# Building Eurostat Knowledge Graph

Alexandros Vassiliades[1][a], Nick Bassiliades[1][b], Georgios Meditskos[1][c] and Kimon Spiliopoulos[2]

[1]*School of Informatics, Aristotle University of Thessaloniki, Greece*
[2]*Quantos Statistics and Information Systems, Athens, Greece*

Keywords: Ontology, Eurostat, OECD, Knowledge Graphs.

Abstract: The evolution of Knowledge Graphs (KGs) has encouraged developers to create more and more context related KGs. This advance is extremely important because Artificial Intelligence (AI) applications can access domain specific information in a machine understandable format. In this paper, we present the conceptual model and semantics of the OWL ontology developed to capture information about the Eurostat website. The KG also contains some knowledge from the Organisation for Economic Co-operation and Development (OECD) website. We also describe how we constructed the ontology schema in order to capture all the data in Eurostat and some of the data in OECD, such as, articles, datasets, and internal connections between them, among others. Moreover, we show how we populated the KG with an automated process, resulting into a KG with more than 820K triples.

## 1 INTRODUCTION

The evolution of Knowledge Graphs (KGs) in the last twenty years allowed developers to construct context related KGs (i.e., KGs that can be used only in specific environments). The creation of context related KGs seems to be the next step for allowing KGs to become the main knowledge representation mechanism for the Web. Our focus is on representing statistical concepts through a high-level representation with statistical articles and statistical datasets, among others. The idea of statistical KGs has been given great attention recently and even generic techniques on how to create a statistical KG were developed (Otte et al., 2022; Smith et al., 2007; Capadisli et al., 2015).

The main idea of this paper is to build innovative semantic approaches to improve data processing and data services, for the information in the Eurostat website[1]. The three main objectives which were pursued:

- Increasing the discoverability and accessibility of data available for analytical purposes.

- Strengthening Eurostat position within the Commission as a provider of statistical data and services for its internal users.

- Improving the methods for extracting information from unstructured data sources – especially data available on the web.

To realise the above, data processing and discovery services need to be further developed. In the current scenario, Eurostat publishes its (open) statistical datasets on its own portal, along with descriptive metadata, which enable users to access and consult data, but also understand the content and the meaning of the data, the expected quality, the formats, the data collection method etc. To further support users when accessing and taking benefit of these data – in other words: searching, discovering, combining and analysing – the deployment of advanced data services, including faceted search, guided query builders, as well as services for data exploration and visual data browsing, needs to be deployed. The aforementioned objectives are part of the *NLP4Stat* project[2] which is an attempt to translate the information in the Eurostat website, into a Linked Open Data counterpart and demonstrate the advantages of doing so.

The problem we are addressing in this paper is how information from websites can be translated into a KG. We rise the question if a generic method could be developed for this purpose. We also support that we have reached a point where, information from very big websites as the one of Eurostat or Organisation for

---

[a] https://orcid.org/0000-0003-4569-503X
[b] https://orcid.org/0000-0001-6035-1038
[c] https://orcid.org/0000-0003-4242-5245
[1] https://ec.europa.eu/eurostat

[2] https://github.com/eurostat/NLP4Stat

Economic Co-operation and Development (OECD)[3], can henceforth be translated into KGs.

In this paper we present the *Eurostat KG*, which constitutes one of the most complete and large KGs (i.e. in regards to quantity of triples), for representing statistical concepts through a high-level representation with statistical articles and statistical datasets. More specifically, the Eurostat KG contains most of the information from the Eurostat website and some information from the OECD website, information about articles, datasets, interconnections between them, relations with external sources, and information for various classifications for the articles and datasets, among others. Therefore, we show how we developed the schema of the KG, and how we captured the data that we used to infer the schema. At this moment the KG contains approximately 820K triples.

The main contributions of this paper are: (i) the Eurostat KG which is one of the biggest, in regards to quantity of triples, and complete KGs for the domain of representing statistical concepts through a high-level representation with statistical articles and statistical datasets, and (ii) an integration mechanism that takes the information from the Eurostat and OECD website (after the data has been scraped), and populates the KG. For better understanding of the importance of the Eurostat KG, we demonstrate a use case in which we can present how the information from the KG can assist a user in a real-life scenario.

The rest of this paper is organized as follows. In Section 2, we present related work on similar KGs. Next, in Section 3 we describe the data we used to extract the conceptual schema of the ontology and to populate the respective KG, we present the ontology schema, how it was populated, and in the end we present a use case scenario. We conclude the paper, with Section 4, where we give a discussion over the results and some future work directions.

## 2 RELATED WORK

In this section we present some KGs in the area of statistics, which can be considered close to our work.

The *Statistical Data and Metadata eXchange* (SDMX) (Sembiring and Uluwiyah, 2015), can be considered as the closest relative to our study. SDMX aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries. Basically, SDMX offers a new standard format in the data dissemination activ-

ities particularly in the exchange of statistical data and metadata though the Web. The main difference is that SDMX is an information exchange protocol for statistical data, while our Eurostat KG is maybe one of the most complete KGs which contain concepts and knowledge about statistics. Similar is the Core Ontology for Official Statistics[4] (COOS), as its main purpose is to serve as an integration model for the core set of ModernStats (Franck et al., 2018) standards backed by elements of well-known standard vocabularies.

STATO[5] is a general-purpose statistics KG. Its aim is to provide coverage for processes such as statistical tests, their conditions of application, and information needed or resulting from statistical methods, such as probability distributions, variables, spread and variation metrics. STATO also covers aspects of experimental design and description of plots and graphical representations commonly used to provide visual cues of data distribution or layout and to assist review of the results. The difference with our Eurostat KG, is that we offer more knowledge, apart from information for statistical articles and dataset which contain the information refered in STATO. Moreover, we offer various classifications for the articles and the datasets, and connections between articles and articles with dataset, that could reveal information for the statistical methods that was not captured in the first place.

STATO has been developed to interoperate with other Open Biological and Biomedical Ontologies (OBO) Foundry ontologies (Smith et al., 2007), hence relies on the Basic Formal Ontology (BFO) (Arp et al., 2015) as a top level KG and uses the Ontology for Biomedical Investigations (OBI) as a mid-level KG (Bandrowski et al., 2016). Therefore, another group of KGs that can be considered in the area of statistical KGs, are the OBI ontology and the Ontology of Biological and Clinical Statistics (OBCS) (Zheng et al., 2014). These KGs contain some statistical concepts for the clinical domain. It is clear that we do not offer the same information with the aforementioned KGs, as we offer a KG that represents statistical concepts through a high-level representation with statistical articles and statistical datasets.

## 3 EUROSTAT KNOWLEDGE GRAPH

Figure 1, displays an overview of the architecture as well as the flowchart of the information among its parts. The architecture for building the KG consists

---

[3]https://www.oecd.org/

[4]https://linked-statistics.github.io/COOS/coos.html
[5]http://stato-ontology.org/

of three parts: the Virtuoso OpenLink Server (OS)[6] that hosts the content (relational) database and the knowledge database (knowledge graph), a set of KG-exploiting applications and the Python environment where all the scripts for scraping information from the Eurostat and external websites and transforming relational content to a knowledge graph are executed.

The Python environment is used to extract, manipulate and store the data, using a connection to the Virtuoso database. Connecting and querying from Python to Virtuoso is done at different stages: (i) to interact with the Content database, from the scraper or the use case applications, (ii) to communicate with the Eurostat KG, when populating it from the Content database or when the Eurostat KG is augmented by the enrichment mechanisms.

Figure 1 exhibits the information and functionality workflow. Content from the Eurostat and external websites (e.g. OECD) is scraped using Scrapy within Python scripts, and the result of the scraping populates the Content database, which is a relational one. The Eurostat KG is then populated by extracting data from the relational tables, using SQL, and creating RDF triples, according to the Eurostat ontology. The content and knowledge DBs are enriched (beyond the scraped content) using various mechanisms. Most of them are inspired by the Use Cases (Section 3.5) and involve various NLP and statistical methods. Furthermore, the KG has been enriched via some SPARQL CONSTRUCT queries that interconnect articles with Eurostat and OECD themes (see Figure 2). Finally, the user interacts with the KG either via applications build on top, using Python or other environments, or directly via the SPARQL endpoint[6].

## 3.1 Inventory of Eurostat Knowledge and Information Resources

The data was scraped from the Eurostat[1] and OECD[3] sites and are stored into the (private) content DB, which contains 66 different tables, with information about categorizations, topics, terminology, named entities (i.e., words that refer to real life entities), links between, glossaries, and information about statistical articles and datasets, among others.

The most important information exists in the articles and datasets that exist in the SQL database. The data about the articles which was scraped from the Eurostat website was separated into two big categories, the first one is called *Statistics Explained Articles* and the second one *Glossary Articles*.

- Statistics Explained Articles are official Eurostat

articles, presented in the Eurostat website containing statistical topics in an easily understandable way. Together, the articles make up an encyclopedia of European statistics for everyone, completed by a statistical glossary clarifying all terms used and by numerous links to further information and the latest data and metadata, a portal for occasional and regular users. An example of such an article related to *agriculture* can be found here[7].

- Glossary Articles cover all statistical and general terms in Statistics Explained in need of a definition or explanation. Because it is quite large, it may be easier to consult instead one of the focused thematic glossaries from the clickable overview below; they are organised according to statistical themes, preceded by a list of abbreviations and further supplemented by special-topic glossaries. An example of such an article related to *social protection* can be found here[8].

Currently there are 892 Statistics Explained Articles and 1314 Glossary Articles. For each one of them the content DB contains their title, abstract, and paragraphs. Moreover, it contains references between them. For instance, an article for *agriculture* will point to other similar articles that are related to agriculture, either glossary or statistics explained articles.

The content DB also contains information about the datasets, such as, information about the taxonomy, the titles, and the url that points to the data of the dataset. An example of a dataset related to *consumers - monthly data* can be found here[9].

## 3.2 Inside Eurostat Knowledge Graph

In this subsection we will analyze in detail the purpose of each class and property in the KG. Currently, there are 1856 classes where the 1811 classes represent datasets, 37 properties - either object type or data type properties, 307419 explicit and 827395 implicit triples. The source files of the KG can be found here[10]. In Figure 3, we can see the upper part of the ontology scheme. Notice that when we constructed the schema of the KG, ontology engineering methods where considered (Iqbal et al., 2013; Kendall and McGuinness, 2019). One can notice that there are 4

---

[6]http://lod.csd.auth.gr:8890/conductor/

[7]https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Performance_of_the_agricultural_sector

[8]https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Social_protection

[9]https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&file=data%2Fei_bsco_m.tsv.gz

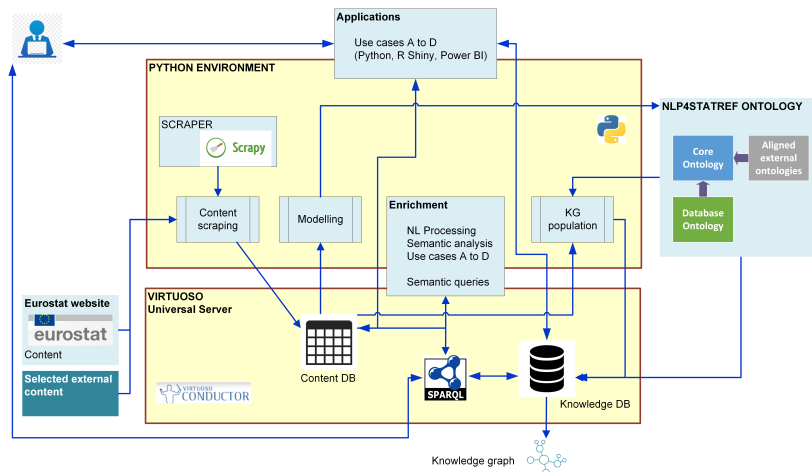[10]https://github.com/eurostat/NLP4Stat

Figure 1: Overview of the method to create the Eurostat KG.

```
construct {
    ?s estat:hasOECDTheme ?OECDTheme .
}
where {
    ?s a estat:Article .
    ?s estat:hasEurostatTheme ?ESTheme .
    ?ESTheme estat:relatedTheme ?OECDTheme .
}
```

Figure 2: An example of a SPARQL CONSTRUCT query that interconnects articles with Eurostat and OECD themes.

main classes in the ontology scheme, the *Glossary-Term* class, the *Content* class, the *Reference* class, and the *Classification* class. The namespaces chosen for this ontology are:

- estat: for the Eurostat ontology entities (i.e., classes and properties).

- estatdata: for the instances of the ontology that were extracted from the Eurostat website.

- oecd: for the instances of the ontology that were extracted from the OECD website.

The GlossaryTerm class refers to the various glossaries that exist in the Eurostat and OECD websites. This class has the subclasses:

(i) The class *FrequentTerm* which contains all the frequent terms that exist in Eurostat and OECD articles, discovered during the topic modelling enrichment task.

(ii) The class *CODEDTerm* which contains information for the Eurostat terms, i.e. it contains information about their title, abstract, content, url, interconnection with other terms, the date they were created and/or updated, and their related theme (themes is a classification scheme of Eurostat).
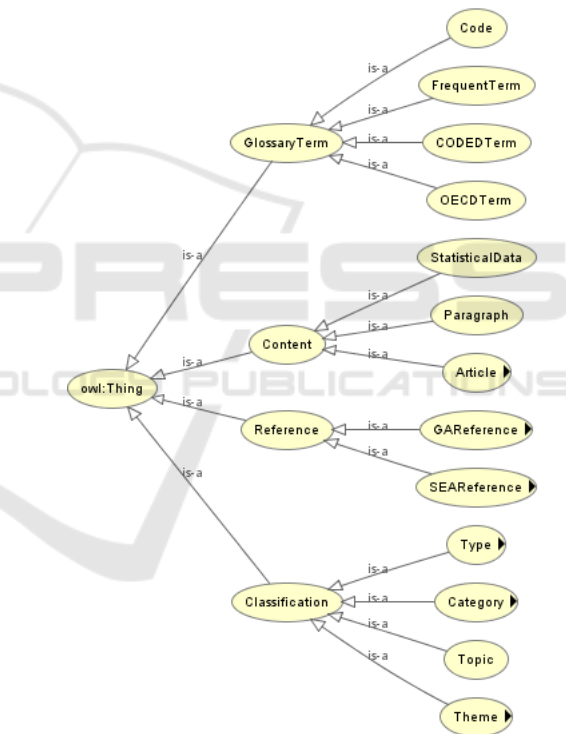


Figure 3: Upper part of the ontology scheme.

(iii) The class *OECDTerm* which contains information for the OECD terms, i.e. information about their title, abstract, content, url, interconnection with other terms, date they were created and/or updated, and related theme (themes is a classification of OECD). Notice that in both cases terms are descriptions of some statistical concepts.

(iv) The class *Code* contains information about the codes given to the datasets (i.e., unique ids).

GlossaryTerm instances are connected to each other through the property *relatedTerm*, mostly due to their common theme(s). Moreover, the other classes are connected though the property *hasGlossaryTerm* with the class GlossaryTerm, which is a super-property with sub-properties *hasCode*, *hasCOD-EDTerm*, *hasOECDTerm*, and *hasFrequentTerm*.

Terms are descriptions of some statistical con-cepts. Themes are a classification that characterizes the articles, given by Eurostat. Moreover, OECD themes are related with Eurostat themes, which cre-ates a linking between OECD and Eurostat articles. Codes are the unique identity codes that Eurostat gives to its datasets. Frequent terms are words that refer to named entities (i.e., words that refer to real-life entities, such as locations, persons, etc.).

In Figure 4, one can see a more detailed analysis of the class *Reference*. The class Reference is further analyzed to the classes *SEAReference* which are the references related to Statistics Explained Articles, and the class *GARerefence* which are the reference related to the Glossary Articles. Each reference can be either an internal reference, which means that it can point to another article or dataset inside Eurostat or OECD, or external which means that it can point to an external source, for example Wikipedia[11]. The leaf classes that we can see both for SEAReference and GARerefence are a classification of the type of reference given by Eurostat, based on the type of the source that the ref-erence points to. For example, the *Legislation* sub-class of SEARerefence refers to external or internal links that contain legal information.

We provide two examples with an external (Figure 5) and an internal (Figure 6) reference for an Statistics Explained Article to ease the understanding of these type of reference. More specifically, Figure 5 shows how an article from the Eurostat website, is related through a reference with an external URL (i.e., a URL that is not part of the Eurostat website). On the other hand, Figure 6 shows how an article from the Eurostat website, is related through a reference with another article from Eurostat website. Notice the example is generic and no specific ids are given to the instances.

Figure 8 depicts a detailed analysis of the class *Content*. The Content class is one of the most im-portant classes of the ontology as it represents knowl-edge about the Statistics Explained and Glossary Ar-ticles. Additionally, it represents knowledge about the datasets. For the articles, we represented their titles, their abstract, their content, their url in the Eurostat website, their internal or external relations, and the dates that they were created and updated. For the Statistics Explained Articles we have also represented

---

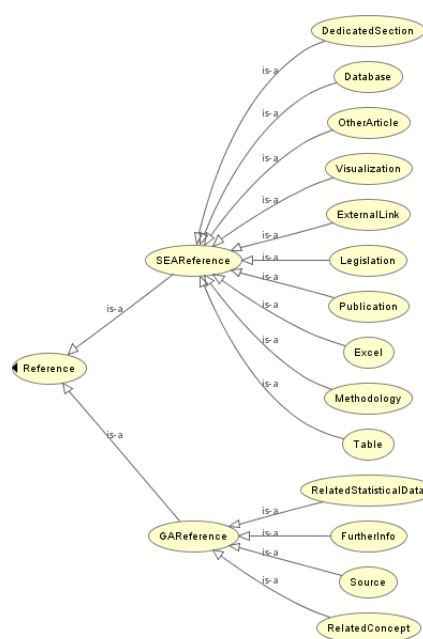[11]https://en.wikipedia.org/wiki/Main_Page

Figure 4: Analysis of the Reference class.

```
estatdata:SEArticleID1 estat:hasReference estatdata:ReferenceID.
estatdata:ReferenceID estat:hasURI estatdata:SEArticleID2.
estatdata:SEArticleID2 estat:hasURL "http/.../"^^xsd:anyURI.
```

Figure 5: Format for external reference.

the knowledge that exists for their paragraphs, and if the article are considered by Eurostat as a background article or not. Background articles is a classification that Eurostat gives to its Statistics Explained Articles considering their importance (i.e., the quality of in-formation they contain). On the other hand, for the datasets we created a taxonomy where the leaf nodes are the datasets, and the intermediate are categories that Eurostat has asserted to its datasets. Figure 7, shows part of the datasets taxonomy.

Apart from the taxonomy the datasets have infor-mation about their title, their description, their code (a unique id that Eurostat gives to each dataset), and their url in the Eurostat website.

Finally, Figure 9 shows the detailed analysis of the class *Classification*. The Classification class, mostly contains information about the various classifications that Eurostat assigns to its entities. For instance, the class *Category* represents information about the cate-gorization that Eurostat assigns to its articles. Similar is the class *Topic* which represents information about the topics that the articles have. The class *Type* has information about the type of information that an ar-ticle contains. For example, if is lexical or not (i.e., it contains text or some equations). The class *Theme* represents information about the Eurostat and OECD themes. The themes are also a classification that char-

```
estatdata:SEArticleID
    estat:hasReference estatdata:ReferenceID .
estatdata:ReferenceID
    estat:hasURI estatdata:SEArticleID.
estatdata:SEArticleID
    estat:hasURL "http/.../"^^xsd:anyURI .
```

Figure 6: Format for internal reference.
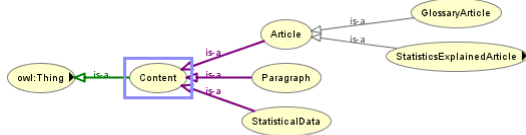


Figure 7: Analysis of the Content class.



Figure 8: Part of the datasets taxonomy.

acterizes the articles, the only difference is that the themes of OECD are related with themes from Eurostat, this is achieved with the property *relatedTheme*.

The Classification class is connected with other classes through property *hasClassification* which has the following sub-properties *hasCategory* (this property also has sub-properties the *hasCategoryOfGlossaryArticle* and *hasCategoryOfStatisticExplainedArticle*), *hasTheme* (this property also has a sub-property *hasOECDTheme*), *hasTopic*, and *hasType*.

## 3.3 Populating Eurostat

The population of the KG was performed after the construction of the ontology schema, and integrates all the information from the SQL knowledge database, that is associated with the classes presented in Subsection 3.2, into the KG. The integration was performed with 4 Python scripts we have developed that map the information from SQL knowledge database to the KG (i.e., it translates the information existing in the relational tables into RDF).

The first script, inserts knowledge about the Statistic Explained and Glossary Articles. Next, the second script inserts knowledge about the glossaries of Eurostat and OECD, and the references of the articles. The third script inserts knowledge about OECD. Finally, the fourth script inserts knowledge about the terms, topics and the classifications that enriched the knowledge graph via semantic analysis and natural language processing techniques, such as Topic Modeling using LDA. The result is a KG with 307419 explicit and 827395 implicit triples that were inferred based on RDFS Plus semantics.

## 3.4 Alignment with External Ontologies

The Eurostat ontology is aligned with various popular external vocabularies to provide interoperability with
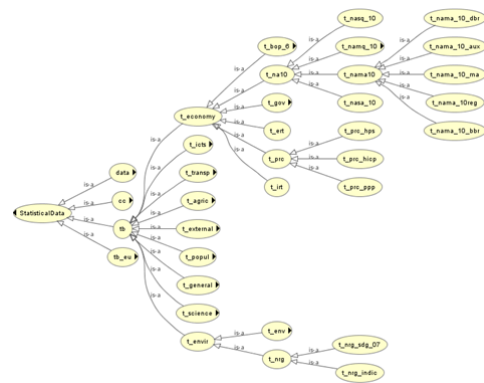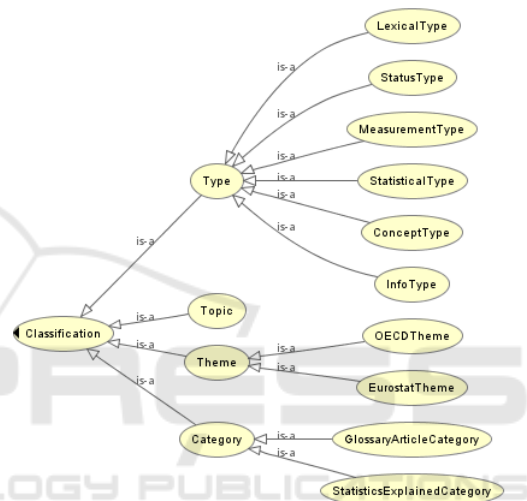


Figure 9: Analysis of the Classification class.

other knowledge graphs and ontologies. The alignment is in the form of mappings axiomatized using the predicates rdfs:subClassOf, rdfs:subPropertyOf, owl:equivalentClass, owl:equivalentProperty, and skos:closeMatch. The external vocabularies are: (i) Data Catalog Vocabulary (DCAT)[12], (ii) Simple Knowledge Organization System (SKOS) (Isaac and Summers, 2009), (iii) DCMI Metadata Terms[13], (iv) Schema.org (Guha et al., 2016), (v) FOAF Vocabulary (FOAF) (Amith et al., 2020), and (vi) RDF Data Cube Vocabulary (Cyganiak et al., 2014).

## 3.5 Use Case

The Eurostat KG was constructed as part of the *NLP4StatRef project* of Eurostat. The NLP4StatRef project, was established to capture four major Use Cases (A - D). Due to space restrictions we concen-

---

[12]dcat:http://www.w3.org/ns/dcat

[13]https://www.ndl.go.jp/jp/dlib/standards/translation/dcmi-terms.htm

trate only to Use Case A in detail. Use Case A tries users experience to interact, interconnect, and re-use the content and data existing in the Eurostat website, through a variety of services including faceted search, guided query builders, as well as services for data exploration and visual data browsing.

**Query Builder:** This is a semantic extension search tool in which the extraction of the information is done from the titles, contents and annotations of the Eurostat articles. The GUI features auto-completion, concepts suggestion, resource type selection (eg., SE article, SE Glossary article, related articles) and also a term understood in an expression. The aim of the tool is to propose to the user related concepts to the one they are enquiring. Based on this knowledge, we suggest polysemous meanings, more generic or more specific concepts, and/or related concepts following a selection of relations. In this way, this tool goes further than what the search bar of the Eurostat website does currently. Below we show how the search results enhance with the use of the Eurostat KG.

- We load the SE Glossary articles data from the knowledge database, in particular, IDs, titles definitions and URLs.

- Similarly, we load the required information from the SE articles, i.e. IDs, titles and URLs, paragraph titles and paragraph contents.

- The result is a list of named texts with as many elements as the texts processed (4292 in the latest run). The elements are themselves lists containing the stemmed terms in each text, the original terms and the URLs where the terms were found.

- The next step is to create three dictionaries, corresponding to 2-, 3- and 4-grams. The keys in n-grams are (n-1) tuples of stemmed tokens. For each key in a dictionary, the value is another (nested) dictionary with the original terms, their counts and the relevant URLs. In the end, the counts are used to calculate probabilities. The example in Figure 10 shows the three values in the key 'collect', 'european','statist' in the 4-gram dictionary, corresponding to the continuations: 'accidents', 'system' and income', with probabilities 0.714, 0.143 and 0.143, respectively.



Figure 10: Query Builder search results without the use of Eurostat KG.

Our tool receives input directly from the KG with SPARQL queries. It accepts content from both SE ar-

ticles and SE Glossary articles and returns very rich 'suggestions', based on n-grams and special dictionaries. The use of SPARQL queries over the KG instead of SQL queries over the relational content DB offered significant (over 75%) performance improvement over the query execution time (see Figure 11).



Figure 11: Query Builder search results using Eurostat KG.

**Faceted Search:** This tool provides options to the user to search along conceptual dimensions / slices of the results. The current faceted search tool of the Eurostat's web site allows only to search for three facets, namely Themes, Collections and publication year. Our tool extends the above facets along hierarchies, such as the alignment of the themes / subthemes taxonomy with the categories tagging. Also, it provides more facets, such as continent - countries, content types, and organisations (e.g. OECD). In the SQL version of the tool only related Statistics Explained articles were shown, whereas using the KG and SPARQL queries we were able to group the resources which are related to the results (SE articles) in SE articles, Glossary articles, Publications, Legislation and Others.

The user interface elements are shown in the following Figure 12. These elements are linked by interactions so that the options available reflect the current selections. The output is a list of the filtered articles, optionally together with the related articles and links. The example below shows only the first SE article found, together with its related data assets, in groups.



Figure 12: Faceted Search results using Eurostat KG.

## 4 DISCUSSION

KGs allow the representation of information from websites into a machine understandable format and, consequently, the exploitation of semantics, i.e., relations that connect entities in the KG with methods that are closer to human thinking. The exploitation of semantics can give great aid to question-answering systems, or to data-driven models trained on them. Also, the translation of data into KGs automatically allows the interconnection and re-usability of the translated data. This is a great advantage, especially in the domain related KGs, as many systems can access and use the data which are in the cloud of KGs.

In this paper we presented the Eurostat KG that contains most of the information from the Eurostat and OECD websites, such as information about articles and datasets, interconnections between them and external sources, and information for various classifications for the articles and datasets, among others. We described how we developed the schema of the ontology, how we captured the data that we used to infer the schema, and how we populated the KG with the aforementioned information.

The creation of the Eurostat KG offers the following: (i) Increases the discoverability and accessibility of data available for analytical purposes, (ii) Strengthens Eurostat position within the Commission as a provider of statistical data and services for its internal users, and (iii) Improves the methods for extracting information from unstructured data sources – especially data available on the web.

As for future work, we plan to create a visualization mechanism that will project pieces of the KG. Moreover, we will link the KG with more external knowledge, for instance from DBpedia and/or ConceptNet, and furthermore with other knowledge graphs e.g. from the EU Open Data portal[14] or to extend the current KG with more knowledge coming from related statistical agencies in Europe or worldwide, the Euro SDMX Registry[15] or the RAMON Metadata Server[16].

## ACKNOWLEDGEMENT

## REFERENCES

Amith, M., Fujimoto, K., Mauldin, R., and Tao, C. (2020). Friend of a friend with benefits ontology (foaf+): extending a social network ontology for public health. *BMC Medical Informatics and Decision Making*, 20(10):1–14.

Arp, R., Smith, B., and Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Mit Press.

Bandrowski, A. et al. (2016). The ontology for biomedical investigations. *PloS one*, 11(4):e0154556.

Capadisli, S., Auer, S., and Ngonga Ngomo, A.-C. (2015). Linked sdmx data. *Semantic Web*, 6(2):105–112.

Cyganiak, R., Reynolds, D., and Tennison, J. (2014). The rdf data cube vocabulary.

Franck, C., Manuel, S., Mauro, B., Francesco, A., and Giuseppina, R. (2018). Modernstats standards supporting the implementation and sharing of statistical services.

Guha, R. V., Brickley, D., and Macbeth, S. (2016). Schema. org: evolution of structured data on the web. *Communications of the ACM*, 59(2):44–51.

Iqbal, R. et al. (2013). An analysis of ontology engineering methodologies: A literature review. *Research journal of applied sciences, engineering and technology*, 6(16):2993–3000.

Isaac, A. and Summers, E. (2009). Skos simple knowledge organization system. *Primer, World Wide Web Consortium (W3C)*, 7.

Kendall, E. F. and McGuinness, D. L. (2019). Ontology engineering. *Synthesis Lectures on The Semantic Web: Theory and Technology*, 9(1):1–102.

Otte, J. N., Beverley, J., and Ruttenberg, A. (2022). Bfo: Basic formal ontology. *Applied ontology*, (Preprint):1–27.

Sembiring, J. and Uluwiyah, A. (2015). Data and metadata exchange design with sdmx format using web service for interoperability statistical data. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 14(2):343–352.

Smith, B. et al. (2007). The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255.

Zheng, J., Harris, M. R., Masci, A. M., Lin, Y., Hero, A., Smith, B., and He, Y. (2014). Obcs: The ontology of biological and clinical statistics. In *Proc. Fifth International Conf. on Biomedical Ontology*, volume 1327.

[14]https://data.europa.eu/en

[15]https://webgate.ec.europa.eu/sdmxregistry/

[16]https://ec.europa.eu/eurostat/ramon