Open eBusiness Ontology Usage:
Investigating Community Implementation of GoodRelations

Jamshaid Ashraf
Digital Ecosystem and Business Intelligence Institute (DEBII), Curtin University of Technology, Perth, Australia
jamshaid.ashraf@gmail.com

Richard Cyganiak
Digital Enterprise Research Institute (DERI), National University of Ireland, Galway
richard@cyganiak.de

Sean O'Riain
Digital Enterprise Research Institute (DERI), National University of Ireland, Galway
sean.oriain@deri.org

Maja Hadzic
Digital Ecosystem and Business Intelligence Institute (DEBII), Curtin University of Technology, Perth, Australia
maja.hadzic@cbs.curtin.edu.au

ABSTRACT

The GoodRelations Ontology is experiencing the first stages of mainstream adoption, with its appeal to a range of enterprises as the eCommerce ontology of choice to promote its offerings and product catalogue. As adoption increases, so too does the need to critically review and analyze current implementation of the ontology to better assist future usage and uptake. To comprehensively understand the implementation approaches, usage patterns, instance data and model coverage, data was collected from 105 different web based sources that have published their business and product-related information using the GoodRelations Ontology. This paper analyses the ontology usage in terms of data instantiation, and conceptual coverage using a SPARQL queries to evaluate quality, usefulness and inference provisioning. Experimental results highlight that early publishers of structured eCommerce data benefit more due to structured data being more readily search engine indexable, but the lack of available product ontologies and product master datasheets is impeding the creation of a semantically interlinked eCommerce Web.

Categories and Subject Descriptors: D.2.5 [Software/Software Engineering]: Testing and Debugging

General Terms: Verification, Design

Keywords: GoodRelations, Instance data analysis, Business ontology, Structured eCommerce data, ontology usage.

1. INTRODUCTION

The Web of data and open ontologies (e.g. FOAF, SIOC, SKOS) have allowed the establishment of a shared understanding between data providers and consumers, in a common format that allows automated processing of information by software agents. Where accepted by the community, an ontology offers opportunity for enhanced information dissemination and commerce. The GoodRelations Ontology (GRO) [1], developed specifically for Web-based eCommerce, is an example of such an ontology that allows businesses describe their product offerings, entities and descriptions. The resulting semantically annotated structured data is then accessible for use in different Semantic Web applications and inclusion in search engine indexes. PingTheSemanticWeb.com[1] has ranked GRO in second position to FOAF as the most widely used ontology. Available since 2008, GROs' schema is mature but uptake reflects that of early adoption.

A review and analysis of the current community implementations of the GRO within its eCommerce environment is timely as it will provide insight into its applicability, conceptual coverage and actual usage within its application domain. This paper reports on the current implementation status of the GRO after investigating 105 publically available data sets. In this first large scale investigation of its kind into the GRO, data providers are categorized and dataset characteristics discussed and the usefulness of currently available data sets is analysed through different use cases in addition to implicit data available through axiomatic triples.

The remainder of the paper is organized as follows. Section 2 introduces the motivation, and the background of this research is discussed in Section 3. In Section 4, we discuss the dataset collection and its characteristics. Section 5 describes dataset investigation, use cases, along with result and observations and impact of reasoning. Related work is presented in Section 6 followed by the conclusion in Section 7.

2. MOTIVATION

The semantic web is providing a level of semantically annotated structured data that is enhancing the level of user experience by sourcing and identifying more accurately, information of interest. Enabled primarily through ontological alignment, semantic annotation is a major contributory factor in the increasing interest in ontology usage by the wider community and had also attracted the attention of early business adopters. Over the last two years the GRO has witnessed wide sectoral appeal and mainstream adoption by eRetailers, such as *BestBuy.com, Overstock.com* and *Oreilly.com*. Announcements from search engine providers Google[2] and from Yahoo[3] to index GRO will for its corporate users extend their consumer reach to a larger audience with their increased appearance in search results. A measure of any ontology popularity is its community acceptance, which in turn will reflect

LDOW2011, March 29, 2011, Hyderabad, India.

[1] Last accessed on Dec 16, 2010

[2] http://googlewebmastercentral.blogspot.com/2010/11/rich-snippets-for-shopping-sites.html

[3] http://developer.yahoo.com/searchmonkey/smguide/gr.html

‘some’ level of use but not the extent of adherence to the schema or the extent of instantiation. A more accurate look at popularity should therefore look at overall ontology population. To date however the literature does not provide evidence of systematic analysis of GRO usage that could provide insight into its adoption and usage status in the emerging eCommerce Web of data. [2] defines ontology population as having occurred when an ontological term (i.e. concept, property or individual) is *used* to annotate the data. An analysis of these terms usage within the GRO would be beneficial for:

eCommerce information producers and consumers: By providing insight into structured data usage as a means to improve the quality and quantity of data being made available to the business consumer.

Ontology engineering: With analysis of ontology population and model coverage to help ontology engineers understand usage pattern, better incorporating stakeholders’ perspectives in ontology evolution [3] and ontology maintenance.

Ontology Mapping: Interaction between different ontology concepts would benefit from understanding the models used and instance data generated. An analysis of the eCommerce Web of data landscape and use of ontologies [4] would be useful.

3. GoodRelations ONTOLOGY OVERVIEW

In the following section, we describe our high level categorization of data providers, brief overview of GR conceptual schema and use of GRO in search indexes.

3.1 Data Providers

Looking at the structured eCommerce data landscape, we can categorize users into three groups based on their publishing approach, usage pattern and data volume.

3.1.1 Large Size Retailers

This group includes large online e-retailers and the retailers who traditionally operate as brick and mortar. They have only recently entered into e-retailing business. Such data sources provide more detailed (rich) product description which is useful for entity consolidation and interlinking with other datasets. Such companies include, among others, BestBuy.com, Overstock.com, Oreilly.com, and Suitcase.com.

3.1.2 Web shops

A large number of the data sources, included in our dataset, comprises small to medium web shops, mainly offering their products and services through web channel only. Most of these web shops use web content management packages[4] such as Maganto, osCommerce, Joomla to add RDFa data in html pages. This approach is quite viable since, in most cases, no special infrastructure arrangement is required.

3.1.3 Data Service providers

To leverage the benefits offered by semantic eCommerce data, businesses are offering data services that build on consolidated semantic repositories. Such providers addditionally use APIs to access and transform proprietary data into RDF, and make them available through their repository. For example, Linked Open Commerce (LOC)[5] contains Amazon.com data despite that fact that Amazon.com has not yet published RDF/RDFa.

3.2 Conceptual Schema

The latest version[6] of GRO comprises 27 concepts (classes), 49 object properties, 43 data properties and 43 named individuals. Keeping backward compatibility intact, the ontology model was updated recently to add new object and data properties based on the experience and feedback gained through real-world implementations. GRO ontology is available at http://purl.org/goodrelations/v1 and **gr** is the prefix used in this paper and also in general practice elsewhere.

3.2.1 Axioms

The GRO comprised classes, properties, individuals and axioms. Axioms allow information to be inferred from a knowledge base through the use of a reasoning engine known as *reasoner*. The expressivity of GRO is based on OWL DLP fragment and contains subclass and subproperty axioms to express the subsumption behaviour in the model. Axiomatic triples in GRO are given in Table 1 to shed light on the possible inference on eCommerce data annotated using GRO and applicable rule sets. There are some RDFS and OWL elements available in the ontology such as rdfs:domain and rdfs:range, not mentioned in the table because they are not considered as part of the reasoning experiment.

Table 1: Axioms in GRO and applicable rule sets

	Axioms	Count	Applicable Rule sets
Class	SubClassOf	13	RDFS
	DisjointClasses	91	OWL2RL
Object Property	SubPropertyOf	4	RDFS
	InverseOf	6	pD*, OWL2RL
	TransitiveProperty	7	pD*, OWL2RL
	SymmetricProperty	2	pD*, OWL2RL
Data Property	SubPropertyOf	13	RDFS

Elements such as rdfs:subClassOf, rdfs:subPropertyOf, owl:inverseOf, owl:TransitiveProperty and owl:SymmetricProperty are all applicable in the same way by ‘entailing new knowledge’. They can be used in forward-chaining to materialize the implied statements and make them explicit , as well as backward-chaining which performs query rewrite to expand query scope and include inferred knowledge. owl.DisjointClasses is different from abovementioned constructs because it is primarily used for data quality and helps to check inconsistencies. Constructs mentioned in Table 1 are covered by almost all of the rule sets including RDFS, pD* [5] and OWL2RL[7]. In our investigation, we employed an RDFS-based reasoning engine with RDFS rules as we believe that this

[4] Complete list of their references are available at http://www.ebusiness-unibw.org/wiki/GoodRelations#Shop_Software

[5] http://www.linkedopencommerce.com

[6] Latest version and model used in our experiment last updated on Nov 26, 2010.

[7] http://www.w3.org/TR/owl2-profiles/

fragment is available in most semantic repositories and plausible for the web.

3.3 Use of GRO by Search Engines

Adoption drivers include the GRO in search engine indexes such as Google and Yahoo, which provide improved visibility of product offering and company web pages in their enhanced organic search result pages [6]. Our investigation found that Yahoo and Google currently includes price, availability (Google only), description and product pictures drawn from GRO annotated structured data as part of their enhanced search results.

4. DATA SET

Our research work requires a collection of eCommerce data from a maximum number of different data sources. The aim is to obtain a good collection of mixed data which are primarily annotated using GRO. Throughout this document we will use the following expressions:

GoodRelations Dataset (GRDS): RDF graph collected from the Web (websites) and stored in a triple store for querying and reasoning.

Data Source: Data source refers to the website (unique domain name server (DNS)) included in GRDS which contains eCommerce data in RDF (any serialization format) or RDFa format based on GRO model.

4.1 Data Set Collection

To analyse the adoption, usage patterns and uptake of GRO in general and by the eCommerce community specifically, we collected data from different data sources and generated GoodRelations Dataset (GRDS). Firstly, we identified the potential data sources which *used* GRO to describe offering or company (Business Entity) or both. Different semantic search engines such as Sindice[8] and Watson[9] which index RDF documents, were used to obtain a list of potential data sources. Traditional search engines such as Google were also used to retrieve RDF documents by using the *filetype:rdf* attribute of advanced search to access RDF documents over the web. Additionally, we also considered the data publisher's list maintained at the GoodRelations's developer wiki site[10]. For our empirical investigation, we collected data from 105 different data sources complying with the criteria mentioned above. The complete list is provided at the end of the document in the Appendix section.

During the data collection process, we noticed that 90% of the websites (data sources) are using the RDFa[11] standard to add structured information to already existing HTML documents. After identifying the potential sources, we evaluated different approaches to identify and retrieve documents which contain GR-marked structured information. Almost all of the sources have *sitemap.xml* files for search engines to crawl web pages and build indexes. However, the links (URLs) provided in the sitemap file are often linked to 'list of product' pages and not to the actual product pages. Since, we are interested in accessing the web pages that have RDFa code in them; we used crawlers to build the list of such URLs and manually verified the list of pages with RDFa. After obtaining the list of URLs, we used REST-based web services, Any23[12] and RDFa Distiller[13] to parse RDFa snippets from HTML documents online and generate RDF graphs (in RDF/XML syntax). We loaded these RDF graphs into OpenLink Virtuoso (Open-Source Edition[14]) triple store to conduct our investigation. From the RDF data management point of view, named graphs are used to group all the triples of one data source under a unique named graph URI. This allowed us to query the dataset vertically as well as horizontally.

Linked Open Commerce[15] (LOC) is an emerging data space which collates eCommerce data from the Web and makes information available for retrieval and viewing through SPARQL endpoint. Despite its presence, collection of data sets was hampered due to i) the unavailability of several data sources in the LOC and ii) the presence of several triples using non-authentic URIs, resulting in an inability to de-referenced the URI, use it in query or obtain provenance details. Roughly, there are 34 data sources, which have published their data in RDF/RDFa format and currently available in the LOC. LOC additionally contains a nominal number of data sources that are made available in RDF/RDFa through the use of middleware APIs' e.g. Amazon.com. Problematic invalid URIs such as those starting with "localhost.localdomain…." were also found[16] but represented minor data quality issue that were easily corrected.

The availability of these two datasets (i.e. LOC and GRDS) for our investigation has provided an optimal search space covering the maximum possible width (GRDS) and depth (LOC) of the structured eCommerce web-of-data. In other words, GRDS covers more data sources, whereas LOC covers more data from a particular data source. In this regard, both datasets complement each other and LOC is used to cross check or find additional information useful in analysis results.

4.2 Dataset Composition Characteristic

In the course of new dataset creation i.e. GRDS, we also looked at the different characteristics of datasets, such as use of different namespaces, use of GR vocabulary and use of annotation properties and described as follows.

4.2.1 Namespace Usage

Table 2 lists all vocabularies and their prefixes found in GRDS. Aside from *gr*, the top three most used vocabularies are *dc* (Dublin Core), *foaf* and *vCard*. There are few vocabularies, such as *vCard* and *dc* that are used with multiple prefixes and the reason for this is the availability of new version with new namespace URI. Quite a few focused vocabularies are used by data sources to annotate the data relevant to their businesses. For example, *frbr* is used by O'Reilly to annotate bibliographic data.

[8] http://www.sindice.com

[9] http://watson.kmi.open.ac.uk/WatsonWUI

[10] http://www.ebusiness-unibw.org/wiki/GoodRelations

[11] http://www.w3.org/TR/rdfa-syntax/

[12] http://any23.org

[13] http://www.w3.org/2007/08/pyRdfa/

[14] http://sourceforge.net/projects/virtuoso/

[15] http://linkedopencommerce.com

[16] Observations were made as of 16 OCT 2010

Table 2 : Prefixes and Namespaces used in GRDS

Prefix	Namespace URI	Data source s (%)
gr	http://purl.org/goodrelations/v1#	100
vCard	http://www.w3.org/2006/vcard/ns# http://www.w3.org/2001/vcard-rdf/3.0[17] (deprecated namespace)	88.57
dcterm dc	http://purl.org/dc/terms/ http://purl.org/dc/elements/1.1/	34.29
foaf	http://xmlns.com/foaf/0.1/	25.71
commerce[18] media use currency	http://search.yahoo.com/searchmonkey/commerce/ http://search.yahoo.com/searchmonkey/media/ http://search.yahoo.com/searchmonkey-datatype/use/ http://search.yahoo.com/searchmonkey-datatype/currency/	14.29
v[19]	http://rdf.data-vocabulary.org	3.81
og[20]	http://opengraphprotocol.org/schema/	0.95
rev[21]	http://purl.org/stuff/rev#	0.95
frbr[22]	http://vocab.org/frbr/core#	0.95
geo[23]	http://www.w3.org/2003/01/geo/wgs84_pos#	0.95

4.2.2 GR Vocabulary Usage

Here, we analyse and discuss the GR vocabulary usage by different implementers. The straightforward approach is to calculate the number of the instances each concept has in the dataset and calculate the properties used in implementation. Even though this approach can help identify the most and the least populated terms, it cannot help us to sufficiently understand the usage patterns across different data sources. For example, if one particular class (concept) is used by a large implementer (e.g. BestBuy.com) for their two hundred thousand plus products, then the count of instances of that class will be high. However, it is possible that only this implementer has used this concept in GRDS. Therefore, we provide the use of ontology terms based on the percentage of data sources that have used it and not on the total number of triples in the dataset.

In this study, we also analysed the GR schema usage from two perspectives. Firstly, in this section, we simply looked at the usage of concepts by looking at the instances and data sources found in GRDS. We believe that this provides a basic understanding about the nature of available data and the frequency of concept and/or property use by different data providers. However, statistical representation based on simple instance and data sources calculation does not provide an insight into the relationships that exist between entities and the implementation of the ontological model in practice. To offer this level of visibility, we have investigated the GR model usage by examining the conceptual coverage of three main pivotal concepts (Business Entity, Offering, Product or Service) and description richness available by exploring (traversing) the relationships available with other concepts through GR properties (see section 5).

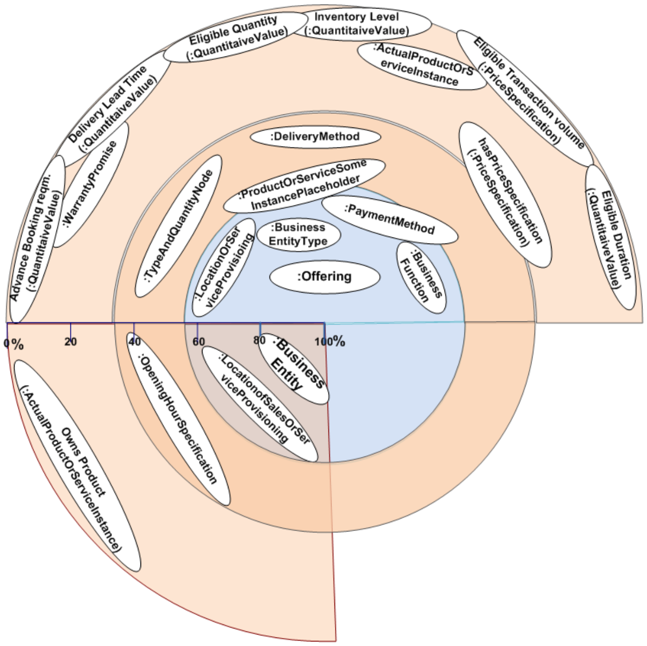

Eligible Quantity (:QuantitaiveValue)
Inventory Level (:QuantitaiveValue)
:ActualProductOrServiceInstance
Eligible Transaction volume (:PriceSpecification)
Delivery Lead Time (:QuantitaiveValue)
:DeliveryMethod
:WarrantyPromise
hasPriceSpecification (:PriceSpecification)
:ProductOrServiceSomeInstancePlaceholder
:TypeAndQuantityNode
:PaymentMethod
Advance Booking reqm. (:QuantitaiveValue)
:Business EntityType
Eligible Duration (:QuantitaiveValue)
:LocationOrServiceProvisioing
:Offering
:Business Function
0%
20
40
60
80
100%
:Business Entity
:OpeningHourSpecification
:LocationofSalesOrServiceProvisioning
Owns Product (:ActualProductOrServiceInstance)

Figure 1: GR Concepts related to two pivotal concepts :Offering and :BusinessEntity

Figure 1, places the concepts on the diagram based on the percentage of their use by different data sources. In this figure, concepts are shown in two groups; namely :Offering[24] and :BusinessEntity. The groups help in visualizing particular fragment of the data in specific context of use. We can see that several concepts are on the border of outermost circle. This indicates that several data sources have not provided fine grain information about their offering but only have made available the basic set of data such as eligible customer type, business function of offering like for selling, leasing or renting. In the lower half part of the Figure 1, we have concepts directly or indirectly linked with business entity (:BusinessEntity).

As shown, 60% of the data sources have provided information regarding their shops (office/branches) and 40% of the data source have further details available such as shops opening and closing time. One of the pivotal concepts i.e :ProductOrService is not shown in Figure 1. The reason for this is that there is no formal Product ontology currently used in GRDS. :ProductOrService sub concepts are being used through offering data to describe whether

[17] W3C has now RDF based vCard however 25.71% of data source are still using deprecated namespace

[18] Yahoo search monkey project defined these namespaces to provide vocabulary to assist developers.

[19] This is Google vocabulary published to be used for structured data with RDFa and microformat.

[20] Facebook Open Graph protocol. Only used by www.lovejoys-ltd.co.uk

[21] Vocabulary for expressing reviews and ratings. Only used by www.overstock.com

[22] Vocabulary for Functional Requirements for Bibliographic Records (FRBR). Only used by www.oreilly.com

[23] Vocabulary for latitude, longitude and altitude in the WGS84 geodetic reference datum. Only used by www.bestbuy.com

[24]Throughout this paper, we assume that http://purl.org/goodrelations/v1# is the default namespace and use prefixes mentioned in Table 2 for other namespaces

the product referred in offering is the actual instance or existentionaly quantified.

4.2.3 Use of annotation properties

The GRO document recommends the use of annotation properties to provide additional information about resources. Almost all the entities in GRDS are annotated with rdfs:label and rdfs:comment properties. We have realized that the usability of these properties is very high in current eCommerce deployment as these are frequently used in queries to retrieve resources of interest. For example, one of the instance of type :Offering has rdfs:label set to "*13 pieces of product "Cash Bases Cost Plus Flip Lid 460, weiss" are on stock*". One possible solution is to use "Lid 460" in the FILTER clause of the SPARQL query to limit the result set to potential candidate offers.

Table 3: Annotation properties use in GRDS

Property	Data sources (%)	Property	Data sources (%)
rdfs:label	94.29	dc:rights	0.95
rdfs:seeAlso	85.71	dc:contributor	0.0
rdfs:comment	84.76	dcterms:license	0.0
rdfs:isDefinedBy	60.95	owl:deprecated	0.0
dc:title	23.81	owl:versionInfo	0.0
dc:creator	22.86	:relatedWebService	0.0
dc:subject	1.90		

Table 3 summarises the use of annotation properties in GRDS. We can see that the majority of the data sources have provided textual descriptions useful for human consumption and user interfaces. GRO has one 'built-in' annotation property (i.e. :relatedWebService, not used by any data publisher in GRDS) to support Semantic Web services discovery and invocation services. This can be a very useful feature to enable automatic service discovery in digital ecosystems [7]. Annotation properties contain literal values and can optionally [8] have language tag (metadata) to explicitly specify the language in which text is written. In Table 4 we have summarized the use of language tags with rdfs:label and rdfs:commet literal values in GRDS.

Table 4: Language tags with literal value use in GRDS

Language tag	Data Sources (%)	Name
en (English)	72.81	
de (German)	8.74	
fr (France)	0.97	slindi.com
en-(Great Britain)	0.97	www.lovejoys-ltd.co.uk
pt (Portugal)	0.97	www.globalautoimports.com.br

As we can see, en[25] (IETF's BCP 47 code of English Language) is the most commonly used natural language for providing textual description of the resources.

5. ANALYSIS

One of the main purposes of making structured data available on the semantic web is to allow users to access accurate (exact) information [9]. The key to accessing exact information is the availability of a conceptual description based on the ontological model. GRO contains concepts and descriptions that help in the publishing and consuming of eCommerce data on the web. We investigated the GRDS by considering simple and common use cases of eCommerce, and observed how data responded to these requirements. Following the GR conceptual model and focusing on pivotal concepts, we issued targeted queries against the dataset and analysed the results. In our investigation, we firstly analyzed the overall conceptual coverage of the model to understand the data landscape. Secondly, we performed a focused analysis to understand the richness of data in GRDS. In the focused analysis, for each use case scenario, we firstly discussed the common understanding of concepts and the set of basic questions one can ask considering the dataset. Secondly, queries were made against these questions to retrieve information and get better data understanding. Finally, discussion is undertaken and noticeable observations are made at the end of each use case.

5.1 Analysis of Concept Coverage

To understand the overall distribution of data and the conceptual coverage of the GR model in GRDS, we use different queries in different combinations. The result is depicted in a chart (see Figure 2). In this chart, the y-axis represents the number of data sources, and x-axis represents used queries (the queries are listed in Table 5). The covered area reflects the information space available in GRDS. For example, point 6 of x-axis shows the number of data sources which provided data for the concepts listed in 6th row of Table 5. The query used for 6th row is available in listing 1.

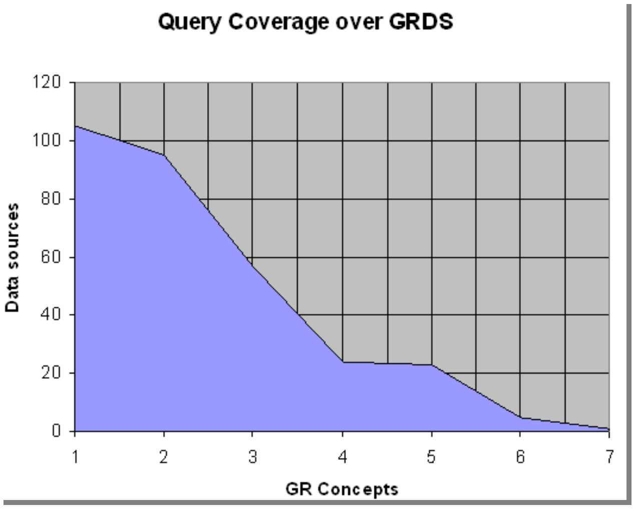


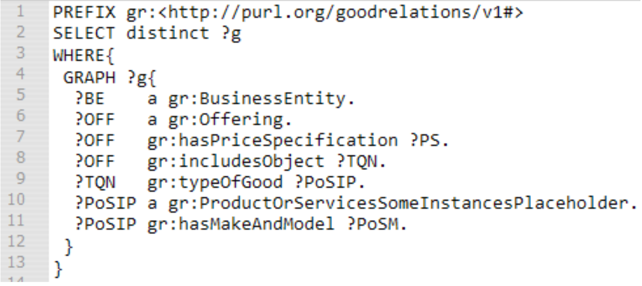
Figure 2: GRDS data coverage

[25] http://www.w3.org/International/articles/language-tags/

Table 5: Each row reflects the concepts used in query

1	BusinessEntity(BE)					
2	BE	:Offering(OFF)				
3	BE	OFF	:TypeAndQuantityNode(TQN)			
4	BE	OFF	TQN	:PriceSpecification(PS)		
5	BE	OFF	TQN	PS	:ProductOrServicesSomeInstancesPlaceholder(PoSIP)	
6	BE	OFF	TQN	PS	PoSIP	:ProductOrServiceModel(PoSM)
7	BE	OFF	TQN	PS	:ActualProductOrServiceInstance	

The highlighted area in the chart gives the kind of structured information that is currently available in eCommerce. Broadly speaking, we can say that on average every data publisher has provided business entity, offering and price details. However, almost no data source has provided any formal specification of the products being offered.



```
PREFIX gr:<http://purl.org/goodrelations/v1#>
SELECT distinct ?g
WHERE{
 GRAPH ?g{
  ?BE    a gr:BusinessEntity.
  ?OFF   a gr:Offering.
  ?OFF   gr:hasPriceSpecification ?PS.
  ?OFF   gr:includesObject ?TQN.
  ?TQN   gr:typeOfGood ?PoSIP.
  ?PoSIP a gr:ProductOrServicesSomeInstancesPlaceholder.
  ?PoSIP gr:hasMakeAndModel ?PoSM.
 }
}
```

Listing 1: Query (representing the concept involved in point 6 of chart's x-axis)

5.2 Use Case Based Analysis

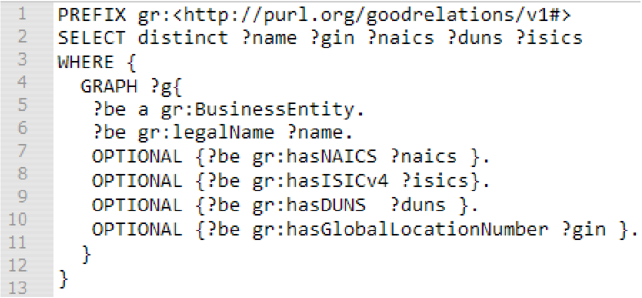
As mentioned earlier, we use generic and simple use case scenarios to understand the richness of semantic eCommerce data.

5.2.1 Finding a Company (Business Entity)

Finding a company is a very common and useful requirement in several situations particularly when seeking a company in a specific vertical industry, company offering specific product, company with specific business role (buyer/seller) or even finding competitors. Intuitively, one could ask many questions to obtain the required information from eCommerce information space. We have intentionally limited our search to the following questions as they are very basic and cover most user requirements.

- Find a company with a specific name
- Find a company in a particular location
- Find a company in a particulate line of business (or service)

These questions also contain basic parameters that if used in different combination can address more advanced requirements. To get the view of the structured information published by different data providers, we access the GRDS using SPARQL query shown in Listing 2.



```
PREFIX gr:<http://purl.org/goodrelations/v1#>
SELECT distinct ?name ?gin ?naics ?duns ?isics
WHERE {
  GRAPH ?g{
   ?be a gr:BusinessEntity.
   ?be gr:legalName ?name.
   OPTIONAL {?be gr:hasNAICS ?naics }.
   OPTIONAL {?be gr:hasISICv4 ?isics}.
   OPTIONAL {?be gr:hasDUNS  ?duns }.
   OPTIONAL {?be gr:hasGlobalLocationNumber ?gin }.
  }
}
```

Listing 2: Query (retrieving company description)

Result and Observations

In GRDS, 93.34% of the data sources (see Table 6) have provided a business name using :legalName property. This property is very helpful when searching for a company with a specific name using the SPARQL filter option. We found few data sources[26] which have not supplied a value for the legal name property. Upon further investigation of these provider's dataset, we found the presence of rdfs:label, vcard:fn properties but no value is attached to these attributes either.

Table 6: Use of location related attributes in GRDS

RDF Terms	**Data sources (%)**	**RDF Terms**	**Data sources (%)**
:BusinessEntity	100	**vCard:Address**	99.5
:legalName	93.34	vCard:country-name	99.5
:hasISICv4	0.95	vCard:locality	99.5
:hasNAICS	0.0	vCard:street-address	85.3
:hasDUNS	0.0	vCard:postal-code	85.3
:hasGlobalLocationNumber	0.0		

The unique identification of a Company (:BusinessEntity) on the Semantic Web using string value is complicated as multiple companies often have the same name. Hence, entity disambiguation [10] is required to qualitatively distinguish identically named companies from each other. GRO has useful attributes to attach concrete information that is helpful in identifying a company easily and accurately. However, in GRDS we found only one data source[27] that provided :ISICv4 code value (i.e 4652) along with company name. We did not find any value of other predicates mentioned in the OPTIONAL clause of the SPARQL query above (see Listing 2).

In GRDS, the second-most used schema after GRO is the vCard[28] that is used to provide the location and address-related information of a company or shop. 99.5% of the data sources have

[26] www.sachse-stollen.de, www.golfhq.com, ww.hagemann24.de, www.globalautoimports.com.br, www.xtremeimpulse.com www.cardgameshop.com, ww.discountofficehomefurniture.com

[27] www.jarltech.com

[28] Two different - one new and other deprecated- URIs are found in GRDS for vCard

provided information about the country and locality, and 85.3% have also provided a street address with postcode (postal address).

Location of Store[29] from the where service is provisioned is marked using :LocationOfSalesOrServicesProvisioining concept . It has relationship with both :BuisnessEntity (through :hasPOS) and :Offering (through :availableAtOrFrom). This allows one to obtain information about the shop by referring to Business Entity or to Offering. Shop location-related information is very helpful in many situations such as when visiting the shop, requesting online delivery or requesting a specific item in a particular location. In GRDS, 71.42% of data sources have provided shop information using :availableAtOrFrom and 44.76% have provided shop information using :hasPOS predicate. 39.04% data sources have provided information using both predicates.

34.28% of the data sources do not have opening and closing time details, number of days operating in a week and the validity of opening hour specification. 100% of the 65.72% who have provided opening hour details have provided :open and :closes time. However, not a single data source (0%) has provided :validFrom and :validThrough duration of opening hour specification. Also, we observe that 96.6% of the data sources have provided opening and closing hours time in UTC format and added 'Z' behind time (e.g 10:10:10Z).

5.2.2 Finding an Offer (Offering)

Making offer-related information available on the web in a structured format is one of the core objectives of GRO. :Offering concept has 13 data properties to describe offer attributes and 16 object properties to allow the creation of several relationships with other related concepts such as price specification, delivery options, payment or delivery charges, payment options, quantity and quality of products included in offer and warranty. As previously alluded to, :Offering is the mostly used concept after :BuisnessEntity and is part of almost all eCommerce use case scenarios.

- Find offering of a specific price range
- Find offering of a specific product and the available quantity
- Find delivery, warranty and payment charges of particular offering

Response to the above question depends on the offering data landscape. We pose different queries and data patterns found is collated in Table 7 & 8.

Table 7: Structure data provided with: Offering data in GRDS

RDF Terms	**Data sources (%)**	**RDF Terms**	**Data sources (%)**
:Offering	100		
:validFrom	82.86	rdfs:comment	77.14
:validThrough	82.86	rdfs:label	8.57
:eligibleRegions	82.86	v:name	0.95
:hasStockKeepingUnit	2.86	v:description	0.95
:hasEAN_UCC-13	1.90	v:price	0.95
:name	0.95	v:category	0.95
:description	0.95	dc:title	0.95
:availabilityStarts	0.95	dc:contributor	0.95
:hasGTIN-14	0.0	dc:date	0.95
:hasMPN	0.0	dc:description	0.95
:condition	0.0	dc:type	0.95
:serialNumber	0.0	dc:duration	0.95
:availabilityEnds	0.0		

Result and Observations

Table 7 lists all the data properties used for offer attributes (GR properties in left column and non GR attributes in right column) with their respective percentages and prefixes. Before explaining Table 7, it is important to mention that in the latest release, :name, :description and two more[30] data properties are added to the updated model to allow publishers to provide lexical information about the offers. Before this update, publishers have used rdfs:label and rdfs:comment to provide descriptive information about offers. This is why we see that 77.14% and 8.57% of data sources have used rdfs:comment and rdfs:label respectively in GRDS. One data source[31] has published data after the latest ontology update and has used :name and :description. 1.90% of the data sources[32] have provided EAN.UCC code and 2.86% have provided stock keeping code. Interestingly, in our dataset, we have one data source[33] that have also used Dublic Core (dc) and Google vocabulary (v) to describe offering data. This should be considered while querying for offer-related information.

From the data retrieval point of view, useful information is available through different relationships between different offers and other related concepts. However, one needs to rely heavily on the textual description of the offer instances to either filter or restrict the search based on string matching approach. Another noticeable observation is that quite a few terms overlap with other vocabularies (such as :name and v:name), and this should be considered when querying or generating customized rules.

Table 8: Object properties and their usage with : Offering in GRDS

RDF Term	**% of Data sources**	**RDF Term**	**% of Data sources**
:Offering	***100***		
:eligibleCustomerTypes	80.95	:hasWarrantyPromise	2.86

[29] It is important to note that herein, when we mention 'Store', it refers to the store, shop, branch office, office or any physical location, where service or product is being provisioned on behalf of Store's Company (Business Entity)

[30] :condition and :serialNumber

[31] www.jing-shop.com

[32] www.BestBuy.com, www.universum-shop.de

[33] http://bitmunk.com

:hasBusinessFunction	77.14	:hasInventoryLevel	0
:availableAtOrFrom	70.48	:advanceBookingRequirement	0
:acceptedPaymentMethods	60.95	:deliveryLeadTime	0
:includesObject	56.19	:eligibleDuration	0
:availableDeliveryMethods	47.62	:eligibleQuantity	0
:hasPriceSpecification	30.48	:eligibleTransactionVolume	0
:includes	3.8	:hasEligibleQuantity	0

Now, we look at the relationship patterns available between offering and other model concepts. In the GR model, there are two possibilities for linking an offering to products. When an offer has a single product, :includes is used, while :includesObject allows complex bundling of products. In GRDS, 59.99% data sources have linked offers with product data. The remaining 40.01% have used the offering concept to attach supplementary information such as eligible customer type, shop location information and supported payment methods. We can see in Table 8 that some of the relationships are available in most of the data sources, and few are not used at all. Evidently, 30.48% of companies have provided price specification details, and 80.95% of the data sources have identified the eligible customer type of the offer using GR predefined individuals such as :BusinessUser, :Endusers and :PublicInstitution. In GRDS, not a single data source has provided information on inventory level, advance booking requirement, delivery lead time, eligible duration of offer, eligible quantity to buy and eligible transaction volume. The absence of such information is common because this kind of information is required only for very unique specific products and usually such products are not offered by web shops.

From a consumer point of view, we would like to find the offers containing some specific product, and if we can find such offers, then we would like to know product price, delivery, payment and similar detail. In GR Primer[34], it is mentioned that at minimum, "*the basic structure of an offering is always a graph that links (1) a business entity to (2) an offering. The offering itself is linked to one or multiple type and quantity nodes and one or more price specification nodes. Each type and quantity node holds the quantity. The unit of measurement for the quality, and the product or service that is included in the offering"*. This means that retrieving the offering with a specific product in mind would require accessing concepts that relate product with offering, and provide details on quantity and unit of measurement. No formal product ontology is currently being used in GRDS. Therefore, we can query offering and filter records based on a textual description attached to offering. The statistics in Tables 7 and 8 show an 'offering' data landscape. One can make the following observations: (1) offerings can be retrieved with their price, quantity and offer start and end date, (2) a filter clause can be applied to properties with literal values (such as :name, :description, rdfs:label and rdfs:comment) to narrow the search to specific offering.

5.2.3 Finding a specific product (Product Find-ability)

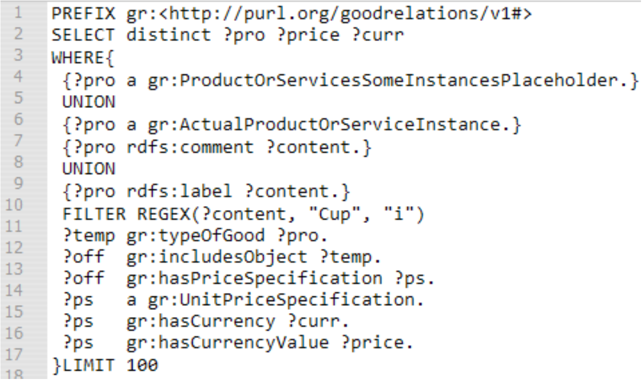
GRO provides three different ways to describe products. Each approach has different structural requirements to allow users to adopt the one which is best for them. In the first approach, GR recommends using an appropriate proper product/service ontology to describe products referred to in an offering. The second and less structural approach is to use GR top level Product or Service (:ProductOrService) concept and related vocabulary to define light weight product ontology tailored to individual specific need. The third and non-structural approach is to describe product information lexically. This approach allows users to restrict search to products having specific words in their textual description. Like previous use cases, we attempt to evaluate how GRDS respond to simple requests related to products such as:

- Find a particular product (e.g. TV or Shoes)
- Find a product with specific requirement (e.g. TV set of 24 inches, HD resolution)

Result and Observations

Upon accessing a dataset to find answers to the above questions (query in Listing 3), we observe that none of the data sources has used any formal product ontology to annotate products and their properties. However, we found that 2.86% of the data sources are using a second approach and using proprietary product ontology to describe quantitative properties. 97.14% of the data sources follow the third approach and publishing textual description of product rather than using any ontology. Two properties which are used for lexical information are rdfs:comment and rdfs:label. In our repository, we found no data source using appropriate product ontology.

Since the core feature of GR is to describe offers, products can be searched either by exploring the offer data or through the products included in the offers. In the absence of proper product ontology, we can search for a particular product by matching the keyword against the lexical information available in offers or product data.



```
PREFIX gr:<http://purl.org/goodrelations/v1#>
SELECT distinct ?pro ?price ?curr
WHERE{
 {?pro a gr:ProductOrServicesSomeInstancesPlaceholder.}
 UNION
 {?pro a gr:ActualProductOrServiceInstance.}
 {?pro rdfs:comment ?content.}
 UNION
 {?pro rdfs:label ?content.}
 FILTER REGEX(?content, "Cup", "i")
 ?temp gr:typeOfGood ?pro.
 ?off  gr:includesObject ?temp.
 ?off  gr:hasPriceSpecification ?ps.
 ?ps   a gr:UnitPriceSpecification.
 ?ps   gr:hasCurrency ?curr.
 ?ps   gr:hasCurrencyValue ?price.
}LIMIT 100
```

Listing 3: Query (retrieving product description)

The above query finds products containing "Cup" in their description, and displays price and associated currency. Query returned 58 products from three data sources[35] with price and currency value.

[34] http://www.heppnetz.de/projects/goodrelations/primer/

[35] www.jarltech.de, www.overstock.com, www.corsetsandcurves.com.au

5.3 Analysis of Axioms for Reasoning

RDFS and OWL defines a set of forward chaining rules [11] which can be used to infer the implicit knowledge in order to provide valid results for the queries. The inclusion of implicit knowledge in the query result is achieved through the use of a *reasoner* with axiomatic triples available in an ontology. Customized rules can be applied for deductive reasoning; however, we focus only on the axioms available in GRO listed in Table 1. Based on the restrictions implemented in the ontology, reasoning process also helps in finding the inconsistencies in instance data. Firstly, we looked at the instance data by applying the axiomatic triple using RDFS rule set. Secondly, using the class disjointness axioms, we performed disjointness checking in GRDS.

5.3.1 Inferencing

We looked at the instance data available in GRDS and applied an axiomatic triple using RDFS rule set to analyze the availability of implied information in triple store. Using the RDFS entailment rules [12] rdfs9 and rdfs7, we were able to retrieve additional information using more generic concepts. This was not available in queries evaluated without reasoning.

Table 9: (a) Implied knowledge (statements), (b) RDFS rule set applicable to GR model

GR Concepts	No Reasoning	With reasoning (rdfs9 rule)
:ProductOrService	0	16093
:PaymentMethod	14	24
:DeliveryMethid	10	16
:PriceSpecification	0	10723
:QuantitativeValue	0	6449

In Table 9, the concepts mentioned in the first column are the more generic concepts (superclass) of their specialised concepts (subclasses). We can see that, with reasoning, we are able to use generalized concepts to access the membership of subclasses.

In addition to subclass axioms, GRO contains subproperty axioms which allow two resources related through subproperty to be implicitly related by superproperty. Here, we used a diagram to represent the subPropertyOf subsumption and transitive behaviour of data type properties. There are 4 object properties which have been added recently (dated 2010-09-16) and no instance data is found with such object properties.

The results are retrieved by enabling RDFS-style reasoning based on backward chaining. This required a query rewrite in order to include the implicit knowledge entailed through RDFS entailment rules. In Figure 3(a), we see that :hasCurrencyValue has two superproperties which means that with RDFS-style reasoner, any query with :hasCurrencyValue in its predicate will return three triples, two additional triples entailed by applying the rule7 and one original triples having :hasCurrencyValue in predicate. The result of applying reasoning over GRDS by using quantitative value concept's data properties is illustrated in Figure 3(b).

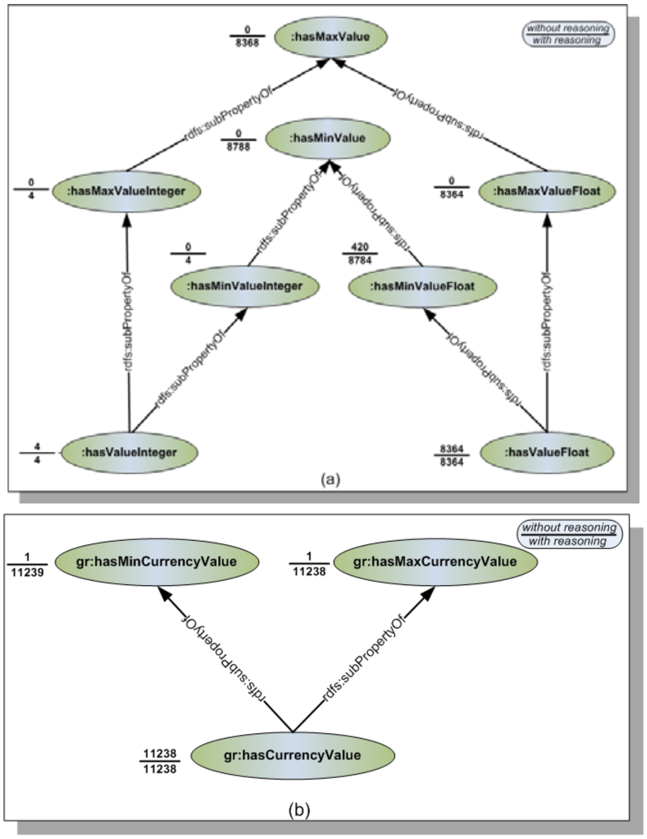


Figure 3: (a) Quantitative value data properties (b) Currency value data properties

In web eCommerce, offering price data value often comes as a fixed price value and produce. It is evident from Figure 4(a) that, except for one instance[36], all data sources have only provided fixed price value of offering. Data consumer will likely use more often this property to access price value and, with RDFS-style reasoner, its superproperties can return the same data. However, in a specific case where price range i.e. :hasMinCurrencyValue and :hasMaxCurrencyValue is provided and not :hasCurrencyValue, then using hasCurrencyValye with or without reasoning will not return any value. Here, custom rules can be applied to return Max price value when there is no :hasCurrencyValue property value available. To handle similar kinds of situations, the GR website provides a set of GoodRelations Optional Axioms[37] to allow users to obtain additional information from the dataset with minimum or, in certain situations, no side-effects.

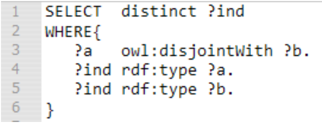
5.3.2 Disjointness checking

In Table 1, we saw the disjoint class axioms in GRO offering model consistency at the instance level. By making two classes disjoint, we are saying that the same individual cannot be an instance of these two (disjoint) classes simultaneously. For example, an individual declared to be instance of class :Offering cannot be declared as an instance of :BusinessEntity because in

[36] http://plushbeautybar.com/services.html#PriceSpec_10

[37] http://www.ebusiness-unibw.org/wiki/GoodRelationsOptionalAxiomsAndLinks

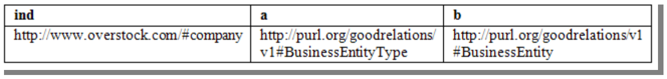
the GR model, both classes (concepts) are defined as disjoint classes. SPARQL query (Listing 4) finds such individuals. Upon accessing GRDS, we found one data source[38] violating GR model.



```
SELECT  distinct ?ind
WHERE{
   ?a   owl:disjointWith ?b.
   ?ind rdf:type ?a.
   ?ind rdf:type ?b.
}
```

Listing 4: SPARQL query

In Figure 4, we can see that the same URI is used as an instance of type :BusinessEntity and :BusienssEntityType; whereas in the model, both classes are declared as disjoint classes.



ind	a	b
http://www.overstock.com/#company	http://purl.org/goodrelations/v1#BusinessEntityType	http://purl.org/goodrelations/v1#BusinessEntity

Figure 4: Individual violating disjointness restriction

6. RELATED WORK

A large amount of research work has been done on ontology evaluation and a survey of different approaches is covered in [13]. In earlier papers, the focus was on analyzing the conceptual model coverage of ontology. Often, test data was used for such evaluation. However, little research work focused on cases where (real) instance data have been used and analyzed from an ontological model perspective.

Generic instance data Evaluation Process (GEP) [14] evaluates the instance data in knowledge management systems. Wine ontology is used with test instance data to discuss the different symptoms, their causes and way to generate potential issues. Findings are categorized into logical inconsistencies, syntax issues and detailed discussion around hypothetical potential issues. The study is of generic nature, and the instance data is evaluated using ontology that is primarily developed for learning purposes and does not reflect the actual usage or state of the instance data on the semantic web.

In [2], authors have analyzed the social and structural relationship available on semantic web considering FOAF vocabulary. The study is performed on approximately 1.5 million FOAF documents to analyze instance data available on the web and their usefulness in understanding social structures and networks. Additionally, the use of different namespaces, concepts and properties is discussed in order to provide a perspective on different FOAF implementations. This research provides only limited analysis since the prime focus was on social network-related instance data.

In [15], the authors provided a detailed study on the quality and state of published RDF data on the semantic web. Linked data principles were used to measure the noise and inconsistency available in a dataset, and reasoning was performed. While highlighting the issues and findings, the researchers have provided guidelines for both data publishers and data consumers to assist in generating and consuming high quality semantic data. Although the experiment is performed on the instance data collected from the web and has provided details on inconsistency and ontology hijacking in general, no particular ontology was considered while analyzing the data. In summary, these studies look at the instance data from a quality perspective or the use of test data for ontology evaluation. Our study performed on data sets from early adopters of open eBusiness ontologies represents a timely contribution and insight into community usage of the GRO.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have analysed the implementation of the GRO through the consolidated of 105 GR data sources into a single data set. We analyzed the use of other ontologies with GRO and categorize data providers. Different use cases were used to better understand and illustrate the schema usage and coverage through ontological instantiation. Data source provide structured data aimed at improving search ranking only with no interlinking currently available between eCommerce datasets or with LOD [16]. Links availability between disparate entities and use of open eBusiness ontologies (such as GRO) could well assist to the integration of disparate information sources.

Overall, the analysis points to early adoption and usage of an ontology that is beginning to achieve mainstream adoption with implementers using the GRO in an à la carte fashion rather than semantics a la mode.

In our future work, we plan to progress in two directions: i) toward a more comprehensive analysis of an expanded dataset. For this, we plan to collect datasets in intervals for duration of 6 months to determine if the status quo remains and, if not, how implementation develops with maturity gain; ii) evaluating the usefulness of structured data on the web. Here, we plan to investigate whether the inclusion of eCommerce structured data (annotated using GRO and other eBusiness ontologies) in search engine indexes (like Google, Yahoo!, etc) leads to improved performance metrics (recall and precision), and increase in business activity, as has already been evident in traffic increase by BestBuy [17].

8. REFERENCES

[1] Hepp, Martin, “GoodRelations: An Ontology for Describing Products and Services Offers on the Web,” in *Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW2008)*, Acitrezza, Italy, 2008, vol. 5268, pp. 332-347.

[2] Li Ding, “How the Semantic Web is Being Used: An Analysis of FOAF Documents,” in *Proceedings of the 38th Annual Hawaii International Conference on*, 2005.

[3] Silva, J, “E-Business Interoperability through Ontology Semantic Mapping,” in *Processes and Foundations for Virtual Organizations*, 2003, vol. 262, pp. 315-322.

[4] De Leenheer, P, “PhD Dissertation, Oncommunity-based Ontology Evolution,” Vrije Universiteit Brussel, 2009.

[5] H. J. ter Horst, “Completeness, decidability and complexity of entailment for rdf schema and a semantic extension involving the owl vocabulary,” *Journal of semantic web*, vol. 3, pp. 79-115, 2005.

[6] Thomas Steiner, “How Google is using Linked Data Today and Vision For Tomorrow,” Future Internet Assembly, Ghent, Belgium, 2010.

[7] Gruber, Tom, “Where the Social Web meets the Semantic Web, Web Semantics,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 1, pp. 4-13, 2008.

[8] G. Klyne, B. McBride, “Resource description framework (RDF): Concepts and abstract syntax.”World Wide Web Consortium", 2004.

[38] http://www.overstock.com/#company

[9] J. Madhavan, and C. Yu, “Structured data meets the web: A few observations,” presented at the IEEE Data Eng, Bull.,, 2006, vol. 29, pp. 19-26.

[10] Tummarello, Stefan, “Sig.ma: Live views on the Web of Data,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 8, no. 4, pp. 355-364, 2010.

[11] Edward Thomas, “Lightweight Reasoning and the Web of Data for Web Science,” in *Intertional Conference on Web Science (WebSci 2010)*, 2010.

[12] Patric Hayes, “RDF Semantics, W3C Working Draft.”2003.

[13] Janez Brank, “A survey of ontology evaluation techniques,” presented at the SIKDD 2005 at multiconference IS 2005, Ljubljana, Slovenia., 2005.

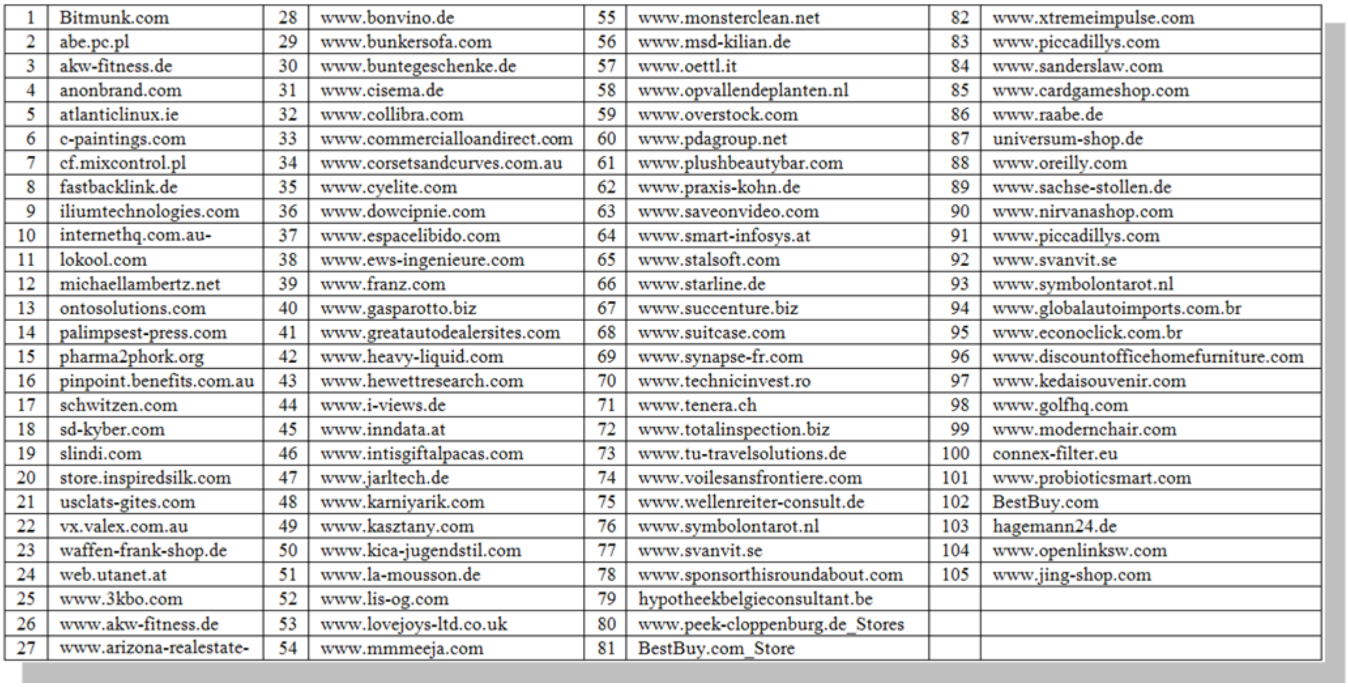
[14] Jiao Tao, “Instance Data Evaluation for Semantic Web-Based Knowledge Management Systems,” in *System Sciences, 2009*.

[15] A. Hogan,, “Weaving the pedantic web,” presented at the In 3rd International Workshop on Linked Data on the Web (LDOW2010) at WWW2010, Raleigh, USA, 2010.

[16] M. Hausenblas, “Exploiting linked data to build applications,” *IEEE Internet Computing*, vol. 13, no. 4, pp. 68-73, 2009.

[17]“http://www.wilshireconferences.com/semtech2010/RWW_070110.pdf.”.

Appendix: List of data sources in GRDS



1	Bitmunk.com	28	www.bonvino.de	55	www.monsterclean.net	82	www.xtremeimpulse.com
2	abe.pc.pl	29	www.bunkersofa.com	56	www.msd-kilian.de	83	www.piccadillys.com
3	akw-fitness.de	30	www.buntegeschenke.de	57	www.oettl.it	84	www.sanderslaw.com
4	anonbrand.com	31	www.cisema.de	58	www.opvallendeplanten.nl	85	www.cardgameshop.com
5	atlanticlinux.ie	32	www.collibra.com	59	www.overstock.com	86	www.raabe.de
6	c-paintings.com	33	www.commercialloandirect.com	60	www.pdagroup.net	87	universum-shop.de
7	cf.mixcontrol.pl	34	www.corsetsandcurves.com.au	61	www.plushbeautybar.com	88	www.oreilly.com
8	fastbacklink.de	35	www.cyelite.com	62	www.praxis-kohn.de	89	www.sachse-stollen.de
9	iliumtechnologies.com	36	www.dowcipnie.com	63	www.saveonvideo.com	90	www.nirvanashop.com
10	internethq.com.au-	37	www.espacelibido.com	64	www.smart-infosys.at	91	www.piccadillys.com
11	lokool.com	38	www.ews-ingenieure.com	65	www.stalsoft.com	92	www.svanvit.se
12	michaellambertz.net	39	www.franz.com	66	www.starline.de	93	www.symbolontarot.nl
13	ontosolutions.com	40	www.gasparotto.biz	67	www.succenture.biz	94	www.globalautoimports.com.br
14	palimpsest-press.com	41	www.greatautodealersites.com	68	www.suitcase.com	95	www.econoclick.com.br
15	pharma2phork.org	42	www.heavy-liquid.com	69	www.synapse-fr.com	96	www.discountofficehomefurniture.com
16	pinpoint.benefits.com.au	43	www.hewettresearch.com	70	www.technicinvest.ro	97	www.kedaisouvenir.com
17	schwitzen.com	44	www.i-views.de	71	www.tenera.ch	98	www.golfhq.com
18	sd-kyber.com	45	www.inndata.at	72	www.totalinspection.biz	99	www.modernchair.com
19	slindi.com	46	www.intisgiftalpacas.com	73	www.tu-travelsolutions.de	100	connex-filter.eu
20	store.inspiredsilk.com	47	www.jarltech.de	74	www.voilesansfrontiere.com	101	www.probioticsmart.com
21	usclats-gites.com	48	www.karniyarik.com	75	www.wellenreiter-consult.de	102	BestBuy.com
22	vx.valex.com.au	49	www.kasztany.com	76	www.symbolontarot.nl	103	hagemann24.de
23	waffen-frank-shop.de	50	www.kica-jugendstil.com	77	www.svanvit.se	104	www.openlinksw.com
24	web.utanet.at	51	www.la-mousson.de	78	www.sponsorthisroundabout.com	105	www.jing-shop.com
25	www.3kbo.com	52	www.lis-og.com	79	hypotheekbelgieconsultant.be		
26	www.akw-fitness.de	53	www.lovejoys-ltd.co.uk	80	www.peek-cloppenburg.de_Stores		
27	www.arizona-realestate-	54	www.mmmeeja.com	81	BestBuy.com_Store		