April 2006), <http://TDAN.com>
4. Heath, T., Bizer, B.: *Linked Data: Evolving the Web into a Global Data Space*, 1st edn. Synthesis Lectures on the Semantic Web: Theory and Technology, vol. 1, pp. 1–136. Morgan & Claypool (2011)
5. Hitzler, P., Parsia, B.: Ontologies and Rules. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, 2nd edn., pp. 111–132. Springer (2009)
6. Jenkins, W., Molnar, R., Wallman, B., Ford, T.: *Property and Casualty Data Model Specification* (2011)
7. Kim, H., Fox, M.S., Sengupta, A.: How To Build Enterprise Data Models To Achieve Compliance To Standards Or Regulatory Requirements (and share data). *Journal of the Association for Information Systems* 8(2), Article 5 (2007)
8. Koutsomitropoulos, A.D., Solomou, D.G., Pomonis, P., Aggelopoulos, P., Papatheodorou, T.S.: Developing Distributed Reasoning-Based Applications for the Semantic Web. In: *Proc. of the 24th IEEE Int. Conference on Advanced Information and Networking (AINA 2010)-Int. Symposium on Mining and Web (MAW 2010)*, WAINA, pp. 593–598 (2010)
9. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: *Big data: The next frontier for innovation, competition, and productivity*. Technical report, Mc Kinsey (2011)
10. Mitchell, I., Wilson, M.: *Linked Data: Connecting and exploiting big data*. White paper. Fujitsu UK (2012)
11. Soares, S.: *IBM InfoSphere: A Platform for Big Data Governance and Process Data Governance*. MC Press Online, LLC (February 2013)
12. Solomou, D.G., Kalou, A.K., Koutsomitropoulos, A.D., Papatheodorou, T.S.: A Mashup Personalization Service based on Semantic Web Rules and Linked Data. In: *Proc. of the 7th Int. Conference on Signal Image Technology and Internet Information Systems (SITIS 2011)*, pp. 89–96. IEEE (2011)
13. The Object Management Group (OMG). *MDA Guide Version 1.0.1*, omg/2003-06-01 edition (2003)