

A Bird's-Eye View of euBusinessGraph: A Business Knowledge Graph for Company Data

Dumitru Roman¹, Vladimir Alexiev², Javier Paniagua³, Brian Elvesæter¹, Bjørn Marius von Zernichow¹, Ahmet Soylu¹, Boyan Simeonov², and Chris Taggart⁴

¹ SINTEF AS, Norway
{firstname.lastname}@sintef.no
² Ontotext, Bulgaria
³ SpazioDati, Italy
⁴ OpenCorporates, UK

Abstract. This poster paper provides an overview of euBusinessGraph—a business knowledge graph for basic company data, together with related artefacts (datasets, ontology, and infrastructure), and its use for creating a prototype for a data marketplace for basic company data. euBusinessGraph was developed by aggregating, linking, and provisioning data from several distributed data sources.

Keywords: Company data · Knowledge Graph · Ontology · Linked data.

1 Introduction

Many data value chains, in various sectors such as procurement and marketing and sales, depend upon company data. Integrating company data from various authoritative and non-authoritative sources is challenging due to heterogeneity and complexity of the data and the lack of generally agreed upon semantic descriptions. Several initiatives (e.g., the Global Legal Entity Identification System (GLEIS)¹) have been established to harmonise and increase the interoperability of corporate data across national borders. Such initiatives are mostly fragmented across borders, limited in scope and size, and siloed within specific business communities. As a step in addressing this challenge, governments are increasingly publishing open data about firmographics. Unfortunately, these datasets are not yet fully harmonised and interoperable because data differs widely in semantics and data formats are often poorly accessible and documented. In other words, the harmonisation of company data is far from a solved problem.

To address the aforementioned issues, we follow the established ontology-based approach for harmonising and integrating data. We built a business knowledge graph (KG)—euBusinessGraph—by aggregating, linking, and provisioning basic company data from several distributed data sources. In this poster paper, we

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹ <https://www.gleif.org>

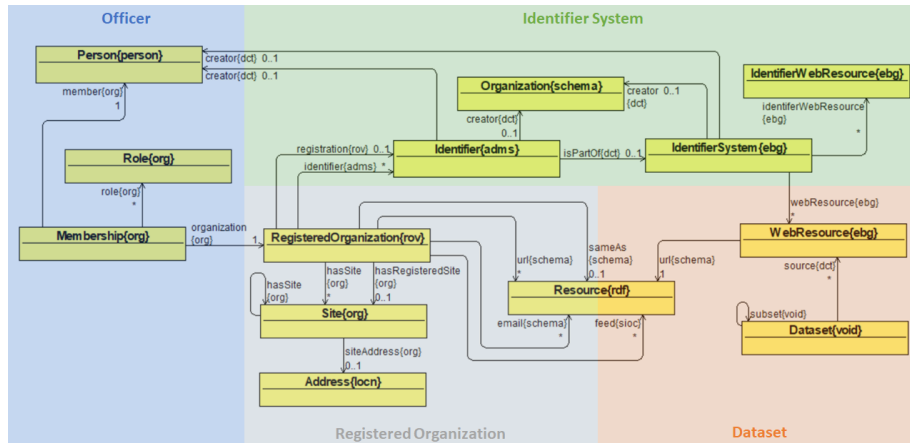


Fig. 1. euBusinessGraph ontology overview – main classes and their relationships.

provide an overview of the underlying datasets, the ontology developed to help with data integration, the data ingestion process, data provisioning infrastructure, and the use of KG for creating a prototype data marketplace for sharing company data.

2 Knowledge Graph Provisioning

The KG was built by aggregating data from four data providers, mapping and translating them into RDF with respect to an ontology and storing and publishing data in a triple store following Linked Data and Semantic Web principles. The data providers for the KG were OpenCorporates², SpazioDati³, Brønnøysund Register Centre⁴, and Ontotext⁵. The data made available by the data providers originally came from both official sources, such as national business registers, and unofficial sources, such as the corporate web.

We developed an ontology—the euBusinessGraph ontology [5]—to represent the basic company information, since existing solutions either insufficiently covered basic company information or were too complex due to many ontological commitments. We applied common techniques recommended by well-established ontology development methods [1]. We used a bottom-up approach by identifying the scope and user group of the ontology, requirements, and ontological and non-ontological resources. Data models of our data providers were one of the guiding elements for the ontology development, since there was a need to harmonise and integrate data models with different sets of attributes, different

² <https://opencorporates.com>

³ <http://spaziodati.eu>

⁴ <http://www.brreg.no>

⁵ <https://www.ontotext.com>

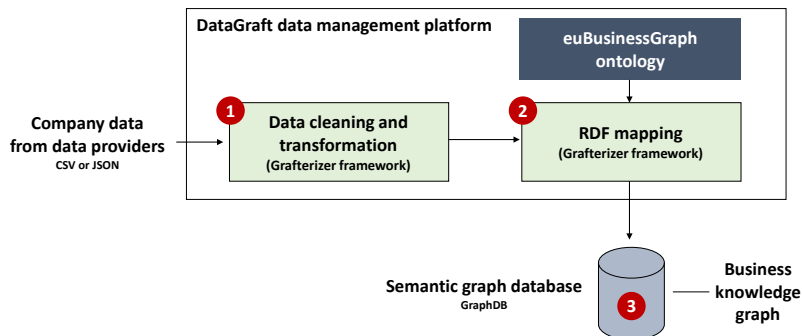


Fig. 2. The data provisioning process for the euBusinessGraph Knowledge Graph.

representations for the same entity and in some cases close but not entirely similar semantics. The resulting euBusinessGraph ontology is composed of 20 classes, 33 object properties, and 56 data properties. Fig. 1 provides an overview of the ontology, depicting the main classes and their relationships (i.e., object properties). Registered organisations are the main entities for which information is captured in the ontology. The main classes include `RegisteredOrganisation`, `Identifier`, `IdentifierSystem`, `Person`, and `Dataset`. Three types of classifications are defined in the ontology for representing the company type, company status, and company activity. These are modelled as SKOS concept schemes. External vocabularies and ontologies were used as needed, for example, the W3C Organisation ontology, W3C Registered Organisation Vocabulary (RegOrg), SKOS, Schema.org, and Asset Description Metadata Schema (ADMS) were reused. The ontology, datasets, and examples are released as open source and are available on GitHub⁶.

We devised a data mapping approach to convert company data from CSV and JSON sources into RDF conforming to the euBusinessGraph ontology using the DataGraft platform [4]. After transforming each dataset into RDF, the resulting data was published to one named graph for each data provider in the GraphDB enterprise semantic graph database. Fig. 2 illustrates the data provisioning process and the tools and services used to generate the business knowledge graph. Grafterizer [6], ASIA [2] and ABSTAT [3] were used to clean, transform, enrich, and convert tabular data to RDF as part of the business knowledge graph construction. The repository hosted on GraphDB contains more than 1,4 Billion RDF triples of company data.

3 Knowledge Graph in Use

We developed a data marketplace prototype on top of the resulting KG. The main motivation behind the development of a data marketplace for basic company data

⁶ <https://github.com/euBusinessGraph/eubg-data>

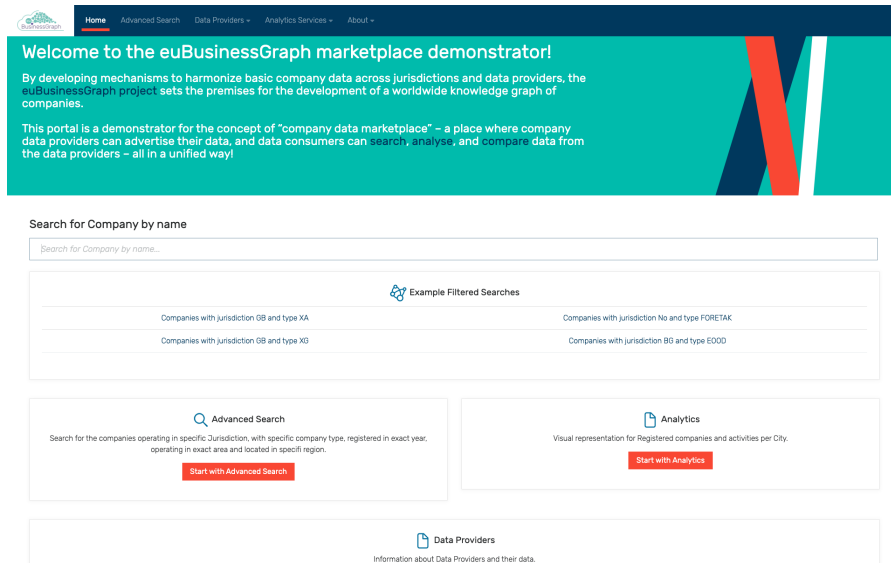


Fig. 3. euBusinessGraph marketplace homepage.

is the democratisation of the company information market, currently dominated by a few large international players that create a market barrier for smaller company data providers. The intention of the marketplace is to enable such smaller players to join a common ecosystem to promote their data offerings, and for data consumers to have a central point where they could easily compare company data offerings. A public prototype of the data marketplace application is available online⁷ and a screenshot of its homepage is shown in Fig. 3.

The available data in the marketplace application includes the most central attributes that reflect how the ontology can be used to describe the semantic relations of company data. Each data provider URI in GraphDB is related to a dataset description that describes the data being offered in the marketplace by inserting `void:inDataset` for each `rov:RegisteredOrganization` in the graph database. The marketplace includes functionality for full-text advanced search and detailed faceted search for exploration of the company knowledge graph. Furthermore, the marketplace offers analytics services such as data aggregation and visualisation (e.g., company activities per city), search for company news articles, and search for company events. The ontology was used in the marketplace to realise scenarios such as:

- **Company search:** *Find a specific company by displaying a page that describes available attributes of the company.* The ontology enables search for detailed company information from different providers, and facilitates data

⁷ <http://marketplace.businessgraph.io>

provenance, as the specific company data from data provider can be traced back to its sources.

- **Advanced company search:** *Find how many companies are in a certain jurisdiction, active or inactive, registered in a certain year, with a certain type, in a certain location or are operating within a certain economic activity.* This scenario is covered by allowing search for companies by certain criteria or facets and dynamic filtering of results. The search functionality of the marketplace demonstrates how the semantic model enables a uniform way of harmonising and representing hierarchical facets for geographical location and economic classification.
- **Analytics related to company data:** *Find out how many companies are registered per year in a specific country and city and are operating in a specific location.* The marketplace application provides the ability to get basic statistics about the company data in the knowledge graph. A bar chart visualization filters information by country, city and activity and gives the user a visual representation of the data.

4 Conclusions

In this poster paper we argued for the importance of harmonised basic company data as a key enabler for different value chains in various sectors that depend on company information and described the euBusinessGraph approach for harmonising basic company data. Our approach is based on a lightweight ontology for aggregating, linking, provisioning, and analysing basic company data. The ontology was developed following best practices for ontology development and published following Linked Data and Semantic Web principles.

Acknowledgements

The work reported in this paper is partly funded by the EC under the H2020 euBusinessGraph project (Grant nr. 732003).

References

1. Corcho, O., et al.: Ontological Engineering: Principles, Methods, Tools and Languages, pp. 1–48. Springer (2006)
2. Cutrona, V., et al.: ASIA: a tool for assisted semantic interpretation and annotation of tabular data. In: Proc. of the ISWC 2019 Satellite Tracks. CEUR-WS.org (2019)
3. Principe, R.A.A., et al.: ABSTAT 1.0: Compute, manage and share semantic profiles of RDF knowledge graphs. In: Proc. of ESWC 2018 Satellite Events. Springer (2018)
4. Roman, D., et al.: Datagraft: One-stop-shop for open data management. Semantic Web **9**(4), 393–411 (2018)
5. Roman, D., et al.: The euBusinessGraph Ontology: a Lightweight Ontology for Harmonizing Basic Company Information. Semantic Web **under review** (2020), <http://www.semantic-web-journal.net/system/files/swj2421.pdf>
6. Sukhobok, D., et al.: Tabular data cleaning and linked data generation with grafterizer. In: Proc. of ESWC 2016 Satellite Events. Springer (2016)