BauDataWeb: The Austrian Building and Construction Materials Market as Linked Data

Andreas Radinger, Bene Rodriguez-Castro, Alex Stolz, and Martin Hepp
Universitaet der Bundeswehr Munich
Werner-Heisenberg-Weg 39
D-85579 Neubiberg, Germany
{andreas.radinger,bene.rodriguez,alex.stolz,martin.hepp}@ebusiness-unibw.org

ABSTRACT

In this paper, we describe the technical approach of and experiences gained in exposing a major share of the European building and construction materials market on the basis of the GoodRelations Web vocabulary for e-commerce. This allows for the fine-grained search for products, suppliers, and warehouses for any building-related sourcing needs. Because building materials show a very high item specificity, they are very interesting for new types of search. Also, transportation costs for building materials are very significant, which makes the distance from the warehouse to the point of consumption a critical dimension of search. Based on existing data sources, we were able to include a rich, machine-readable description of individual product features using the FreeClassOWL ontology, which allow for multi-dimensional search. The result is one of the largest and richest public datasets for a well-defined trade sector that is available on the Semantic Web.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

E-Commerce, Linked Data, GoodRelations, FreeClass, Building Materials, Product Classification Standards

1. INTRODUCTION

Search and matchmaking between two business parties over the current Web are still very time-consuming if (1) the objects of the search are very specific; (2) multiple characteristics of these objects determine their relevance (e.g. not just the price); (3) information from multiple sources needs to be combined to assess the relevance or execute the query; and (4) the locations of buyer, seller, warehouse, and point of consumption matter.

One of the domains that aligns strongly with these four characteristics is that of building and construction materials. Because building materials tend to present several functional characteristics and show a very high item *specificity* [10], they are very interesting for new types of search. Also, transportation costs for building materials are very significant, making the distance from the warehouse to the point of consumption a critical dimension of search. Or what in the context of Web search, is sometimes attributed to the *findability* of a resource [13].

An important source of building and construction materials on the Web is the Eurobau portal[1]. As the creators and administrators of the portal states[2], "Eurobau.com is a comprehensive internet portal for the commercial construction industry, with 200,000 users/month." It compiles construction materials data from ten European countries, available in eight languages, with Austria being one of the main contributors to the repository.

To facilitate browsing, navigation and integration of the data, Eurobau employs an open source classification standard developed specifically for the building materials domain, known as FreeClass[3] [4]. FreeClass provides a comprehensive range of categories aimed at covering all relevant attributes and application areas of building materials. Thanks to the FreeClass classification standard and a Web application, users of the Eurobau portal can browse and explore all the available data. However, in its current form, the data can only be viewed, accessed or searched in a limited fashion. The most notable limitations include for example (a) searching for building materials over more than one FreeClass category (i.e., *all* materials in categories 14-15-10 "insulation plaster" *and* 14-15-20 "restoration plaster"); (b) querying over all materials that meet a specific set of attributes or requirements (i.e., *all* bricks whose width is *greater than* 10 centimeters); (c) finding manufacturers within certain distance because no *geographical coordinates* (or geo-code) information is stored; or (d) the fact that data can not be combined easily with other data repositories on the Web.

In order to increase the applicability and reusability potential of the data hosted by the Eurobau portal, we describe in this paper the technical approach of and experiences gained in exposing this major share of the European building and construction materials market on the basis of the GoodRelations Web vocabulary for e-commerce [8].

ISEM'13, September 04 - 06 2013, Graz, Austria
Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-1972-0/13/09 ...$15.00.
http://dx.doi.org/10.1145/2506182.2506186

[1] http://www.eurobau.com/

[2] http://www.inndata.at/

[3] http://www.freeclass.eu/

Based on existing data resources, we were able to include a rich, machine-readable description of individual product features using the RDF data model, which allows for multi-dimensional and fine-grained search for products, suppliers, and warehouses for any building-related sourcing needs. The result is one of the largest and richest public RDF datasets for a well-defined trade sector that is available on the Semantic Web, covering a major share of the respective European market.

The work involved throughout this process has been carried out in the context of the BauDataWeb project, an initiative of the Research Promotion Agency (FFG)[4] of the Austrian Government, to bring together academic research and industry. The most significant components involved are summarized on the project overview site[5] and they include: (a) FreeClassOWL, a fully GoodRelations-compliant ontology for describing construction and building materials and services, derived from the FreeClass classification standard; (b) the Eurobau Utility ontology, which defines a few key extensions to GoodRelations for this particular vertical domain; (c) the BauDataWeb RDF dataset, containing over 88 million triples of real business data with a high domain density, which includes geographical data for warehouse locations (of manufacturers and resellers) and in a large part, product features using the FreeClassOWL ontology.

An additional and intended capability of this BauDataWeb RDF dataset is that it can be combined with other data sources on the Semantic Web, such as: (a) DBpedia[6], for information about business entities, population or transportation infrastructure; (b) LinkedGeoData[7], which provides open geographical data in a structured format; (c) governmental information; or (d) real estate offers. To that effect, the BauDataWeb RDF dataset is available via a SPARQL endpoint and all deliverables are fully compliant with current open W3C Semantic technologies and Linked Data principles.

To illustrate our contribution, we equipped BauDataWeb with a demonstrative application to serve as both, a user-friendly interface and an evaluation tool to the new functionality provided. All of this work is discussed in detail throughout the rest of the paper as follows: Section 2 analyzes the main components of the project involved in the creation of a BauDataWeb RDF dataset; Section 3 provides a quality assessment of the produced RDF and presents a demonstrative application to illustrate some of the new functionality gained; Section 4 reviews existing relevant resources considered in the context of this work; and finally Section 5 concludes the paper with our final remarks.

2. PRODUCING BAUDATAWEB

This section focuses on the various aspects relevant to the methodology employed for the creation of the FreeClassOWL ontology, the Eurobau Utility ontology, the BauDataWeb RDF dataset and the strategy for the continued management of this data repository. Figure 1 provides a high-level architecture view of the interaction among these components and it will help to anchor the discussion going forward.

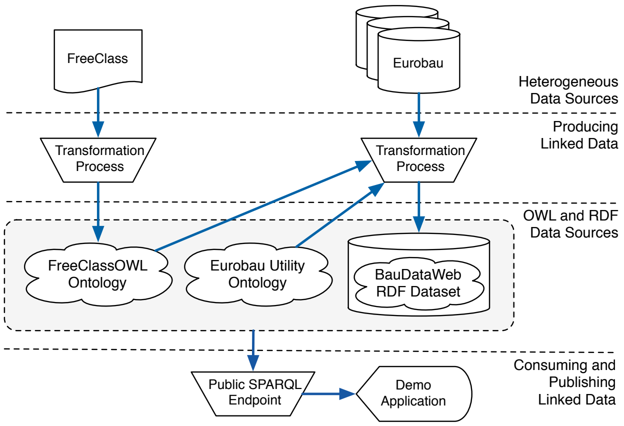

Figure 1: Architecture of BauDataWeb

Element	Amount	Comment
Hierarchy Level	(Total) 2,838	
- Segments	11	1st level
- Families	86	2nd level
- Classes	480	3rd level
- Commodities	2,261	4th level
Properties	174	
- Object	172	
- Datatype	2	
Predefined Values	1,423	Qualitative Values
Languages	8	en, de, cs, it, ro, sk, hu, pl

Table 1: Statistics of FreeClass

2.1 The FreeClass Classification in OWL

FreeClass is an open-source product classification standard for the domain of building and construction materials [4]. All categories (nodes) in the classification are arranged in a hierarchical tree structure. Analyzing the classification, we noticed the following main characteristics: more than 2,800 categories in four hierarchical levels, 170 properties, and 1,400 predefined values. FreeClass provides also description labels for all of these elements that in most cases are available in eight languages (Czech, English, German, Hungarian, Italian, Polish, Romanian, and Slovak). See Table 1. Note that languages are listed using their 2-letter codes as per ISO 639-2[8]. One of the goals of the BauDataWeb project was to develop a GoodRelations-compliant ontology based on the FreeClass product classification for describing construction and building materials, which will be referred to hereon as FreeClassOWL. All categories in the FreeClass classification hierarchy are defined by an eight-digit numeric code, where every two digits classify a new level and the level can be determined by the length of the numeric code. The first level is composed of eleven different categories.

2.1.1 GoodRelations Ontology

GoodRelations is a light-weight Web ontology for e-commerce on the Semantic Web [8]. The ontology pro-

[4] http://www.ffg.at/innovationsscheck
[5] http://semantic.eurobau.com/
[6] http://dbpedia.org/
[7] http://linkedgeodata.org/
[8] http://www.loc.gov/standards/iso639-2/

vides comprehensive support for the representation of the most frequently used concepts involved in the description of an offering, such as product or service details, business entities, prices, and terms and conditions among others. To further categorize products and to describe them more precisely, GoodRelations also allows to extend products with classes and features of comprehensive product classification standards (e.g. eClassOWL [6] or the Product Types Ontology[9]). The resulting FreeClassOWL ontology relied on this extension capability of GoodRelations to accommodate product classifications. GoodRelations elements will be referred to using the commonly accepted namespace prefix *gr:*. Thus, `http://purl.org/goodrelations/v1#ProductOrService` is equivalent to *gr:ProductOrService*.

2.1.2 FreeClassOWL Ontology

The process of creating a valid, consistent OWL ontology from the FreeClass product classification included several steps. The main steps are aligned to the suggested guidelines for creating vertical product ontologies that extend and comply with GoodRelations[10]: namely, (1) defining all classes of products and services needed; (2) collecting all relevant properties for each class; (3) for every property, determining which type of property it is (either object or data type); (4) collecting all relevant value types or predefined values. To assist with this process, we relied on a set of *four* files in comma separated value (CSV) format that described all elements in the full open FreeClass classification, using a proprietary encoding developed by our project industrial partner, *inndata Datentechnik GmbH*[2]. The transformation into a GoodRelations compliant OWL ontology was implemented using a Python script, which ultimately delivered the FreeClassOWL ontology as one file in RDF/XML format. The namespace URI created for the FreeClassOWL ontology is `http://www.freeclass.eu/freeclass_v1#` and its elements will be referred to by using the namespace prefix *fc:*. Hence, `http://www.freeclass.eu/freeclass_v1#hierarchyCode` is equivalent to *fc:hierarchyCode*. The rest of this section describes these main steps in more detail and the most relevant design decisions applied.

2.1.3 Classes of FreeClassOWL

The first step was to generate all classes of the FreeClassOWL ontology. To do so, the GenTax (Generic/Taxonomic) approach [9] was used given that it allows to preserve the taxonomic structure of the original FreeClass categories in the class structure of the resulting ontology. The GenTax method derives two OWL classes from each FreeClass classification category. The first is a context-specific class, in this case in the domain of building materials, and is a specialization of *gr:ProductOrService*. The second represents a broader taxonomic concept that preserves the hierarchy of the original FreeClass structure using the *rdfs:subClassOf* relation. The naming convention for the identifier of these two types of classes starts with *"fc:C_"*, followed by an *"ID"* and ends with *"-gen"* or *"-tax"* respectively.

For traceability purposes to the original schema, a series of design decisions were applied in the conversion process: (1) the *ID* component of every class name, is mapped to a *strong identifier* field in the source CSV files. The strong identifier is unique to the specific FreeClass category that the associated *fc:C_ID-gen* and *fc:C_ID-tax* pair of OWL classes represent; (2) every class *fc:C_ID-gen* is, in addition of being a subclass of *gr:ProductOrService*, also a subclass of the corresponding *fc:C_ID-tax* class, to preserve its alignment to the category in the original FreeClass classification that it was derived from; (3) the taxonomic class *fc:C_ID-tax*, stores the eight-digit numeric code of the original FreeClass category via the annotation property *fc:hierarchyCode*; and (4) the context specific class *fc:C_ID-gen*, stores the URL of the HTML page associated to the category in the FreeClass Web portal[3] via the property *rdfs:seeAlso*. These provenance traces between the original FreeClass product classification and the classes in the FreeClassOWL ontology, allow for exploiting the original hierarchy for queries and other operations on the data annotated with the resulting ontology.

The eight language descriptions supported in FreeClass, are preserved in all FreeClassOWL classes via the property *rdfs:label*, including the appropriate two-letter language tag in the XML Schema data type string. Additional comments for further describing the FreeClassOWL classes, are given only in English via the property *rdfs:comment*. In certain cases, where languages are not given, special filters avoid the generation of empty property value literals.

2.1.4 Properties of FreeClassOWL

The conversion of properties in the FreeClass product classification to OWL properties in the FreeClassOWL ontology was perhaps the most manually intensive task. The task implied the analysis of every property in the source CSV files and it turned out particularly challenging because ultimately, it required certain knowledge or familiarization with concepts specific to the building materials industry. The range of possible values observed for a property, determined the appropriate type of OWL property that it should be transformed into in FreeClassOWL, i.e. either a data type or an object property. Three distinct sets of range values were observed, namely (1) strings of characters; (2) numeric values representing some type of unit of measurement (i.e. centimeters, cubic meters, etc.); or (3) predefined values representing features applicable to certain properties (i.e. "water repellent", "corrosion protection", "iron", etc.).

In terms of the FreeClassOWL ontology, FreeClass properties whose range of values aligned to (1) above, were transformed into OWL data type properties, while properties whose values aligned to either (2) or (3), were transformed into OWL object properties. Furthermore, based on the guidelines to build GoodRelations-compliant product ontology extensions[10], OWL data type properties became subproperties of *gr:datatypeProductOrServiceProperty*; and OWL object properties whose values aligned to either (2) or (3), became subproperties of either *gr:quantitativeProductOrServiceProperty*, or *gr:qualitativeProductOrServiceProperty* respectively. (From here on, the suffix *"-POSP"* will be used as a shortcut for *"-ProductOrServiceProperty"* on the three properties involved)

At the end of the process, the 174 properties of FreeClass, were transformed into two data type properties, and 81 quantitative and 91 qualitative object properties in the resulting FreeClassOWL ontology.

The naming convention for the identifier of all FreeClass properties in FreeClassOWL employs the pattern *fc:P_ID*,

[9] `http://www.productontology.org/`

[10] `http://wiki.goodrelations-vocabulary.org/Documentation/Extensions`

where the *ID* portion is mapped to a *strong identifier* field in the source CSV files. The strong identifier is unique to the specific FreeClass property represented by the associated *fc:P_ID* OWL property.

There is one more property explicitly added to the FreeClassOWL ontology that did not fall into the approach described above. This is the annotation property *fc:hierarchyCode*. This annotation property stores the 8-digit numeric code of a FreeClass category and its purpose and intended use has already been addressed in the previous section.

All eight language descriptions are supported following the same approach outlined for the classes of FreeClassOWL scenario, via the properties *rdfs:label*. Additional comments are supported via *rdfs:comment*, with the exception that comments on properties were available only in German.

2.1.5 Domain and Range of Properties

As subproperties of one of the properties *gr:datatype-*, *gr:quantitative-*, or *gr:qualitativePOSP*, the domain of every *fc:P_ID* property of FreeClassOWL was set to the class *gr:ProductOrService*.

The range of the two FreeClassOWL properties subsumed by *gr:datatypePOSP*, should align to the range of an OWL data type property that is, an RDF literal or a simple XML Schema data type. The range of values observed in the original FreeClass schema for the two properties eligible as a OWL data type property, was in fact a string.

The range of the 81 FreeClassOWL properties subsumed by *gr:quantitativePOSP* is the class *gr:QuantitativeValue*. To comply with GoodRelations, such a property should specify a numeric value and the unit of measurement that the value represents. GoodRelations supports this requirement via the property *gr:hasUnitOfMeasurement*. The range of *gr:hasUnitOfMeasurement* is by convention a three-letter code from the UN/CEFACT [15] standard of units of measurement. The units of measurement found for the properties of the original FreeClass product classification, used an internal ad-hoc encoding. In the conversion process all codes found, were mapped to the corresponding UN/CEFACT standard common codes.

The range of the 91 FreeClassOWL properties subsumed by *gr:qualitativePOSP* is the class *gr:QualitativeValue*. The expected range of such a property is a predefined ontology individual. FreeClassOWL includes many of these individuals, which are discussed in detail in the next section.

2.1.6 Individuals of FreeClassOWL

The creation of ontology individuals was the last step of the transformation process of the FreeClass product classification. Another set of elements found in the FreeClass schema consisted of predefined values. In fact, there were 1,423 predefined values. As stated earlier, predefined values represent particular features applicable to certain properties of FreeClass and they delimit the range of possible values for such properties.

In FreeClassOWL and based on the guidelines for ontology extensions compliant with GoodRelations, predefined values were transformed into OWL individuals of the class *gr:QualitativeValue*.

The naming convention for the identifier of all individuals in FreeClassOWL follows the pattern *fc:V_ID*. The *ID* part is mapped to a *strong identifier* field in the source CSV files that is unique to the associated predefined value.

Analogous to previous cases, a textual description of predefined values was preserved via *rdfs:label* while other comments were stored via *rdfs:comment*. The former was available in eight languages, while the latter only in German.

2.2 Eurobau Utility Ontology

BauDataWeb required a few domain-specific conceptual elements particularly relevant in the building materials industry that are not defined in GoodRelations. For this purpose we developed a small ontology, the Eurobau Utility ontology. The new namespace URI of the ontology was set to `http://semantic.eurobau.com/eurobau-utility.owl#` and the namespace prefix assigned to it is *ebu:*. The ontology consists of three OWL individuals and two OWL object properties.

The three new individuals are instances of the GoodRelations class *gr:DeliveryMethod*, and represent three standardized procedures for transferring products to a destination location. They are: *ebu:DeliveryModeOwnFleetWithCrane*, *ebu:DeliveryModeRentalTruck*, and *ebu:DeliveryModeOwnFleetWithLiftgate*.

The two new object properties represent two different viewpoints on the distance between two locations. One refers to the distance measured in a hypothetical straight line (as if no obstacles existed between the points of origin and destination) and the other refers to the distance over an existing connecting path (i.e. a road suitable for driving). They are named *ebu:eligibleDeliveryDistance* and *ebu:eligibleDeliveryDistanceRoad*, respectively. The range of both properties is set to the class *gr:QuantitativeValue*, whose individuals are intended to be expressed as a floating point number measured in kilometers.

As stated in the introductory section, *findability* of building materials is an important aspect to be considered in this domain, and it can be severely affected by different delivery options or distances. A key goal of the Eurobau Utility ontology is to allow a more granular description of these characteristics so that they can be exploited by queries or other operations on data annotated with the ontology.

2.3 BauDataWeb Dataset

FreeClassOWL and the Eurobau Utility ontology introduced in the previous sections served as data models for the BauDataWeb dataset, the Semantic Web representation of Eurobau data (cf. Figure 1). The Eurobau dataset provides data related to the domain of construction materials. In particular, it comprises information about brands and companies, warehouse locations, product models, product model variants, availability, transportation and delivery, and geopositional data. The comprehensive dataset comprises 2,210,824 offerings, 67,218 product models and variants, 87 manufacturers/brands, 16 resellers, and 194 warehouse locations. In the following, we describe our approach of transforming the Eurobau datasets into RDF, as outlined in Figure 1.

2.3.1 Data Transformation

Unlike the source files of the ontologies, the instance data of Eurobau resides in a relational database hosted at Eurobau.com, spread across seven interlinked database tables.

As a preliminary step, the data of the Eurobau datasets has to be cleansed before being fed into the transformation process component. The source data suffers from poorly normalized tables, pragmatic design decisions that have evolved over time, and lack of standards compliance. For that purpose, some rules for improving the data quality right from the relational database are applied. The cleansing rules include stripping whitespaces from string values, overcoming inconsistent usage of units of measurement, or mappings between the non-standard use of country codes and respective codes in the ISO-3166-2[11] standard.

The result of applying custom data transformation rules on the instance data and of generating annotations based on the FreeClassOWL ontology leads to a considerable number of small RDF files. The produced files are intended for representing different types of resources as highlighted in Figure 2, which shows the relationships between files at the manufacturer and at the reseller side. Instead of labeling the nodes in Figure 2 according to the file names used in the dataset, we decided to assign meaningful labels to them (e.g. *Product Offering* instead of *Offering.rdf*). This allows us to accommodate the relationships that exist between the conceptual elements described in the files.

The involved conceptual elements comply with the Agent-Promise-Object principle, a basic structure of offers that underlies the GoodRelations ontology. The principle implies that an agent (*gr:BusinessEntity*, represented by *Company*) makes a promise (*gr:offers* with *gr:Offering*, represented by *Product Offering*) to transfer property rights (in our context, "to sell") for an object (*gr:ProductOrService*, again represented by *Product Offering*). In addition, *Company*, occurring twice in the architecture of Figure 2 (i.e. at manufacturer and at reseller side), defines information about the location from where the promise is made; whereas *Product Offering* contains the object that is promised, i.e. the product. Products can have model data attached to them using *gr:hasMakeAndModel*. The difference between products and product models is subtle but important. While a product embodies a tangible object, product models describe the make and model (or prototype, datasheet) of products with shared characteristics. Thus, only products are subject to physical transfer between business parties.

In the RDF materialization of the Eurobau data, we distinguish two types of models, namely product models (*Product Model*) and product model variants (*Product Model Variant*). Both are defined in terms of the GoodRelations *gr:ProductOrServiceModel* class. Nonetheless, the latter serves to describe variants of the former. Each variant inherits all data type, quantitative, and qualitative values that are defined by the related (or base) product model. Furthermore, *Product Models* are used to ensure the linkage to their manufacturers. Thus, the same property values do not have to be defined repeatedly which preserves space and conforms to the idea of best data modeling practice.

An example in Turtle syntax that we put online[12] demonstrates (a) how product features with corresponding values are attached to a product model variant, and (b) how links to the related product model (using *gr:isVariantOf*), the URL of the product image, and the related HTML page on the Eurobau portal are created.

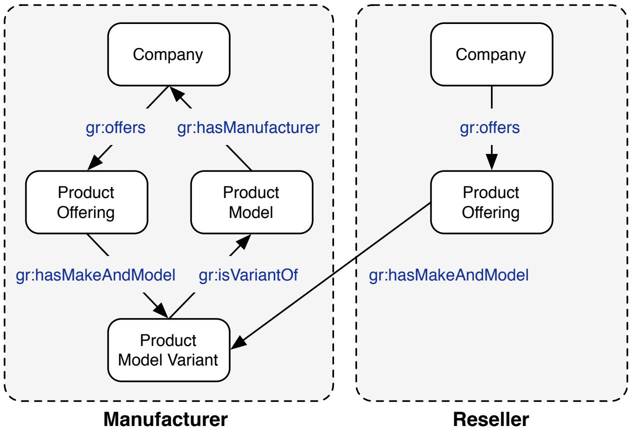

Figure 2: Architecture of the BauDataWeb dataset pointing out relationships between conceptual elements in the files for manufacturers and resellers

Characteristic	Amount
Models with FreeClass class	Base: 66 % Variants: 65 %
Models (variants only) with FreeClass properties	60%
FreeClass properties per model	5 (Median)

Table 2: Coverage of FreeClass in the Eurobau dataset

Overall, more than 65% of the product models are described using FreeClassOWL classes, whereas 60% of the model properties are assigned from FreeClassOWL (see Table 2). On average (median), a product model consists of five properties. The obtained Semantic Web dataset comprises ca. 2.2 million individual RDF/XML files plus a few large, compressed data dump files in N-Triples syntax that aggregate several RDF files into a compound data source. That aims to simplify and speed up the crawling and loading of the full dataset.

2.3.2 Data Management

One specific requirement of the project demanded the deployment of the RDF files on a subdomain[5] of Eurobau.com. The host of this subdomain consisted of a Microsoft Windows platform running Microsoft Internet Information Services (IIS) as Web server, and under the management of our industrial partner.

From that platform, a Semantic sitemap[13] [2] is served that provides an overview (site index) of sub-datasets, i.e. datasets that contain data that belong to a single manufacturer or reseller. The sub-datasets describe the locations from where the RDF files are served, i.e. organized and stored in meaningful folders on the server.

Besides the storage of physical files on a server, it is sensible to also provide a SPARQL endpoint for the publication and easy accessibility of the data. Most SPARQL implementations offer means for loading huge amounts of data. The complete BauDataWeb dataset has been made available in the following public SPARQL endpoint: `http:`

[11] `http://www.iso.org/iso/home/standards/country_codes.htm#2012_iso3166-2`

[12] `http://rdf-translator.appspot.com/convert/xml/n3/html/http://semantic.eurobau.com/at/baumit_com/articles/A_114507.rdf`

[13] `http://semantic.eurobau.com/sitemap.xml`

//linkeddata.uriburner.com/sparql.

Concerning data management tasks, a command-line script written in Python handles the full conversion. The script acts as the transformation process component of the BauDataWeb dataset as illustrated in Figure 1. It is invoked by specifying relevant companies together with the product models, product model variants, and, if applicable, product offerings, to be made available on the Web of Data. Cron jobs scheduled at regular intervals (e.g. once a month) help to keep the dataset up-to-date. The script indicates updates of datasets that just became available by simply changing the *lastmod* element of the sitemap. That means that clients are not supposed to download the full catalog of datasets, but only those datasets that have in fact changed.

3. EVALUATION

In this section, we evaluate the contributions of our work. Firstly, we validate the potential of our approach by addressing the limitations raised in the introduction. Secondly, we provide a fully-fledged demonstrator that provides a user-friendly interface for the consumption of BauDataWeb data. Lastly, we show that the deployment of the datasets complies with general Linked Data guidelines.

3.1 Exploring Novel Use Cases

In the introduction, we argued that the traditional approach based on data in a relational database suffers from several drawbacks that our proposal is able to solve. To shortly recap the limitations, they were given as the difficulty of (a) searching for building materials spanning more than one FreeClass category, (b) querying over all materials that meet a specific set of attributes or requirements, (c) finding manufacturers within certain geographic distances, and (d) combining construction material data with other data sources on the Web.

We can show that BauDataWeb can contribute to overcome these limitations from various angles. To address issues (a), (b), and (d), we crafted SPARQL queries that we made available online[14] as interactive queries (queries a, b and c), and which are subsequently explained at an abstract level. In order to solve issue (a), a SPARQL query can be defined that selects all building materials relative to one or more FreeClassOWL classes, for example by using UNION clauses. Even better yet, if the RDF store is capable of RDFS entailment, the taxonomy hierarchy in FreeClassOWL provided by the GenTax approach can be utilized to infer all materials defined as instances of classes and subclasses of a specific FreeClass category. This way, we are able to harness knowledge given by and deducible from the declared class axioms. For example, materials referring to different branches in the class hierarchy can be searched on the basis of inference techniques. For overcoming problem (b) instead, a SPARQL query based on basic graph pattern matching is sufficient that matches over triples including specific properties and attributes, optionally supplied with FILTER functions to express constraints such as "above 10 centimeters in length". To address issue (d), we again crafted a SPARQL query which includes data provided by external SPARQL endpoints using query federation introduced by SPARQL 1.1. It looks up DBPedia URIs for addresses contained in the BauDataWeb dataset. Unlike all data that is isolated in a relational database, as it is the case for Eurobau, structured data published on the Semantic Web by virtue of appropriate ontologies helps to unlock useful information that can be made available to other services for an immediate benefit. Therefore, the publication of Eurobau data at high granularity on the Web can be seen as an important step towards enabling better exploratory searches over Linked Open Data in the context of building and construction materials.

As issue (c) is concerned, it has actually been implemented for the demo application that is presented in the upcoming Section 3.2. The use case takes advantage of geopositional data of warehouse locations stored in the dataset.

3.2 BauDataWeb Demo Application

The demo application presented herein provides a user-friendly way of interacting with the BauDataWeb dataset. It combines the benefit of searching for an item classified according to FreeClassOWL with a straightforward way of presenting products grouped according to the companies (warehouses) offering the products. Companies are ordered by their vicinity to a custom location which can be specified by the user. The tool furthermore allows to filter query results by relevant properties from FreeClassOWL. The demonstrator is available online[15].

The entry point to the usage of the tool is to select a FreeClass code from a tree navigation representing the hierarchy of the industry standard. Next, all FreeClass properties related to the selected product category are gathered using a SPARQL query. Quantitative properties with upper and lower boundary values are presented in text fields and can be edited by the user. Similarly, eligible qualitative values are provided as drop-down lists to supply criteria to narrow down the result set. The final steps involve specifying optional search parameters and entering address information about the delivery location, which steps are well-supported and self-explanatory. The resulting SPARQL query is executed taking into consideration all parameters supplied with the input form.

Beyond the demonstrator, in Figure 3, we show a map that displays all warehouse locations in Austria and its surroundings. In order to obtain this map, we translated all geo-tagged company data in the RDF files to KML (Keyhole Markup Language), a XML format to encode geographic information supported by Google Maps. For the conversion we made use of the Geo2KML[16] Web service. A map with all warehouse locations that aligns with the excerpt from Figure 3 is available at http://goo.gl/maps/SpPNO.

3.3 Deployment of BauDataWeb

To deploy *all* deliverables of the BauDataWeb project, we adhered to available best practices not only on their development phase, but also on the publishing and deployment phase. If the former focused on the *production* of a valid, well-formed RDF data repository, the focus of the latter is set on enabling and facilitating a useful and practical *consumption* of this repository. As an evaluation benchmark to fulfill these goals, we aimed to comply with the four Linked Data principles [5]. In short: (1) use URIs as names to iden-

[14] http://www.ebusiness-unibw.org/tools/baudataweb-queries/

[15] http://www.ebusiness-unibw.org/tools/freeclass-search/

[16] http://graphite.ecs.soton.ac.uk/geo2kml/

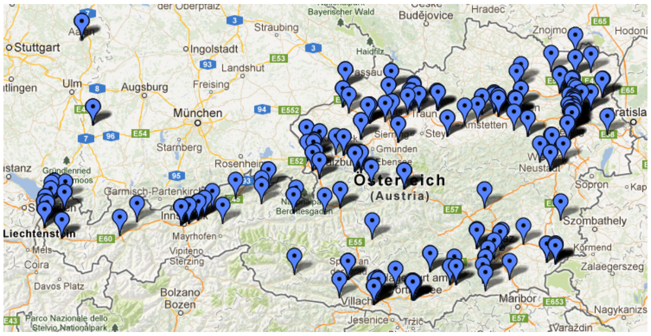

Figure 3: Warehouse locations displayed on a Google map

tify everything; (2) use HTTP URIs so that those names can be *dereferenced* (or looked up); (3) when someone looks up a URI, provide useful information (i.e. HTML or RDF/XML); and (4) include links to other URIs so that they can discover more things.

Principles (1) and (2) are satisfied provided that all data resources involved in BauDataWeb (e.g. offerings, product models, online Web documents...), use unique, *cool* URIs [14] and are *dereferenceable* via HTTP.

Compliance to principle (3) is supported by the configuration of the *content negotiation*[17] functionality of HTTP. In general terms, content negotiation allows to format or present the data of a Web resource differently, depending on the type of HTTP request made by the agent looking up the URI that links to such resource. In the context of BauDataWeb, content negotiation was configured to deliver the correct representation of FreeClassOWL (RDF/XML or HTML format), depending on the type of request.

Lastly, principle (4) acknowledges three types of RDF links in order to discover additional data about a Web resource. These are *relationship*, *identity*, and *vocabulary* links [5]. The RDF data of BauDataWeb includes vocabulary links, such as those required in order for both, the FreeClassOWL and BauData Utility ontologies, to specialize and extend elements in the GoodRelations ontology.

In addition to these four principles, other recommendations were applied to facilitate the consumption of the BauDataWeb repository, such as Semantic sitemaps [2] and Cross-Origin Resource Sharing (CORS) [16].

4. RELATED WORK

There are several topics closely related to our work on this project. They include, at a low level, closer to our methodology, (1) the re-engineering of non-ontological resources as OWL ontologies; (2) the transformation of various data sources into structured data in RDF format; and at a high level, as an overall solution, (3) additional existing use cases where Semantic Web technologies may have been applied to the e-commerce backbone of the construction industry marketplace.

The vast majority of existing approaches on the re-engineering of non-ontological resources into OWL ontologies are surveyed in [17]. One of such approaches is the GenTax methodology [7], which enables the conversion of product classification standards with potentially inconsistent semantic relations among their classes, into fully compliant OWL ontologies. The alignment between the GenTax approach and the GoodRelations ontology is also well-defined, which simplifies the creation of ontologies that extend and comply with the latter. The FreeClass classification lay at the intersection of all of these characteristics and thus, as discussed in Section 2, GenTax was applied for building the FreeClassOWL ontology.

For the transformation of heterogeneous data sources into RDF data, some of the most prominent tools were examined such as, RDF Refine[12], Any23[18] (Anything to triples), XLWrap [11], RDF123 [3], D2RQ [1], or R2RML[19]. At the same time, we had to accommodate certain requirements and constraints specific to the BauDataWeb project, which included (1) traces of proprietary information in the data sources; (2) certain inconsistencies on the data sources that hindered the level of automation; (3) deployment of deliverables on a Windows platform under the management of our industrial partner; (4) minimal performance impairment and run-time intrusion into the relational databases that support the Eurobau.com portal by the new semantic technologies; or (5) support for a command-line interface for executing all data transformation processes. These requirements deemed unfeasible for different reasons in each case, the reuse of the tools examined, which led us to the development of custom, *fit for purpose*, conversion scripts and utility tools.

Lastly, if to the application of e-commerce solutions for the construction industry, we add the adoption of Semantic Web technologies, the space for finding related work can get fairly reduced. Yet, several examples emerged as part of the First Intl. Workshop on Linked Data in Architecture and Construction held in 2012[20]. All works from the event, showcase compelling scenarios where the use of OWL ontologies and Linked Data could assist to overcome some of the reoccurring interoperability issues found in the architecture, engineering and construction (AEC) sector. And although this domain overlaps significantly with that of BauDataWeb, their focus leans towards the management of information exchange across the many actors involved in the execution of AEC projects. A topic much broader than the development of new e-commerce solutions for the construction materials marketplace, the core motivation of this project.

The conclusion at the end of this review indicates that BauDataWeb is, to the best of our knowledge, the first use case that relies on open standard Semantic Web technologies, to bring new, more granular and better articulated search, exploration and exploitation capabilities (both as a stand-alone solution or as part of the Web of Data), to a significant data repository of the building and construction materials domain, a prominent B2B/B2C (Business-to-Business/-Consumer) e-commerce marketplace.

5. CONCLUSIONS

In the context of building and construction materials, the Eurobau portal already provides means to search for granular descriptions of products, suppliers and warehouse locations. Yet, the technical limitations of the traditional Web

[17] `http://www.ietf.org/rfc/rfc2616.txt`

[18] `http://any23.org/`

[19] `http://www.w3.org/TR/r2rml/`

[20] `http://multimedialab.elis.ugent.be/ldac2012/`

and its heterogeneous data silos attributable to relational databases, hamper the creation of advanced usage scenarios involving better search and browsing experiences, typically enabled by structured data, enrichment with other data sources, and capabilities to infer additional knowledge.

In this paper, we presented the results of a project carried out with an industry partner to publish the comprehensive Eurobau dataset, which covers the Austrian building and construction materials industry to a large extent, in the Web of Linked Data on the basis of the GoodRelations vocabulary for e-commerce. Our most notable achievements include the development of the FreeClassOWL ontology, the Eurobau Utility ontology, and the BauDataWeb RDF dataset, which was derived from the comprehensive Eurobau dataset, giving rise to more than 88 million triples of real business data with a high domain density. One of the core challenges in the context of the resulting BauDataWeb dataset was to ensure easy data access and reliable, efficient data management.

In order to evaluate our work, we showcased novel use cases in SPARQL that previously, based on past technical capabilities, would not have been easily possible. Furthermore, we presented a user-friendly online demonstrator for searching within the BauDataWeb dataset and displayed warehouse locations on a Google map relying on geo coordinates in the dataset. We showed that the deployment of BauDataWeb complies with best practices of publishing Linked Data sources on the Web.

To put our solution into context, we contrasted it to related works in the fields and domains of construction industry, ontology creation from non-ontological resources, and general approaches to transform datasets into RDF.

In summary, our approach facilitates novel search and product matchmaking scenarios in the domain of building and construction materials.

6. ACKNOWLEDGMENTS

The work on BauDataWeb was mainly funded by the Austrian Research Promotion Agency (FFG) under project grant "icontent.document" (no. 819773) and Otto Handle, Inndata Datentechnik GmbH, Pacherstraße 24, 6020 Innsbruck, Austria, handle@eurobau.com. We would also like to thank the support of: the German Federal Ministry of Education and Research (BMBF) under the KMU Innovativ program as part of the Intelligent Match project (FKZ 01IS10022B); and the EUREKA and European Commission under the Eurostars program as part of the Ontology-based Product Data Management (OPDM) project (FKZ 01QE1113D).

7. REFERENCES

[1] C. Bizer and A. Seaborne. D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs. In *Poster Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan, 2004.

[2] R. Cyganiak, H. Stenzhorn, R. Delbru, S. Decker, and G. Tummarello. Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web. In *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*, pages 690–704, Tenerife, Canary Islands, Spain, 2008.

[3] L. Han, T. Finin, C. S. Parr, J. Sachs, and A. Joshi. RDF123: From Spreadsheets to RDF. In *Proceedings of the 7th International Semantic Web Conference (ISWC 2008)*, pages 451–466, Karlsruhe, Germany, 2008.

[4] O. Handle. *Konzeption und Realisierung eines branchenübergreifenden Produktklassifikationssystems für das Bauwesen unter Nutzung der produktspezifischen Fachkompetenz der Baustoffindustrie*. Master's thesis, MCI Management Center Innsbruck, 2007.

[5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 1st edition, 2011.

[6] M. Hepp. eClassOWL: A Fully-Fledged Products and Services Ontology in OWL. In *Poster Proceedings of the 4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, 2005.

[7] M. Hepp. Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards. *Journal on Semantic Web & Information Systems (IJSWIS)*, 2(1):72–99, 2006.

[8] M. Hepp. 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 (EKAW 2008)*, pages 332–347, Acritezza, Italy, 2008.

[9] M. Hepp and J. de Bruijn. GenTax: A Generic Methodology for Deriving OWL and RDF-S Ontologies from Hierarchical Classifications, Thesauri, and Inconsistent Taxonomies. In *Proceedings of the 4th European Semantic Web Conference (ESWC 2007)*, pages 129–144, Innsbruck, Austria, 2007.

[10] P. L. Joskow. Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence. *Journal of Law, Economics, & Organization*, 4(1):95–117, 1988.

[11] A. Langegger and W. Wöß. XLWrap - Querying and Integrating Arbitrary Spreadsheets with SPARQL. In *Proceedings of the 8th International Semantic Web Conference (ISWC 2009)*, pages 359–374, Chantilly, VA, USA, 2009.

[12] F. Maali and R. Cyganiak. RDF Refine - a Google Refine extension for exporting RDF. Available online at `http://refine.deri.ie/`.

[13] P. Morville. *Ambient Findability: What We Find Changes Who We Become*. O'Reilly Media, 1st edition, 2005.

[14] L. Sauermann, R. Cyganiak, D. Ayers, and M. Völkel. Cool URIs for the Semantic Web. W3C Interest Group Note, W3C, December 2008. Available online at `http://www.w3.org/TR/cooluris/`.

[15] United Nations Economic Commission for Europe (UNECE). Recommendation No. 20: Codes for Units of Measure Used in International Trade. Recommendation, UN/CEFACT Information Content Management Group, 2006.

[16] A. van Kesteren. Cross-Origin Resource Sharing. W3C Candidate Recommendation, W3C, January 2013. Available online at `http://www.w3.org/TR/cors/`.

[17] B. Villazón-Terrazas. *A Method for Reusing and Re-Engineering Non-Ontological Resources for Building Ontologies*. IOS Press Inc., 2012.