Semantic Annotation and Publication of Linked Open Data

Serena Sorrentino[1], Sonia Bergamaschi[1],
Elisa Fusari[2], and Domenico Beneventano[1]

[1] DIEF - University of Modena and Reggio Emilia
via Vignolese 905, 41100 - Modena, Italy
{sonia.bergamaschi,serena.sorrentino}@unimore.it
[2] Graduate Student at DIEF - University of Modena and Reggio Emilia,
via vignolese 905,41100 Modena, Italy
57915@studenti.unimore.it

Abstract. Nowadays, there has been an increment of open data government initiatives promoting the idea that particular data produced by public administrations (such as public spending, health care, education etc.) should be freely published. However, the great majority of these resources is published in an unstructured format (such as spreadsheets or CSV) and is typically accessed only by closed communities. Starting from these considerations, we propose a semi-automatic experimental methodology for facilitating resource providers in publishing public data into the Linked Open Data (LOD) cloud, and for helping consumers (companies and citizens) in efficiently accessing and querying them. We present a preliminary method for publishing, linking and semantically enriching open data by performing automatic semantic annotation of schema elements. The methodology has been applied on a set of data provided by the Research Project on Youth Precariousness, of the Modena municipality, Italy.

1 Introduction

Nowadays, the availability of freely accessible information on the Web is constantly growing. In particular, recently, there has been an increment of open data government initiatives (e.g., data.gov for US and data.gov.uk for UK, dati.gov.it for Italy etc.) promoting the idea that certain data produced by public administrations (such as public spending, health care, education etc.) should be freely published in order to allow companies and citizens to browse, analyze and reuse them [12].

As a result, numerous open data sources are available on public organization's web sites. However, the great majority of these resources is published in an unstructured format (such as spreadsheets or CSV) and is typically accessed only by closed communities. Indeed, even if freely available on the Web, there are no connections among them and their structural and semantic heterogeneity makes it difficult to perform automatic or semi-automatic cross-data analysis, thus preventing to obtain high value information.

B. Murgante et al. (Eds.): ICCSA 2013, Part V, LNCS 7975, pp. 462–474, 2013.

The Linked Open Data (LOD) paradigm represents the key solution to improve and enrich the use of open data and to help consumers (citizens and companies) to access their integrated information. In the Semantic Web research area, the term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web [4]. LOD extends the linked data paradigm by publishing data which are freely available to everyone and for any purpose. LOD data sets are represented and published in the RDF[1] standard format and any resource (e.g. things, persons, etc.) has a dereferenceable URI (Uniform Resource Identifier)(i.e., a string of characters used to identify a name or a resource) as global identifier. In this way, open data sets can be exposed on the Web and data consumers could use the current Web infrastructure to obtain relevant information about any resource. The RDF data sets, then, can be queried by using the standard SPARQL query language.

Nevertheless, providing a standard way to represent and query public data is not enough: a full and easy access to open data sources requires the opportunity to integrate multiple LOD data sets belonging to the same knowledge domain. However, while the LOD cloud is rich of instance links (e.g., *owl:sameAs* relationships), schema level mappings (e.g., *rdfs:subClassOf* relationships) which are fundamental for performing dataset integration are almost absent. In this context, *semantic annotation* of schemas, i.e. the explicit association of one or more *meanings* to a schema element with respect to a reference lexical thesaurus is a key tool. Its effectiveness has been proved in the task of discovering schema and ontology mappings, i.e. semantic correspondences at the schema-level [3].

Starting from these considerations, we present a preliminary and experimental semi-automatic methodology to perform: RDF-ization (i.e., RDF translation) of open data sets; semantically annotation of their schema elements; publication on the Web; linking in the unified LOD cloud. In particular, during the process we make use of different already developed methods and tools.Our methodology represents a first step towards an automatic LOD integration system allowing users to publish any kind of public data, dynamically integrating two or more LOD data sets by exploiting semantic information and querying them without any pre-configured statistic analysis.

The methodology has been applied on a real case: the data we used were provided by the Research Project on Youth Precariousness, of the Modena municipality, Italy, which is carrying out an investigation about the precarious situation of young people living in the Modena district[2].

The rest of the paper is organized as follows. In Section 2, we give a step-wise description of the methodology by illustrating its main features, functionalities and the tools employed. Section 3 describes and analyzes related work. Finally, in Section 4, we give our concluding remarks and describe future work.

[1] `http://www.w3.org/RDF/`

[2] The project is carried out by the Councillor for Youth Policies Fabio Poggi, in collaboration with Prof. Claudio Baraldi and Dr. Federico Farini of the Department of Language and Culture, University of Modena, Italy.

2 Annotation and Publication of Linked Open Data

Our goal was to provide a standard methodology to facilitate both source providers and consumers in publishing, semantically enriching and querying LOD data sets. To this aim, we studied a general methodology consisting of four main steps:

1. *RDF-ization* for modeling the data in a structured format and convert them into RDF;
2. *Semantic Enrichment* for understanding the semantics of source schema elements;
3. *Web Publishing* for making data accessible through a SPARQL query endpoint;
4. *Linking and mapping* for discovering instance level-links and semantic mappings between the public data sets and other LOD resources.

Figure 1 shows the process of annotation and publication of the Youth Precariousness data set, and the interaction with the tools and resources exploited during the methodology. In the following, we describe each of these steps in details. We start by briefly describing the data set we used.

2.1 The Youth Precariousness Data Set

Our methodology has been applied on a set of data collected within the Modena (Italy) district project *Youth Precariousness*. This project aims to analyze the actual situation of job and emotional insecurity that young people are living in Modena. The data about the Youth Precariousness were collected by means of a paper questionnaire and were stored in Excel spreadsheets. The questionnaire were filled up by young people with age between 20 and 35 years. The questionnaires were anonymous and composed by 29 questions including the personal data of the interviewed such as age, birth place, the actual employment situation (i.e., employed, unemployed or student), school and university career and questions about parents job, family and friends in order to assess their social environment. Moreover, it included psychological questions about what the interviewed expects from the future: uncertainties, difficulties, expectations etc (see [10] for more details).

At present, the data collection phase is still in progress. However, for our purpose, it was enough to apply our methodology on the first 315 collected questionnaires. The questionnaire data were stored within an Excel spreadsheet, which simply map each question with the corresponding answer (e.g., question Q3 answer a2).

2.2 RDF-ization

In [4], Sir Tim Berners Lee introduces a "'5-star rating system'" for open data. This system can be summed up as follow:

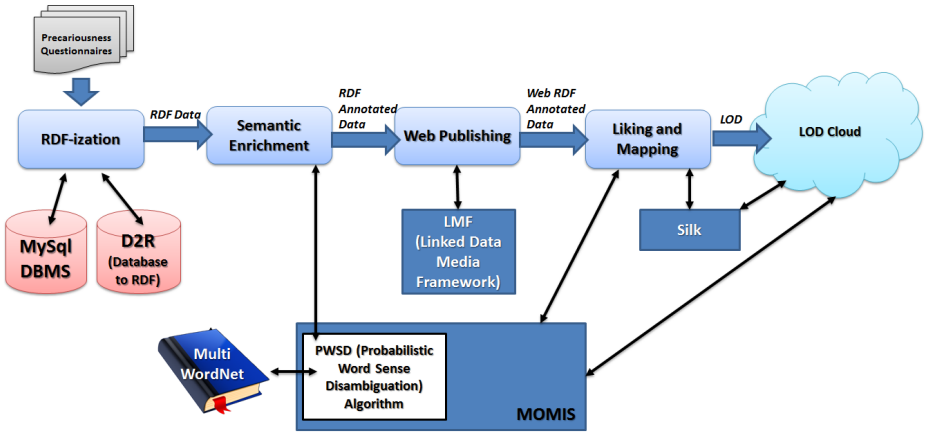

Fig. 1. The Annotation and publication process

1. one star: the resource is available on the web (whatever format);
2. two stars: the resource is available as structured data (e.g. excel instead of the scan of a table);
3. three stars: the resource is available as a non-proprietary structured data format (e.g. csv instead of excel);
4. four stars: the resource is available in the RDF format and use URLs to identify things;
5. five stars: the resource is linked to other open data.

Our methodology aims to make available the Youth Precariousness data set as a five stars open data. Thus, first of all we need to convert the data set into the RDF standard knowledge representation language. To represent data into RDF, we need first to convert them in a relational database and then to exploit one of the several freely available open source automatic tools for relational-RDF translation. The relational database has been realized in a semi-automatic way: starting from the Excel spreadsheet, we analyzed the data in order to design the corresponding Entity/Relationship diagram [19]. This process is fundamental and it has been performed manually with the support of Entity/Relationship editors. In particular, during this step, we identified the main concepts and the relationships among them.

Then, by using the open source MySQL Workbench tool[3], we automatically generated and populated a relational database storing the public data. From the collected questionnaires we created a database composed by 18 table for a total of approximately 5400 records.

Finally, we employed the D2R (Database To RDF)[4] open source software to convert the database into RDF. D2R is an HTTP server that allows us to convert

[3] `http://www.mysql.com/products/workbench/`
[4] `http://d2rq.org/`

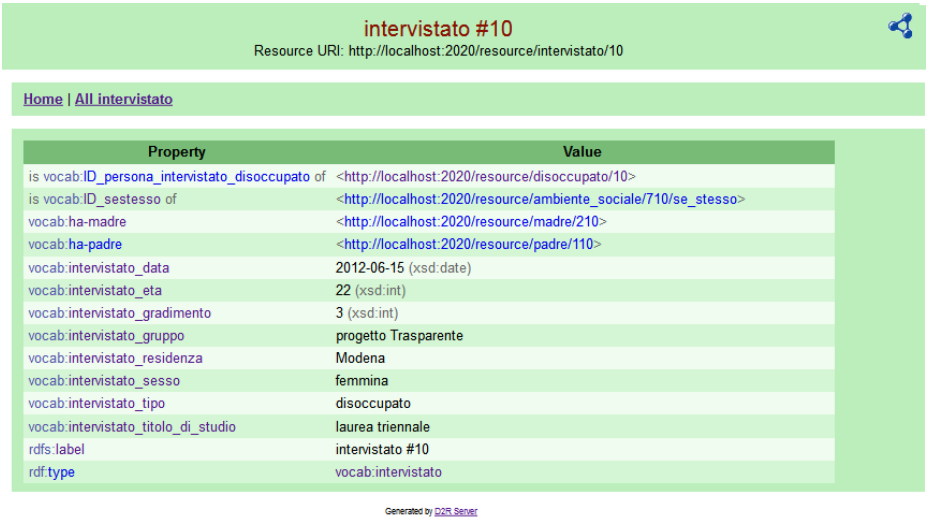

Fig. 2. Example of RDF data representation by using D2R

relational data in RDF triples through its specific internal language called D2RQ: a mapping file is created which maps each database table in a corresponding RDF class and each column in a RDF property. During this phase, the only information to be provided is the resource URI, which is mandatory for creating globally unique resource identifiers.

Figure 2 shows the RDF representation of an excerpt of the Youth Precariousness data set schema where each class and attribute has been converted into RDF by using the D2R vocabolary "*vocab*".

2.3 Semantic Enrichment

In order to efficiently use LOD data sets, consumers need to deeply understand the semantics of source schemas. Moreover, the hidden meanings associated to schema elements can be exploited for discovering semantic mappings and thus performing integration of different LOD data sets. Indeed, semantics facilitate and speed up the recognition of correspondences between data, contextualizing and enriching the information available.

To add semantics at the schema level, we use semantic annotation. *Semantic annotation* is the process of explicit alignment of one or more meanings to schema element labels (classes and attributes names). Manual semantic annotation is a time consuming and not scalable task. However, automatic semantic annotation is difficult due to the problem of term ambiguity. Thus, to perform automatic or semi-automatic annotation, a method for *Word Sense Disambiguation* (WSD), i.e. for identifying the sense of a term in a context [16], has to be devised.

We decided to utilize the PWSD (Probabilistic Word Sense Disambiguation) algorithm [18] developed in the MOMIS data integration system [2], which annotates schema elements with one or more meanings.

PWSD performs annotation with respect to the lexical reference database WordNet [15]. The strength the WordNet is that it provides a set of possible meanings, called synsets, for each term (nouns, adjectives, verbs and adverbs) and includes a wide network of semantic relationships among these meanings (e.g., hypernymy/hyponymy relationship defines between two concepts where one is more general/specific of the other) which can be used to infer semantic mappings among schema elements.

PWSD is composed by five different algorithms: the Structural Disambiguation algorithm which exploits terms that are related by a structural relationship (e.g., is-a relationships); the WordNet Domains Disambiguation algorithm which tries to disambiguate terms by exploiting domains information supplied by WordNet Domains [11]; the Gloss Similarity and the Iterative Gloss Similarity algorithms based on string similarity techniques; WordNet first sense heuristic rule selecting the first WordNet meaning (that is the more used in English) for a term. Before performing annotation, all the schema labels are preprocessed by using the normalization techniques described in [20], in order to expand abbreviations, remove stop words, and identify compound terms.

However, we cannot directly apply PWSD to the schema labels of our target data set which are in Italian, while the original version of WordNet includes English terms only. Therefore, we extended and modified the original implementation of PWSD in order to deal with Italian terms.

First of all, we needed to select an Italian thesaurus: we decided to use MultiWordNet [5] which is a multilingual lexical database containing an Italian version of WordNet strictly aligned with the English WordNet (e.g., the Italian synset {corte, tribunale} is aligned with the English synset {court, tribunal, judicature}). Moreover, it includes the access to other versions of WordNet in several languages as well as all the relationships that exist between the various translations of the same word.

Then, as the PWSD algorithm has been designed for annotating English terms, we need to verify its applicability for annotating Italian terms in the context of Liked Open Data. We experimentally verified that the Structural, Gloss Similarity, Iterative Gloss Similarity and the First Sense algorithms can be directly applied to Italian terms as they exploit term features (like the structural relationships and the glosses) that do not depend on the language. As regards the WordNet Domains algorithm, it can be easily adapted to Italian terms: indeed, in MultiWordNet, the domain information has been automatically transferred from English to Italian, resulting in an Italian version of the resource WordNet Domains: for instance, as the English synset {court, tribunal, judicature} was associated with the domain LAW, also the corresponding Italian synset {corte, tribunale}, results automatically associated with the LAW domain.

[5] `http://multiwordnet.fbk.eu/english/home.php`

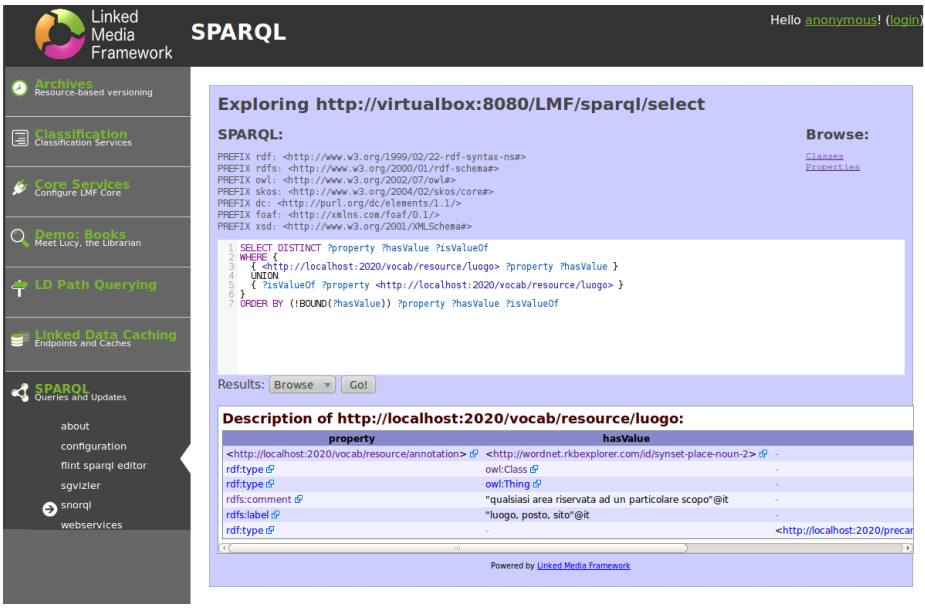

Fig. 3. Screen shot of the SPARQL service of LMF showing the annotation tag for the class label "luogo"

PWSD is a probabilistic algorithm, i.e. it associates to each annotation a probability value indicating the reliability of the annotation itself. The probability is used to filter annotations having reliability under a given threshold. An evaluation of PWSD, is out of our scope and can be found in [20]. However, to give an idea of its performance in the case of Italian schemas, we evaluated precision and recall of the annotation process on our resources: by using a probability threshold of 0.30, it obtained 0.71 in precision and 0.57 in recall. The recall value is not so high due to the application of a threshold greater than the one usually used in [20] (i.e., 0.15). However, this helps to reduce the risk of wrong annotations that might propagate errors in all the derived mappings in the LOD cloud.

To represent the semantic annotations in RDF, we used the class properties *owl:AnnotationProperty*, *rdfs:label* and *rdfs:comment*. In particular, we added as label the element name and its synonyms terms and as comment the Italian gloss (i.e., the definition of the meaning) taken from MultiWordNet.

As the English WordNet is available in the LOD cloud in RDF/OWL format[6] by using MultiWordNet, we can further enrich schema elements by linking Italian annotations to the corresponding English WordNet URI.

The RDF/OWL WordNet schema has three main classes: Synset, WordSense and Word. Each instance of Synset, WordSense and Word has its own URI. There

[6] `http://www.w3.org/TR/wordnet-rdf/`

is a pattern for the URIs so that it is easy to determine from the URI the class to which the instance belongs. The navigable URI provides some information on the meaning of the entity it represents. For example, the following URI

`www.w3.org/2006/03/wn/wn20/instances/synset-bank-noun-2`

is an instance of the class Synset representing the second meaning of the noun "bank". We used a custom class of *owl:AnnotationProperty* in order to link an *rdfs:Class* in our public schema with an instance of the class Synset in WordNet. The custom property *<owl:AnnotationProperty rdf:about="&vocab;annotation"/>* was called *vocab:annotation*, where the namespace "vocab" refers to the vocabulary that is automatically created by D2R during the RDF database conversion. By using this property, we can insert the link to the RDF WordNet synset in a navigable way. For instance, the semantic enrichment of the Italian schema element "luogo" is represented in the following way:

```
<owl:Class rdf:about="&vocab;luogo">
    <rdfs:label xml:lang="it">
        luogo, posto, sito
    </rdfs:label>
    <vocab:annotation rdf:resource="&wordnet;synset-place-noun-2/>
    <rdfs:comment xml:lang="it">
        qualsiasi area riservata ad un particolare scopo
    </rdfs:comment>
</owl:Class>

<!--http://wordnet.rkbexplorer.com/id/synset-place-noun-2 -->
<owl:Thing rdf:about="&wordnet;synset-place-noun-2">
    <rdf:type rdf:resource="&rdfs;Resource"/>
</owl:Thing>
```

where the annotation tags, actually, link the meaning from the WordNet thesaurus to the schema elements (in the form of URIs).

2.4 Web Publishing

The set of thesaurus-based annotation tags previously obtained represent the semantics of the schema. The following step is to make the data public on the Web. To this aim, we need an RDF repository exposing an HTTP de-referenceable SPARQL endpoint, so that the published data set can be referred and linked to other resources from the LOD cloud. To this aim, we evaluated two open source tools providing both the functionalities of RDF data storing and RDF querying: Fuseki[7] and LMF (LinkedData Media Framework)[8].

Fuseki is a SPARQL Server implemented by Apache-Jena, which allows us to query and analyze RDF data through a query engine called ARQ. Its main advantage is that it provides several query functionalities, such as grouping operators or counting functions, which can be used to perform statistic queries (e.g.,

[7] `http://jena.apache.org/documentation/serving_data/index.html`
[8] `http://code.google.com/p/lmf/`

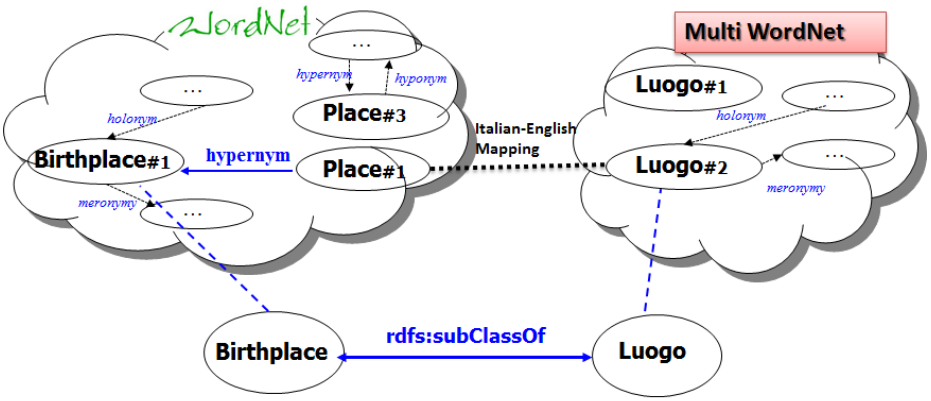

Fig. 4. An example Mapping discovery from semantic annotations: for the label "Luogo"

"how many interviewed are students? how many are employed and how many are unemployed?").

LMF is an application server employing the query service Snorql[9]. With respect to Fuseki, it permits to compose only simple SPARQL queries (e.g., it does not support grouping operators). However, it provides navigation functionalities that allow us to explore the WordNet URIs and see all the information related to its resource (see Figure 3).

For our purpose, we decided to employ LMF for its navigation functionalities. However, a composition of both the tools represent an interesting future work.

2.5 Linking and Mapping

After having published the data on the Web, we need to link the data set to other LOD resources. Creating links and mappings between the published resources is a key part of the Linked Open Data (LOD) paradigm [7]. We can identify two kinds of connections:

- *Instance-Level links* which are established between LOD data set instances (e.g., *owl:sameAs* established between two instances representing the same real world objects);
- *Schema-level mappings* which are established between schema concepts (e.g., *rdfs:subClassOf* used to state that all the instances of one class are instances of another);

In the LOD cloud, instance-links represent the great majority of connections. In our case, thanks to the annotation tags, our schema is automatically linked to the

[9] `http://data.semanticweb.org/snorql/`

RDF version of WordNet. To link the data set to other LOD resources, we can use Silk [14], a popular semi-automatic framework providing several patterns and property based techniques for helping data set providers in discovering instance links. We used Silk to discover links between our public data and DBPedia [5] which essentially, makes the content of Wikipedia available in RDF.

As observed before, schema-level mappings are almost absent in the LOD cloud even if, as previously described, they represents a fundamental means to integrate different LOD resources. Silk does not provide the semantic techniques needed to discovery semantic mappings among LOD schemas. We can discover semantic mappings among different LOD schemas exploiting the techniques developed for ontology and schema matching systems [9].

As will be pointed out in Section 4, this problem represents a core challenge and a future work of our research area. However, in this section, we want to present our approach to discover mappings starting from the previous obtained semantic annotations.

We can automatically discover RDF relationships by exploiting the wide semantic network provided by WordNet. In particular, let c_1 and c_2 be classes of two different schemas and $m(c_1)$ and $m(c_2)$ their meanings in WordNet, we consider the following possible RDF relationships:

- c_2 *rdfs:subClassOf* c_1, defined if $m(c_1)$ is a *hypernym* of $m(c_2)$ in WordNet;
- c_1 *rdfs:subClassOf* c_2, defined if $m(c_1)$ is a *hyponym* of $m(c_2)$ in WordNet;
- c_1 *owl:equivalentClass* c_2, defined if $m(c_1)$ is a *synonym* of $m(c_2)$ in WordNet.

Figure 4 shows an example of mapping discovery starting from semantic annotations: let us suppose that we want to discover the mappings between our Youth Precariousness data set and another English LOD schema. First of all we annotate the elements of both schemas with respect to MultiWordNet (in case of Italian terms) and WordNet (in case of English terms). By using the direct correspondence between Italian and English WordNet synsets, we can discover that there exists a hyponym relationship between the meaning associated to "luogo" (i.e., "place" in English) and the annotation of "Birthplace". Thus, we can automatically infer that there exists also an rdfs:SubClassOf relationship between these two RDF classes.

To perform semantic driven mapping discovery, we will use the open-source MOMIS data integration system which has been designed and tested by our research group. An open-source version of MOMIS is actually delivered and maintained by the academic Spin-Off DataRiver[10]. However, MOMIS will need to be reviewed and adapted in order to deal with RDF and LOD resources.

3 Related Work

As we have previously seen, the creation of semantic mappings at schema level has a fundamental role in the integration and alignment of LOD resources from

[10] `http://www.datariver.it/`

different domains. Several research groups have developed tools, frameworks and platforms for the integration of Linked Open datasets at the semantic level.

BLOOMS (Bootstrapping-based Linked Open Data Ontology Matching System)[13] is a system for the alignment of LOD ontologies at schema level. It utilizes a bootstrapping approach based on the Wikipedia category hierarchy. Essentially, BLOOMS for each matching candidate ontology class C identifies all its corresponding Wikipedia articles and construct a forest (i.e., a set of trees, one for each article) T_C, by selecting the Wikipedia super categories of each selected Wikipedia article. Then they compare each couple of forests (e.g., T_C and T_B for the classes C and B) in order to evaluate whether or not two classes should be aligned. The main drawback of this approach is that, it considers all the possible meanings (i.e., Wikipedia articles) for each ontology class, thus increasing the complexity of the method and the risk to discover wrong mappings. On the contrary, in our approach we address the problem of term ambiguity by performing automatic WSD. Moreover, they only present a system for LOD mapping discovery while we propose a complete methodology for translating, publishing and mapping LOD datasets.

Another system for publishing LOD resources is AGROPub [17], which facilitates integration of LOD agro-environmental resources. AGROPub comprises services and tools that enable resource providers to semantically annotate their resources by relevant concepts from selected agro-environmental domain ontologies, to generate and publish RDF descriptions of the resources to LOD and to link the published resources to related resources from LOD. Moreover, it provides services and tools that enable consumers of the agro-environmental resources to search and annotate published resources by adding their own annotations. However, the semantic annotations have to be added manually by the users (resource provides or consumers) by using the GUI, and no automatic or semi-automatic annotation method is provided.

Stratosphere[11] is an open-source cluster/cloud computing framework for Big Data analytics. In [12] this system has been extended for the integration of large data sets belonging to the Linked Open Data. In particular, it has been applied to integrate open governmental data with other popular LOD resources, such as DBpedia[12] and Freebase[13]. It addresses the problem of semantic and structural heterogeneity among different LOD resources by developing data cleansing operators for the Stratosphere framework. The integration methodology is mainly based on the analysis of the instances of the data sets and on the use of entity (e.g., persons, cities etc.) extraction and record linkage (i.e., identify the real world entities across the different data sources) techniques. However, in our case this method could not be applied as the great majority of data are numeric or do not correspond to entities.

Finally, WebSmatch [6] is a flexible environment for Web data integration with a service oriented architecture. It has been applied on the real scenario of Data

[11] `https://www.stratosphere.eu/`
[12] `http://it.dbpedia.org/`
[13] `http://www.freebase.com/`

Publica, a French company, providing added value over public data sets they crawl, such as visualization of data source or data integration. WebSmatch has been employed all over the process of metadata extraction, matching and visualizing data sources. For the matching phase, it exploits YAM++ (Yet Another Matcher) [8], a tool for pattern matching and alignment of ontologies, which combines different matching techniques mainly based on the string similarity, dictionary and thesauri like WordNet and instance based techniques. However, also in this case it does not make us of WSD techniques.

4 Conclusion and Future Work

In this paper, we presented an experimental and preliminary methodology to publish and link public open data to the LOD cloud. Moreover, we propose an automatic and multilingual method to semantically enrich LOD data sets by performing semantic annotation of schema elements with respect to the MultiWordNet lexical thesaurus. The process has been applied to public data coming from the Research Project on Youth Precariousness of the district of Modena, Italy. However, it might be easily adapted for any other public data set.

As previously described, during the process we employed different open source software. The main drawback in using different tools is the need of creating custom interfaces in order to allow the automatic communication among them. Future work will be devoted to implement and integrated system providing all the functionalities supplied by the different tools used during the process in order to allow data providers and consumers to interact with them by using an integrated GUI.

Furthermore, we will investigate the application of traditional data integration and schema matching systems in the context of Linked Open Data: we will extend the MOMIS data integration system by adapting its schema matching method in dealing with RDF data; moreover, by using the MOMIS provenance module [1], we will add to the public data further RDF metadata describing the provenance of data (i.e., where data came from and how they were derived and modified over time) in order to provide consumers with valuable information that can be exploited during the LOD navigation.

Acknowledgments. Our sincere thanks to the Councillor for Youth Policies Fabio Poggi[14] and to Dr. Sergio Ansaloni, Head of Stradanove[15], Studies and Documentation Centre on Youth, for providing us the public data. This work is partially supported by the BIOGEST-SITEIA laboratory (`www.biogestsiteia.unimore.it`), funded by Emilia-Romagna (Italy) regional government.

[14] `http://www.comune.modena.it/politichegiovanili/info/assessorato`
[15] `http://www.stradanove.net/`

References

1. Beneventano, D.: Provenance based conflict handling strategies. In: Yu, H., Yu, G., Hsu, W., Moon, Y.-S., Unland, R., Yoo, J. (eds.) DASFAA Workshops 2012. LNCS, vol. 7240, pp. 286–297. Springer, Heidelberg (2012)
2. Bergamaschi, S., Castano, S., Vincini, M.: Semantic integration of semistructured and structured data sources. SIGMOD Record 28(1), 54–59 (1999)
3. Bergamaschi, S., Po, L., Sala, A., Sorrentino, S.: Data source annotation in data integration systems. In: Fifth International Workshop on Databases, Information Systems and Peer-to-Peer Computing, DBISP2P (2007)
4. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)
5. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia - a crystallization point for the web of data. J. Web Sem. 7(3), 154–165 (2009)
6. Coletta, R., Castanier, E., Valduriez, P., Frisch, C., Ngo, D., Bellahsene, Z.: Public data integration with websmatch. CoRR, abs/1205.2555 (2012)
7. Cruz, I.F., Palmonari, M., Caimi, F., Stroe, C.: Towards "on the go" matching of linked open data ontologies. In: LDH, pp. 37–42 (2011)
8. Duchateau, F., Coletta, R., Bellahsene, Z., Miller, R.J. (not) yet another matcher. In: Cheung, D.W.-L., Song, I.-Y., Chu, W.W., Hu, X., Lin, J.J. (eds.) CIKM, pp. 1537–1540. ACM (2009)
9. Euzenat, J., Shvaiko, P.: Ontology matching. Springer, Heidelberg, DE (2007)
10. Fusari, E.: Linked open data: pubblicazione, arricchimento semantico e linking di dataset pubblici attraverso il sistema momis. Master Degree Thesis (2012), `http://www.dbgroup.unimore.it/tesi/FusariElisa-tesi2012.pdf`
11. Gliozzo, A.M., Strapparava, C., Dagan, I.: Unsupervised and supervised exploitation of semantic domains in lexical disambiguation. Computer Speech & Language 18(3), 275–299 (2004)
12. Heise, A., Naumann, F.: Integrating open government data with stratosphere for more transparency. J. Web Sem. 14, 45–56 (2012)
13. Jain, P., Hitzler, P., Sheth, A.P., Verma, K., Yeh, P.Z.: Ontology alignment for linked open data. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 402–417. Springer, Heidelberg (2010)
14. Jentzsch, A., Isele, R., Bizer, C.: Silk - generating rdf links while publishing or consuming linked data. In: ISWC Posters&Demos (2010)
15. Miller, A.: Wordnet: A lexical database for english. Communications of the ACM 38(11), 39–41 (1995)
16. Navigli, R.: Word sense disambiguation: A survey. ACM Comput. Surv. 41(2) (2009)
17. Nešić, S., Rizzoli, A.E., Athanasiadis, I.N.: Publishing and linking semantically annotated agro-environmental resources to LOD with aGROPub. In: García-Barriocanal, E., Cebeci, Z., Okur, M.C., Öztürk, A. (eds.) MTSR 2011. CCIS, vol. 240, pp. 478–488. Springer, Heidelberg (2011)
18. Po, L., Sorrentino, S.: Automatic generation of probabilistic relationships for improving schema matching. Inf. Syst. 36(2), 192–208 (2011)
19. shan Chen, P.P.: The entity-relationship model: Toward a unified view of data. ACM Transactions on Database Systems 1, 9–36 (1976)
20. Sorrentino, S., Bergamaschi, S., Gawinecki, M.: Norms: An automatic tool to perform schema label normalization. In: ICDE, pp. 1344–1347 (2011)