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
m.hadzic@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 product catalogue. As adoption increases, so too does the need to review and analyze current implementation of the ontology to better inform 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 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, Linked Data, Instance data analysis, Business ontology, Structured eCommerce data, Ontology usage.

1. INTRODUCTION

The Web of data and open ontologies (e.g. FOAF, SIOC, SKOS) promotes the establishment of a shared understanding between data providers and consumers in a common format that allows the automated processing of information by software agents. Where accepted by the community, an ontology offers the opportunity for enhanced dissemination and commercial use of information. The GoodRelations Ontology (GRO) [1], developed specifically for Web-based eCommerce, is an example of such an ontology that allows businesses to 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 second 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, dataset characteristics discussed and the usefulness of currently available data sets analysed through different use cases. Implicit data available through axiomatic triples is also considered.

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

2. MOTIVATION

The semantic web provides a level of semantically annotated structured data that enhances the level of user experience by more accurately sourcing and identifying information of interest. Enabled primarily through ontological alignment, semantic annotation is a major factor contributing to the increasing interest in ontology usage by the wider community and one which had also attracted the attention of early business adopters. Over the last two years, the GRO has witnessed this sectoral appeal with 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 the GRO will, for its corporate users, extend consumer reach to a larger audience with their increased appearance in search results. A measure of the popularity of any ontology is its community acceptance, which reflects 'some' level of use but not the extent of adherence to the schema or the extent of instantiation. A more accurate examination of popularity should therefore consider the

---

[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

overall ontology population. To date, however, the literature does not present evidence of systematic analysis of GRO usage that could provide this 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 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: better incorporating stakeholders' perspectives in ontology evolution [3] and ontology maintenance by analysing the ontology population and model coverage to help ontology engineers understand usage patterns;

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 also be useful.

3. GoodRelations ONTOLOGY OVERVIEW
In the following section, we describe our high level categorization of data providers, and present a brief overview of the GR conceptual schema and use of the 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 retailers who are traditionally premises-based and have only recently entered the eRetailing business. Such data sources provide more detailed (rich) product description which is useful for entity consolidation and interlinking with other datasets. Such companies include 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, offering their products and services mainly through web channels. Most of these web shops use web content management packages[4] such as Maganto, osCommerce and Joomla to add RDFa data in html pages. This approach works well since no special infrastructure arrangement is required in most cases.

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. Moreover, the providers use APIs to access and transform proprietary data into RDF before making them available through their repositories. For example, Linked Open

Commerce (LOC)[5] contains Amazon.com data although Amazon.com has not yet published RDF/RDFa.

3.2 Conceptual Schema
The latest version[6] of the GRO comprises 27 concepts (classes), 49 object properties, 43 data properties and 43 named individuals. To cater for backward compatibility, the ontology model was recently updated with the addition of new object and data properties based on implementation feedback. Note, **gr** is the GRO prefix used in general practice and throughout this paper. For the full specification of the GRO, the reader is referred to http://purl.org/goodrelations/v1.

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 a *reasoner*. The expressivity of the GRO is based on an OWL DLP fragment and contains subclass and subproperty axioms to express the subsumption behaviour in the model. Axiomatic triples in the GRO are given in Table 1 to shed light on the possible inference on eCommerce data which has been annotated using the GRO and applicable rule sets. RDFS and OWL elements such as rdfs:domain and rdfs:range, which are available in the ontology, are omitted from the table as they were not included in 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 were considered as they are associated with new knowledge. They can be used both in forward-chaining, to materialize the implied statements thereby making them explicit, and in backward-chaining performing query rewrites to expand query scope and include inferred knowledge. owl.DisjointClasses differs because it is used primarily for data quality and checking for 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 because it is generally available in most semantic repositories.

---

[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/

3.3 Use of the GRO by Search Engines

The adoption of the GRO is driven by the level of enhanced visibility that a company's products and general profile can receive as a result of its GRO marked-up data being included in the search engine indexes of large providers such as Google and Yahoo [6]. Yahoo and Google currently include price, availability (Google only), description and product pictures drawn from the GRO annotated structured data as part of their enhanced search results. BestBuy.com, a major implementer of the GRO, has announced[8] an increase of 30% traffic across their store's pages which contain the GRO annotated structured information. However, the literature does not contain any further study where the BestBuy findings are benchmarked and compared with others.

4. DATA SET

The eCommerce data sets constructed were collected primarily from those annotated with the GRO and represent the maximum number of accessible data sets. Throughout this paper, we use **GoodRelations Dataset or** GRDS to refer to the RDF graph collected from the various websites and stored in a triple store for querying and reasoning, and **Data Source** to refer to the websites' unique domain name server (DNS) included in GRDS, which contains eCommerce data in RDF (any serialization format) or an RDFa format based on the GRO model.

4.1 Data Set Collection

To analyse the adoption, usage patterns and uptake of the GRO in general, and by the eCommerce community in particular, data sets were collected from multiple sources and consolidated to create the GoodRelations Dataset[9] (GRDS). Potential data sources that used the GRO to describe the offerings or company (Business Entity) were the primary identification drivers. Different semantic search engines such as Sindice[10] and Watson[11] 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 list of data publishers maintained at the GoodRelations's developer wiki site[12]. For our empirical investigation, data was collected from 105 different data sources complying with the stated criteria. The complete list of these data sources is provided in the Appendix.

During the collection process, we noticed that 90% of the websites (data sources) were using the RDFa[13] standard to add structured information to existing HTML documents. The majority of sources had *sitemap.xml* files that allowed search engines to crawl the web pages and build indexes. However, the links (URLs) provided in the sitemap files were often linked to a list of products pages and not to the actual product pages themselves. Being interested in accessing the web pages that had

embedded RDFa code, we relied upon crawlers to build the list of such URLs before manually verifying the constructed list. With a valid list of URLs REST-based web services, Any23[14] and RDFa Distiller[15] were used to parse RDFa snippets from online HTML documents and generate RDF graphs (in RDF/XML syntax). Graphs were then loaded into the OpenLinks Virtuoso[16] triple store to create the GRDS experimental data set. From an RDF data management perspective, named graphs are used to group triples from one data source under a uniquely named graph URI, allowing the dataset to be queried vertically and horizontally.

Linked Open Commerce[17] (LOC) represents an emerging data space which collates eCommerce data from the Web and makes information available for retrieval and viewing through a SPARQL endpoint. Despite its presence, collection of data sets within this environment proved difficult owing 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-reference the URI, use it in a query or obtain provenance details. The LOC environment has approximately 34 data sources which publish and make their data available in RDF/RDFa format. Moreover, LOC contains a nominal number of data sources that are made available in RDF/RDFa but through the use of middleware APIs' such as Amazon.com. Invalid URIs such as those starting with "localhost.localdomain…." were also found[18] to be problematic, but overall represented minor data quality issue that was easily overcome.

The inclusion and use of these two datasets (i.e. LOC and GRDS) in our experiment provides the best opportunity for an optimal search space covering the maximum possible width (GRDS) and depth (LOC) of the structured eCommerce web-of-data. In essence, GRDS covers a greater number of data sources while LOC has greater coverage of data from data sources. Hence, both datasets complement one another with LOC providing an ability to cross check or find additional information which is useful for the analysis of results.

4.2 Dataset Composition Characteristic

During GRDS data set composition, the different characteristics of the datasets such as use of different namespaces, GR vocabulary and annotation properties were considered. The results from each investigation are described below.

4.2.1 Namespace Usage

Table 2 lists all vocabularies and their prefixes found in GRDS. Apart from *gr*, the top three most used vocabularies are *dc* (Dublin Core), *foaf* and *vCard*. Some vocabularies, such as *vCard* and *dc* were found to be used with multiple prefixes due to the availability of a new version with new namespace URIs. A larger percentage of focused vocabularies were 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.nytimes.com/external/readwriteweb/2010/07/01/01readwriteweb-how-best-buy-is-using-the-semantic-web-23031.html

[9] http://debii.curtin.edu.au/~jamshaid/GRDS-dump-v0.1.rar

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

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

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

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

[14] http://any23.org

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

[16] Open Source Edition, http://sourceforge.net/projects/virtuoso/

[17] http://linkedopencommerce.com

[18] Observations were made as of 16 OCT 2010

4.2.2 GR Vocabulary Usage

Here, an analysis of the GR vocabulary usage by different implementers was undertaken. The straightforward approach was used to calculate the number of instances each concept has in the dataset and calculate their properties used in implementation. While this approach helps to identify both the most and least populated terms, it does not provide sufficient understanding about usage patterns across different data sources. For example, if one particular class 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. It is also equally possible that this particular implementation has used this concept in GRDS. Therefore, we consider the usage of ontology terms based on the percentage of data sources where it is in fact used, rather than on the total number of triples in the dataset.

Table 2 : Prefixes and Namespaces used in GRDS

Prefix	Namespace URI	Data sources (%)
gr	http://purl.org/goodrelations/v1#	100
vCard	http://www.w3.org/2006/vcard/ns#	88.57
	http://www.w3.org/2001/vcard-rdf/3.0[19]	
	(deprecated namespace)	
dcterm	http://purl.org/dc/terms/	34.29
dc	http://purl.org/dc/elements/1.1/	
foaf	http://xmlns.com/foaf/0.1/	25.71
commerce[20]	http://search.yahoo.com/searchmonkey/commerce/	14.29
media	http://search.yahoo.com/searchmonkey/media/	
use	http://search.yahoo.com/searchmonkey-datatype/use/	
currency	http://search.yahoo.com/searchmonkey-datatype/currency/	
v[21]	http://rdf.data-vocabulary.org	3.81
og[22]	http://opengraphprotocol.org/schema/	0.95
rev[23]	http://purl.org/stuff/rev#	0.95
frbr[24]	http://vocab.org/frbr/core#	0.95
geo[25]	http://www.w3.org/2003/01/geo/wgs84_pos#	0.95

Our study also analysed the usage of GR schema from the perspectives of concept usage related to the instances and data sources found within the GRDS.

---

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

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

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

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

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

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

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

This allowed 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 insight into the relationships that exist between entities and the implementation of the ontological model in practice. To achieve this level of visibility, we investigated the level of 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).
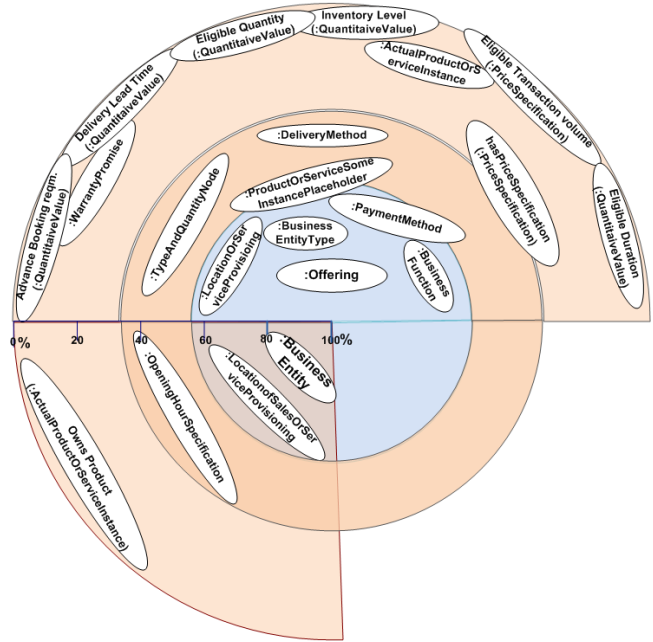
Figure 1: GR Concepts related to two pivotal concepts :Offering and :BusinessEntity

Figure 1 positions concepts on the diagram based on the percentage of their use across different data sources. Concepts are shown as the two groups of :Offering[26] and :BusinessEntity, with the groups intended to assist with visualizing the specific content and use of a particular data fragment. Several concepts appear on the edge of the outermost circle, indicating that several data providers have not made available any fine grained information about their offerings, although they have provided basic data sets for eligible customer types or business functions for activities such as selling, leasing or renting. The lower half of Figure 1 details concepts linked directly or indirectly to the business entity (:BusinessEntity). Overall, 60% of the data sources provided information regarding their stores (office/branches) and 40% of the data sources have made further details available such as the opening and closing times of stores. One of the pivotal concepts i.e :ProductOrService is not shown in Figure 1 as no formal product ontology is currently being used by GRDS. It was found that the :ProductOrService subconcepts were used throughout the

---

[26]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

offering data to describe whether the product referred to in the offering is the actual instance or existentially quantified.

4.2.3 Use of annotation properties

GRO 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. eCommerce deployment frequently used these in queries to retrieve resources of interest due to the properties being highly usable. 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. The 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 could prove to 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 indicated, en[27] (IETF's BCP 47 code of English Language) is the natural language most commonly used for providing textual description of the resources.

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

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. The GRO contains concepts and descriptions that help in the publishing and consuming of eCommerce data on the web. We investigated the GRDS by considering common eCommerce use and observed the data response 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 in order to understand the data landscape. Secondly, we performed a focused analysis so as to understand the richness of data in GRDS. As part of 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 of the dataset. Secondly, queries were constructed from these questions to retrieve information and provide a better understanding of data. Finally, an analysis was conducted 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, different queries in different combinations were used. Figure 2 depicts the results in chart format. The y-axis represents the number of data sources, and x-axis shows the number of used queries (the queries are listed in Table 5). The shaded 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.
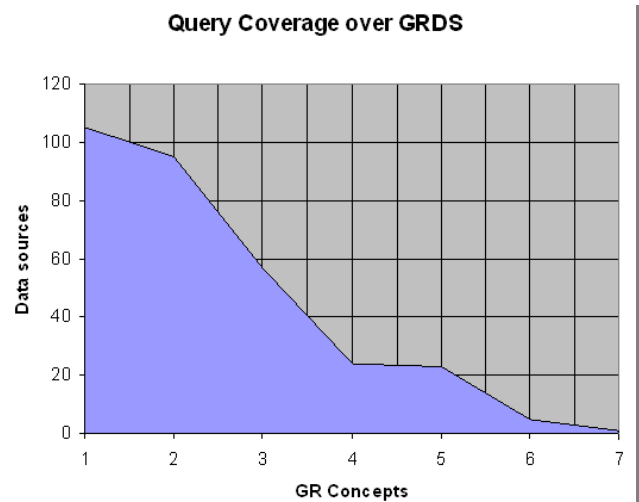
Figure 2: GRDS data coverage

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 details the type of structured information currently available in eCommerce. Broadly speaking, we observe 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.

```
1  PREFIX gr:<http://purl.org/goodrelations/v1#>
2  SELECT distinct ?g
3  WHERE{
4   GRAPH ?g{
5     ?BE    a gr:BusinessEntity.
6     ?OFF   a gr:Offering.
7     ?OFF   gr:hasPriceSpecification ?PS.
8     ?OFF   gr:includesObject ?TQN.
9     ?TQN   gr:typeOfGood ?PoSIP.
10    ?PoSIP a gr:ProductOrServicesSomeInstancesPlaceholder.
11    ?PoSIP gr:hasMakeAndModel ?PoSM.
12   }
13  }
```

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

5.2 Use Case-Based Analysis

As previously mentioned, we use generic use case scenarios to illustrate the extensive use of semantic eCommerce data.

5.2.1 Finding a Company (Business Entity)

Finding a company is a very common and useful requirement in multiple situations, particularly when seeking a company in a specific vertical industry, a company offering a specific product, a company with a specific business role (buyer/seller), or even competitors. Intuitively, one could ask many questions to obtain the required information from the 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 particular line of business (or service)

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

```
1  PREFIX gr:<http://purl.org/goodrelations/v1#>
2  SELECT distinct ?name ?gin ?naics ?duns ?isics
3  WHERE {
4   GRAPH ?g{
5     ?be a gr:BusinessEntity.
6     ?be gr:legalName ?name.
7     OPTIONAL {?be gr:hasNAICS ?naics }.
8     OPTIONAL {?be gr:hasISICv4 ?isics}.
9     OPTIONAL {?be gr:hasDUNS  ?duns }.
10    OPTIONAL {?be gr:hasGlobalLocationNumber ?gin }.
11   }
12  }
13 }
```

Listing 2: Query (retrieving company description)

Result and Observations

With reference to Table 6, within GRDS, 93.34% of the data sources provided a business name using the :legalName property. This property is very helpful when searching for a company with a specific name using the SPARQL filter option. A few data sources[28] were found which did not supply a value for the legal name property. Further investigation of these providers' datasets found the presence of rdfs:label, vcard:fn properties but also with no attribute value.

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 a string value is complicated as multiple companies often have the same name. Entity disambiguation [10] is required to distinguish identically named companies from each other. Despite the fact that the GRO has useful attributes that assist in identifying a company easily and accurately, we found only one data source[29] in the GRDS that provided both the :ISICv4 code value (i.e. 4652) and company name. We did not find any value for the other predicates mentioned in the OPTIONAL clause of the SPARQL query above (see Listing 2).

In the GRDS, the second-most widely used schema (after GRO) is vCard[30] which provides the location and specific address of a company or shop. 99.5% of the data sources provided information

---

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

[29] www.jarltech.com

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

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

Location of Store[31], indicating where the service/product is provided, is annotated using the :LocationOfSalesOrServicesProvisioining concept . It has relationships with both :BusinessEntity (through :hasPOS) and :Offering (through :availableAtOrFrom). This allows information about the shop to be accessed by referring to the Business Entity or Offering. Shop location-related information is very helpful in many situations such as when visiting the shop, requesting online delivery or searching for a particular 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 working time details, or number of operating days per week. All of the 65.72% who have provided opening hour details also provided :open and :closes time. However no data source provided :validFrom or :validThrough opening hour specification. We also observed that 96.6% of the data sources have provided opening and closing times in UTC format and added 'Z' after time (e.g 10:10:10Z).

5.2.2 Finding an Offer (Offering)

Making offering-related information available on the web in a structured format is one of the core objectives of GRO. The :Offering concept has 13 data properties that describe offering attributes and 16 object properties allowing several relationships to be established 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 stated, :Offering is the most widely used concept after :BuisnessEntity and is found in almost all eCommerce use case scenarios. Such as:

- 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

Responding to these question is dependent on the offering data landscape as illustrated in Table 7 and 8 with the different query and data patterns found.

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 details the data properties used for offering attributes (GR properties in column 1 and non GR attributes in column 3, with their respective percentages and prefixes. It is important to mention that the latest GR release has :name, :description and two more[32] data properties added to model allowing publishers to provide lexical information about offerings. Prior to this update publishers used rdfs:label and rdfs:comment to provide descriptive information about offerings. This is reason that 77.14% and 8.57% of data sources have used rdfs:comment and rdfs:label respectively in GRDS. One data source[33] which published data after the latest ontology update has used :name and :description. 1.90% of the data sources[34] have provided a EAN.UCC code and 2.86% a stock keeping code. Interestingly, with GRDS we found one data source[35] that had also used Dublic Core (dc) and Google vocabulary (v) to describe offering data, which we additionally considered when querying for offering-related information.

From the perspective of data retrieval, useful information is available through relationships between different offerings and other related concepts. However, in order to either filter or restrict the search based on string matching, the textual description of the offer instance has to be relied upon. Another finding was that the prevalence of terms overlapping across vocabularies (such as :name and v:name), which also had to 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

---

[31] 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 the service or product is being provided on behalf of the store's Company (:BusinessEntity)

[32] :condition and :serialNumber

[33] www.jing-shop.com

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

[35] 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

Relationship patterns between offerings and other model concepts were looked at. The GR model provided two means of linking offering to products. When an offer has a single product, :includes is used, and for complex product bundling, :includesObject is used. In GRDS, 59.99% of data sources linked offers with product, and the remaining 40.01% use the offering concept to attach supplementary information such as eligible customer type, shop location information and supported payment methods. Tables 8 outlines the availability of relationships across the data sources with some not being used at all. 30.48% of companies provided price specification details, while 80.95% of the data sources identified the eligible customer type of the offer using GR predefined individuals such as :BusinessUser, :Endusers and :PublicInstitution. Within the GRDS, no data source provided information on the inventory level, advance booking requirement, delivery lead time, eligible duration of offer, eligible quantity to buy or eligible transaction volume. These omissions are not uncommon as this kind of information is required only for specific (unique) products not offered by web shops.

Consumers normally like to find offers containing some specific product and if found, they look for the product price, delivery method, payment details etc. The GR Primer[36] mentions 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*". Retrieving offers with a specific product in mind would therefore require accessing concepts that relate product with offering and provide details on quantity and unit of measurement. As no formal product ontology is currently being exploited in the GRDS, we could only query offering and filter records based on a textual description attached to the offering. Tables 7 and 8 show an 'offering' data landscape. The following observations can be made: i) offerings can be retrieved with their price, quantity and offer start and end date; and ii) a filter clause can be applied to properties with literal values (such as :name, :description, rdfs:label and rdfs:comment) to narrow the search for specific offering.

5.2.3 Finding a specific product (Product Findability)
GRO provides three means of describing products. Each approach has different structural requirements, allowing users to select the most appropriate. The first recommends using an appropriate

product/service ontology to describe products referred to in an offering. The second and less structural approach allows lightweight product ontology tailored to individual specific needs using the GR top level Product or Service (:ProductOrService) concept and related vocabulary. The third, non-structural approach allows the description of product information lexically. This approach allows users to restrict their search to products with specific terms in their textual description. As in the previous use cases, we evaluated how GRDS responded to product related requests 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

The query in Listing 3 was used to 'ask' the question which highlighted the lack of any formal product ontology to annotate products and their properties. We did however find that 2.86% of data sources used the second approach of using proprietary product ontology to describe quantitative properties. 97.14% of the data sources follow the third approach, publishing textual description of the product rather than the ontology. Two properties found in use for lexical information are rdfs:comment and rdfs:label. In general however we found no evidence within the GRDS of data sources using an appropriate product ontology.

Since GR has aits core the description of offers, products can be searched either by exploring their offer data or through products included in the offers. In the absence of a proper product ontology we queries particular products by matching the keyword against the lexical information available in offer or product data.

```
1  PREFIX gr:<http://purl.org/goodrelations/v1#>
2  SELECT distinct ?pro ?price ?curr
3  WHERE{
4    {?pro a gr:ProductOrServicesSomeInstancesPlaceholder.}
5    UNION
6    {?pro a gr:ActualProductOrServiceInstance.}
7    {?pro rdfs:comment ?content.}
8    UNION
9    {?pro rdfs:label ?content.}
10   FILTER REGEX(?content, "Cup", "i")
11   ?temp gr:typeOfGood ?pro.
12   ?off  gr:includesObject ?temp.
13   ?off  gr:hasPriceSpecification ?ps.
14   ?ps   a gr:UnitPriceSpecification.
15   ?ps   gr:hasCurrency ?curr.
16   ?ps   gr:hasCurrencyValue ?price.
17 }LIMIT 100
18
```

Listing 3: Query (retrieving product description)

The query in Listing 3 finds products containing "Cup" in their description, and displays price and associated currency, returning returned 58 products from three data sources[37] with price and currency value.

5.3 Analysis of Axioms for Reasoning
RDFS and OWL define a set of forward chaining rules [11] which can be used to infer implicit knowledge and provide valid query results. The inclusion of implicit knowledge in the query result is

---

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

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

achieved by using a *reasoner* with axiomatic triples available in an ontology. Despite the availability of customized rules for deductive reasoning, we focused on the axioms (listed in Table 3) available in GRO only. Based on ontological restrictions, the reasoning process helps to identify inconsistencies in instance data. We first looked at the instance data by applying the axiomatic triple using the RDFS rule set before using the class disjointness axioms, to perform disjointness checking in GRDS.

5.3.1 Inferencing

Implied information in the GRDS was investigated by applying an axiomatic triple using the RDFS rule set to analyze the availability of implied information at the data instance level. Using the RDFS entailment rules [12] rdfs9[38] and rdfs7[39], we were able to retrieve additional information using more generic concepts. This was not possible for those queries evaluated without reasoning.

Table 9: Implied knowledge (statements)

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

The concepts mentioned in the first column of Table 9 reflect the more generic concepts (superclass) of their specialised concepts (subclasses). With reasoning, we are able to use generalized concepts to access subclass membership.

In addition to subclass axioms, the GRO contains subproperty axioms which allow two resources, which are related through subproperty, to be implicitly related by superproperty. Figure 3 represents the subPropertyOf subsumption and transitive behaviour of data type properties. Of the 4 object properties recently added (2010-09-16), no instance data was found. As RDFS-style reasoning based upon backward chaining was used, it was necessary to rewrite queries so as to include implicit knowledge generated through RDFS entailment rules. In Figure 3(a) :hasCurrencyValue has two superproperties demonstrating that with RDFS-style reasoners any query with :hasCurrencyValue in its predicate will return three triples, two additional triples entailed by applying rule 7 and one original set of triples having the :hasCurrencyValue predicate. Figure 3(b) shows the results of applying reasoning over GRDS by using quantitative value concept's data properties.

In web eCommerce, price data is often presented as a fixed price value and the user focuses on the :hasCurrencyValue property. Figure 4(a) draws attention to the fact that, apart from for one instance[40], all data sources have provided only a fixed price value for their offerings.
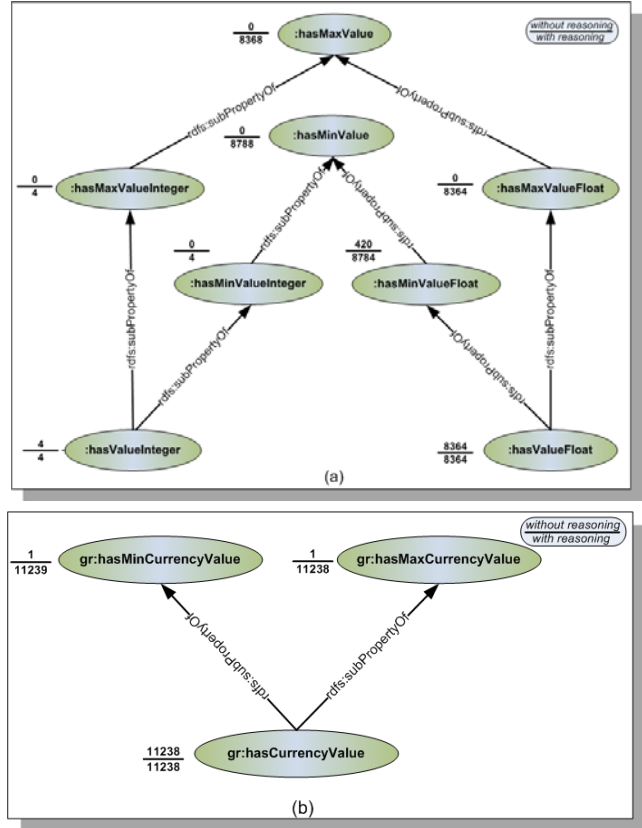
Figure 3: (a) Quantitative value data properties (b) Currency value data properties

Data consumers will most likely use this property to access price value and, with the 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 :hasCurrencyValue 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[41] to allow users to obtain additional information from the dataset with minimal side-effects.

5.3.2 Disjointness checking

In Table 1, we saw that the disjoint class axioms in the GRO offer model consistency at the instance level. By making two classes disjoint, the same individual cannot be an instance of both (disjoint) classes simultaneously. For example, an individual declared to be an instance of class :Offering cannot be declared as an instance of :BusinessEntity since, in the GR model, both classes are defined as disjoint classes. The SPARQL query in Listing 4 finds such individuals and within GRDS we identified one data source[42] violating the GR model.

---

[38]IF(uuu rdfs:subClassOf xxx AND vvv rdf:type uuu) THEN (vvv rdf:type xxx)

[39]IF(aaa rdfs:subPropertyOf bbb AND uuu aaa yyy) THEN (uuu bbb yyy)

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

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

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

```
1  SELECT  distinct ?ind
2  WHERE{
3      ?a   owl:disjointWith ?b.
4      ?ind rdf:type ?a.
5      ?ind rdf:type ?b.
6  }
```

Listing 4: SPARQL query

In Figure 4, 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 conceptual model analysis coverage of the ontology using test data only. There is little evidence in the literature of work that focuses on cases where instance data based upon actual field implementation has been used and analyzed from an ontological model perspective. Generic instance data Evaluation Process (GEP) [14] evaluates instance data in knowledge management systems. The Wine ontology is used with test instance data to discuss the different symptoms, their causes and ways to generate potential issues. Findings are categorized into logical inconsistencies, syntax issues and detailed discussion around hypothetical potential issues. The study is generic in nature and the instance data is evaluated using an ontology primarily developed for learning purposes but which does not reflect the actual usage or state of the instance data on the semantic web.

[2] has analyzed the social and structural relationship available on semantic web by considering FOAF vocabulary. The study was 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 a limited analysis since the primary focus was on social network-related instance data.

[15] 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 was 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 for data analysis. In summary, these studies examine 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 analysed the implementation of the GRO by consolidating 105 GR data sources into a single data set. We analyzed the use of other ontologies with the GRO and categorized data providers. Different use cases were used to better understand and illustrate the schema usage and coverage through ontological instantiation. Data sources provide structured data aimed at improving search ranking only with no interlinking currently available between eCommerce datasets or with LOD [16]. The availability of links between disparate entities and the use of open eBusiness ontologies (such as the GRO) could well assist to integrate 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 intend to progress in two directions: i) toward a more comprehensive analysis of an expanded dataset. For this, we plan to collect datasets at intervals for a duration of six months to determine whether the status quo remains unchanged and, if not, how implementation develops with increased maturity; ii) evaluating the usefulness of structured data on the web. Here, we plan to investigate the impact of eCommerce structured data (annotated using GRO and other eBusiness ontologies) in search engine indexes (like Google, Yahoo!, etc) and measure the increase in business activity, as has already been evident in the traffic increase for BestBuy [17].

8. ACKNOWLEDGMENTS

9. 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, On community-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.

[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		