Biswanath Dutta, (2017) "Examining the interrelatedness between ontologies and Linked Data", Library Hi Tech, Vol. 35 Issue: 2, pp.312-331, https://doi.org/10.1108/LHT-10-2016-0107

EXAMINING THE INTERRELATEDNESS BETWEEN ONTOLOGIES AND LINKED DATA

Biswanath Dutta
DRTC, Indian Statistical Institute
Bangalore – 560059, India
Email: dutta2005@gmail.com

Abstract

Purpose- Ontology and Linked Data are the two prominent Web technologies that have emerged in the recent past. Both of them are at the center of Semantic Web and its applications. Researchers and developers from both academia and business are actively working in these areas. The increasing interest in these technologies promoted the growth of linked datasets and ontologies on the Web. In the current work, we investigate the possible relationships between them. Our effort is to investigate the possible roles that ontologies may play in further empowering the Linked Data. In a similar fashion, we also study the possible roles that Linked Data may play to empower ontologies.
Design/methodology/approach- The work is mainly carried out by exploring the ontology and Linked Data based real-world systems, and by reviewing the existing literature.
Findings- The current work reveals, in general, that both the technologies are interdependent and have lots to offer to each other for their faster growth and meaningful development. Specifically, anything that we can do with Linked Data, we can do more by adding an ontology to it.
Practical implications- We envision that the current work, in one hand, will help in boosting the successful implementation and the delivery of semantic applications; on the other hand, it will also become a food for the future researchers in further investigating the relationships between the ontologies and Linked Data.
Originality/value- So far, as per our knowledge, there are very little works that have attempted in exploring the relationships between the ontologies and Linked Data. In this work we illustrate the real-world systems that are based on ontology and Linked Data; discuss the issues and challenges and finally illustrate their interdependency discussing some of the ongoing researches.

Keywords Ontology, Linked Data, Relationships between Ontology and Linked Data, Applications, Challenges, Data integration, Schema alignment.

1. Introduction

Ontology, a formal explicit specification of a shared conceptualization (Studer *et al.*, 1998), is at the center of Semantic Web (SW) (Berners-Lee *et al.*, 2001) and applications. It is a vocabulary where the terms are expressed formally (using knowledge representation formalism, such as, Description Logics) and defined explicitly (in terms of their properties and constraints), which make them machine processable and interpretable. Ontologies are useful for various purposes, for instance, annotating the documents, semantic information retrieval, reasoning and inferencing and so forth (discussed further in Section 3.2). An immense amount of research works are undergoing in the area of ontology, for instance, ontology development and ontology design approaches, ontology evaluation and ontology alignment and mapping (Adhikari *et al.*, 2015). Varieties of ontologies are available on the web ranging from general purpose ontologies (aka top-level ontologies, e.g., Cyc (OpenCyc for the Semantic Web, 2012; Suggested Upper Merged Ontology, 2016; Masolo, *et al.*, 2003)) to domain ontologies (e.g. spatial ontology, Gene Ontology (2016), food ontology) and application ontologies (e.g. restaurant ontology, recipe ontology).

Linked Data (LD) refers to a structured data published on the web following a set of principles designed to promote the interlinking between the things (aka resources) and consequently between the various data sets on the web. Here data refer to anything (W3C, 2015), such as date, time, title, numbers, chemical properties, images, video clippings, etc., that one can conceive of. A large number of researchers and practitioners both from academia and business are actively working in this area. Some of the notable Linked Data sets are DBPedia (2016) Linked Data set, Freebase (2009), Geonames (2017), MusicBrainz (2000), etc. The goal of the LD initiative is to create a huge data infrastructure, on top of which various applications can be built, for instance, MashUp applications and question answering systems (further discussed in Section 4.3). The expectation is that the LD technology will enable the applications, say, question answering systems to process the complex queries, such as, *give me the books on topic T written by an Indian author A who worked with an Italian professor A' in University U during the time period t*.

From the above discussion, precisely we can say that while the objective of an ontology is to assist the software program in semantic operations, the objective of LD is to assist in developing a global data infrastructure. Based on these technologies, several real-world applications are developed as discussed in Sections 3 and 4. In the recent time, we see some works questioning the relationships between the ontologies and LD. For example, in Studer *et al.* (2011), the authors have raised a question "did Linked Data kill ontologies?" To study the relationships between ontologies, annotation and LD, Janowicz and Hitzler (2013) have raised several questions, such as "are ontologies an additional layer on top of data models; are they data models themselves; are Linked Data entities instances of ontological classes or just annotated using ontologies which exist in their own realm; what difference does this make; what is the role of semantics & reasoning for querying and information retrieval? […]." In the current work, we aim to examine the possible relatedness between an ontology and LD. Our effort is to investigate the possible roles that an ontology may play in further empowering the LD. In a similar fashion, we investigate the possible roles that LD may play to empower and strengthen the ontology and its development. To achieve this, we study the strengths and weaknesses of both ontology and LD and explore whether they can be benefitted from each other. The main contributions of the current work are as follows: illustrating a set of ontology- and LD-based real-world systems; discussing the issues and challenges that both ontology and LD face today; and more importantly, presenting their complementary features reporting some of the ongoing-related research works. We envision that the current work, in the one hand, will help in accelerating the development of ontology and LD and their applications; on the other hand, it will also become a food for the future researchers in further exploring the relationships between these two prominent SW technologies.

The rest of the work is organized as follows: Section 2 discusses the related works; Section 3 elaborates the current state of the ontology. It illustrates some of the ontology-based real-world systems and applications and also discusses some of the challenges that an ontology, especially at the development phase, faces today. Similar to Section 3, Section 4 investigates the current state of LD. The section first briefly discusses LD, its usefulness and some of its real-world applications. And then elaborates some of the challenges that LD faces today. Section 5 explores the interrelatedness between an ontology and LD. It illustrates and explains how an ontology and LD can be benefited from each other. Finally, Section 6 concludes the paper and provides some observations. It also discusses some research questions that need to be further investigated.

2. Related Work

So far, there are very few works have attempted to explore the relationships between an ontology and LD. In this section we briefly discuss them.

Jain, Hitzler, Yeh, Verma and Sheth (2010) have explored the various limitations of Linked Data cloud (LDC), for instance, lack of conceptual description of data sets, lack of expressivity, the absence of schema-level links and so forth. They have advocated for the use of an upper-level ontology to alleviate these limitations. According to them in the absence of an ontology, LD, in its current form, is merely weakly linked "triple collection" and will only be of very limited benefit for the artificial intelligence or SW communities. Poggi *et al.* (2008) have discussed how linking of data to ontologies would help for designing effective systems for ontology-based data access. Halpin and Presutti (2009) have illustrated how an ontology can make LD more self-describable and can also support inferencing (for instance, to verify the class membership of various resources). They have argued that an ontology should serve as a fundamental contribution of modeling LD. Studer *et al.* (2011) in a presentation have briefly discussed the relationships between an ontology and LD by analyzing their characteristics and possible uses. One of the most prominent works, which is closely related to the current work, is by Janowicz and Hitzler (2013). In this work, they have tried to explore the relationships between LD, semantic annotations and ontologies. The relationships are explored from two perspectives: is LD created by extending and instantiating ontologies; or by relating them through semantic annotations, i.e., instead of directly instantiating data from the classes, annotating data using classes from an ontology? They have finally concluded by saying that "[…] for the Linked Data, it is time to give up on the idea of context-free ontologies as models of the physical world and instead define a multitude of purpose and data-driven micro-ontologies." In Dutta (2014), the author has attempted to investigate some of these questions.

The basic differences between the previous works and the current work are: in the current work, we investigate and discuss the relationships between ontologies and LD from two perspectives: the strengths and weaknesses of them; and the way they can be mutually benefitted from each other in overcoming some of the issues and challenges they are facing today.

3. Ontology: where are we?

3.1 What is an Ontology?

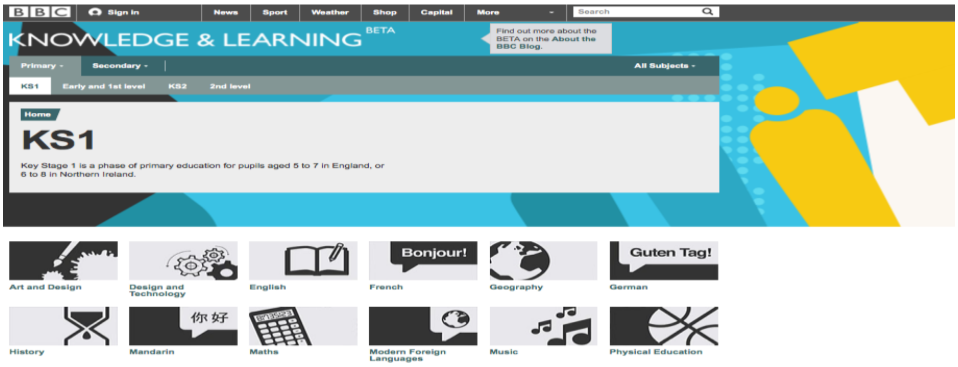
In Information Science and Computer Science, ontology is considered as an engineering artifact. It is referred as a formal naming and definition of the types, properties and relationships of the entities that really or fundamentally exist in a particular domain of discourse. The most prominent definition of ontology was provided by Gruber (1993). According to him, ontology is an "explicit specification of a conceptualization." Studer *et al.* (1998) extended Gruber's definition stating that "an ontology is a formal, explicit specification of a shared conceptualization." So, in simple words, we can say that an ontology is a formally represented knowledge of a domain of discourse (aka universe of discourse) based on a shared conceptualization. Here, conceptualization refers to an abstraction, a simplified view of the domain of discourse motivated by some purposes. The formal and explicit specification of the conceptualization of the domain of discourse makes the constituents of an ontology machine interpretable (Dutta, 2008).

Uniform Resource Identifier (URI) and Resource Description Framework (RDF) are the two key web technologies for the construction of an ontology in OWL, a W3C recommended Web Ontology Language. URI (Berners-Lee, 2006) is to name things (the resources) globally and also to uniquely identify them. The URIs may also be the Internationalized Resource Identifiers, i.e., web addresses that use the extended set of natural-language scripts supported by Unicode. RDF is a data model. It is usually seen as a directed graph with labeled nodes and arcs. RDF enables to describe the resources in the form of a subject (i.e. the resource), predicate (i.e. the property of the resource) and object (i.e. the value of the property).

3.2 Ontology applications

Ontology is in the core of the semantic-based applications. It has immense significance in semantic applications. For instance, as a controlled vocabulary, which can be used by both humans and computers to communicate and access information, and for knowledge sharing within and between the domains enabling the semantic interoperability (Ouksel and Sheth, 1999). An ontology can be used for representing and storing data, reasoning and inferencing knowledge. It can also be used to organize, navigate and manage web content and can be used as a tool for NLP tasks, such as, for sense disambiguation (Sanderson, 1994). In the following, we illustrate some of the real-world applications that are based on the ontologies. Note that we do not attempt to provide a comprehensive review of the state of the art of ontology-based applications.

Content organization – an ontology can be used for content organization and navigation. One such real-world example is BBC's Education system (Figure 1). The system uses a curriculum ontology (available at: www.bbc.co.uk/ontologies/curriculum) to organize the learning contents. The ontology provides the data model and vocabularies for describing the national curricula within the UK. Besides the education system, BBC also uses the ontologies for organizing contents, such as music, general news, etc.



Find out more about the BETA on the About the BBC Blog.

Key Stage 1 is a phase of primary education for pupils aged 5 to 7 in England, or 6 to 8 in Northern Ireland.

Figure 1. BBC's educational site (http://www.bbc.co.uk/education)

Entity Markup – an ontology can be used to markup entities (e.g. person, organization, location, music) exist in the web pages. The marked-up web pages are easy to interpret by software programs. Google search engine uses the marked-up information for displaying the content in search results in a useful way, for instance, by showing the rich snippets (a small piece or brief extract). Figure 2 presents a snippet of recipe retrieved from Google. In this context, we can mention schema.org (2011), a vocabulary supported by the major search engines such as Google, Yahoo! and Bing. It is designed to create structured data markup. The content creators can use this vocabulary to markup a wide range of entities such as a person, organization, location, event, book, recipe, music, video and so forth.

Thai **Mango Salad recipe** - Canadian Living

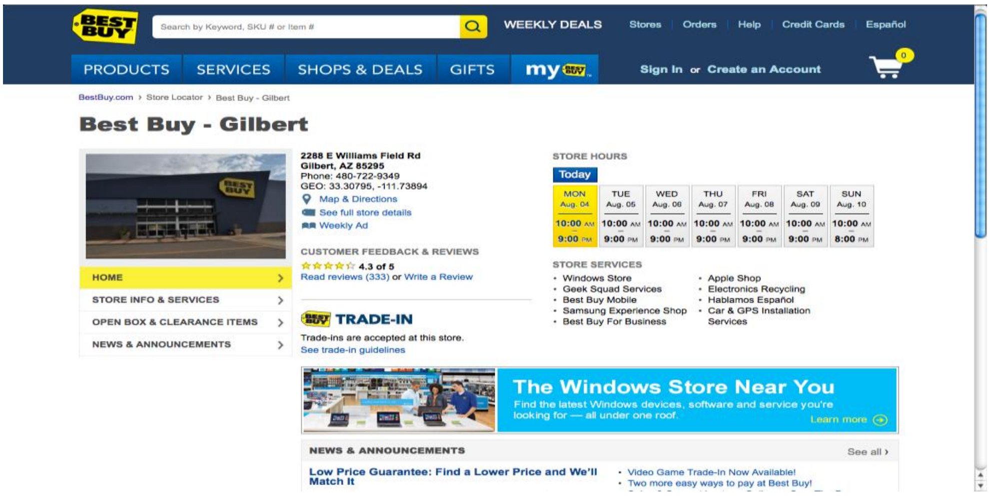
www.canadianliving.com/food/thai_**mango**_**salad**.php

★★★★☆ Rating: 4 - 38 votes - 251 cal

The traditional Thai **salad** is sweet and spicy, filled with finely chopped pork, dried ... Add **mangoes**, red pepper, carrots, mixed greens, **green** onions and mint; ...

Figure 2. Google snippet for a salad recipe

Content publication – the use of an ontology in content publication increases the visibility of the sites and the content. The use also increases the ranking of the sites significantly in the search results. Many online commercial websites are using the ontologies to structure and publish their content, for instance, Bestbuy (2009) (Figure 3). It uses GoodRelations (2011), a standard vocabulary to describe product, price, store and company data.



Best Buy - Gilbert

2288 E Williams Field Rd
Gilbert, AZ 85295
Phone: 480-722-9349
GEO: 33.30795, -111.73894

CUSTOMER FEEDBACK & REVIEWS

★★★★☆ 4.3 of 5

TRADE-IN

Trade-ins are accepted at this store.

STORE HOURS

Today

MON Aug. 04	TUE Aug. 05	WED Aug. 06	THU Aug. 07	FRI Aug. 08	SAT Aug. 09	SUN Aug. 10
10:00 AM – 9:00 PM	10:00 AM – 9:00 PM	10:00 AM – 9:00 PM	10:00 AM – 9:00 PM	10:00 AM – 9:00 PM	10:00 AM – 9:00 PM	10:00 AM – 8:00 PM

STORE SERVICES

- Windows Store
- Geek Squad Services
- Best Buy Mobile
- Samsung Experience Shop
- Best Buy For Business
- Apple Shop
- Electronics Recycling
- Hablamos Español
- Car & GPS Installation Services

The Windows Store Near You

Find the latest Windows devices, software and service you're looking for — all under one roof.

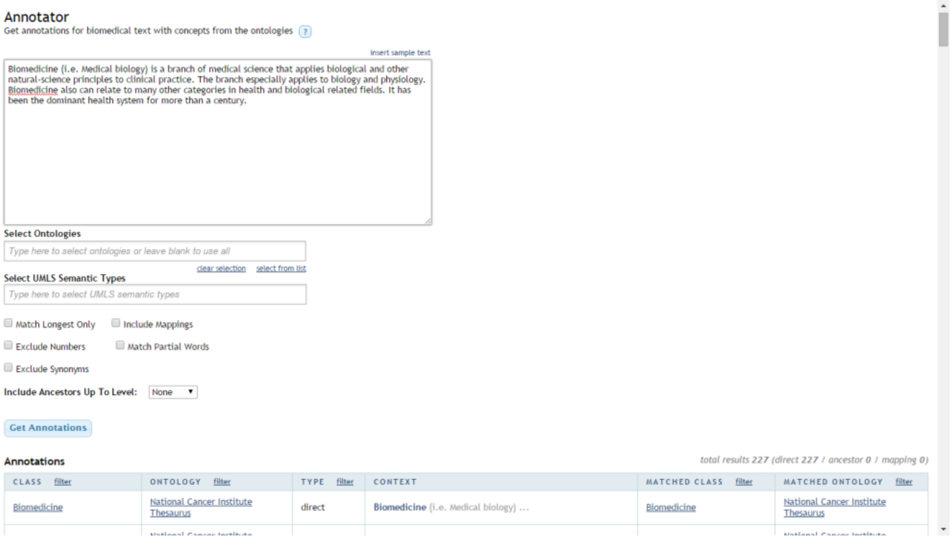
NEWS & ANNOUNCEMENTS

Low Price Guarantee: Find a Lower Price and We'll Match It

- Video Game Trade-In Now Available!
- Two more easy ways to pay at Best Buy!

Figure 3. Best Buy using GoodRelations

Content annotation – an ontology can be used in annotating content. One such example is BioPortal Annotator (BioPortal, 2005). The annotator annotates biomedical text with the concepts from the ontologies. To annotate the content, we need to enter text in the text box and press the submit button. The system then matches words in the text to the terms in the ontologies by doing an exact string comparison (i.e. a "direct" match) between the text and the ontology term names and synonyms. Figure 4 presents a screenshot of the annotator system. It shows the annotation result for a piece of text that we copied from Wikipedia. The result shows the details of the classes from the text and their corresponding matching classes from the ontologies.



Biomedicine (i.e. Medical biology) is a branch of medical science that applies biological and other natural-science principles to clinical practice. The branch especially applies to biology and physiology. Biomedicine also can relate to many other categories in health and biological related fields. It has been the dominant health system for more than a century.

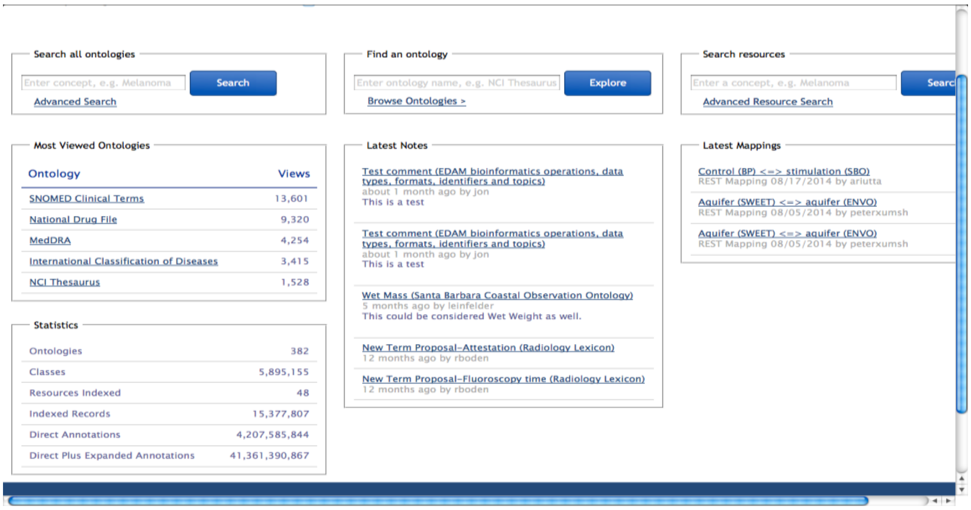
☐ Match Longest Only ☐ Include Mappings
☐ Exclude Numbers ☐ Match Partial Words
☐ Exclude Synonyms
Include Ancestors Up To Level: None

total results 227 (direct 227 / ancestor 0 / mapping 0)

CLASS filter	ONTOLOGY filter	TYPE filter	CONTEXT	MATCHED CLASS filter	MATCHED ONTOLOGY filter
Biomedicine	National Cancer Institute Thesaurus	direct	Biomedicine (i.e. Medical biology) ...	Biomedicine	National Cancer Institute Thesaurus

Figure 4. BioPortal annotator

Content navigation – an ontology can be used for content navigation. For example, BioPortal, the largest repository for biological ontologies, provides enhanced content navigation and search facilities as shown in Figure 5.



Ontology	Views
SNOMED Clinical Terms	13,601
National Drug File	9,320
MedDRA	4,254
International Classification of Diseases	3,415
NCI Thesaurus	1,528

Ontologies	382
Classes	5,895,155
Resources Indexed	48
Indexed Records	15,377,807
Direct Annotations	4,207,585,844
Direct Plus Expanded Annotations	41,361,390,867

Test comment (EDAM bioinformatics operations, data types, formats, identifiers and topics)
about 1 month ago by jon
This is a test

Test comment (EDAM bioinformatics operations, data types, formats, identifiers and topics)
about 1 month ago by jon
This is a test

Wet Mass (Santa Barbara Coastal Observation Ontology)
5 months ago by leinfelder
This could be considered Wet Weight as well.

New Term Proposal–Attestation (Radiology Lexicon)
12 months ago by rboden

New Term Proposal–Fluoroscopy time (Radiology Lexicon)
12 months ago by rboden

Control (BP) <=> stimulation (SBO)
REST Mapping 08/17/2014 by ariutta

Aquifer (SWEET) <=> aquifer (ENVO)
REST Mapping 08/05/2014 by peterxumsh

Aquifer (SWEET) <=> aquifer (ENVO)
REST Mapping 08/05/2014 by peterxumsh

Figure 5. Content navigation in BioPortal

Semantic integration of data - the ontologies can be used for semantic integration of data frommultiple sources. The researchers and practitioners in the field of database and information integration have produced a large set of research works to facilitate the interoperability between different systems (Noy, 2004). Similarly, in the area of ontology, many researchers are working toward the semantic integration of data. One such recent work is the Knowledge Base of Biomedicine (KaBOB) (Livingston *et al.*, 2015). It is a knowledge base of semantically integrated biomedical data from 18 databases (e.g. Xenarios *et al.*, 2000; Law *et al.*, 2014; Becker *et al.*, 2004; UniProt Gene Ontology Annotation, 2016). KaBOB uses 14 Open Biomedical Ontologies, for instance, Basic Formal Ontology (2002), BRENDA Tissue/Enzyme Source (Gremse *et al.*, 2011) and Chemical Entities of Biological Interest (2017). The multiple ontologies are used primarily to provide a common representation model of data. The use of biomedical ontologies allows making queries in terms of biomedical concepts, for instance, genes and gene products, proteins, interactions and processes.

Besides the above applications, there are many others where the ontologies are used. For instance, for topic exploration (Doms and Schroeder, 2005; Weijian Xuan *et al.*, 2009), query enhancement (McGuinness, 1998; Arenas *et al.*, 2014), software development (Uschold, 2008; Bertoa *et al.*, 2006; Alonso, 2006) and query expansion (Fu *et al.*, 2005; Jain *et al.*, 2012).

3.3 Ontology: some truths and challenges

In the above section, we have illustrated the several real-world uses of ontologies in designing the semantic applications. Nevertheless, besides the above success stories of ontological uses, an ontology, especially its developmental phase, faces various challenges. Some of such challenges are discussed below:

- *Expensive* – an ontology construction is an expensive affair. A usable ontology demands lots of human resources, infrastructural support and time (Obrst *et al.*, 2014; Dutta *et al.*, 2015).
- *Growth is slow* – an ontology development is a mental process. A quality ontology (i.e. a usable ontology) attracts an immense amount of human labor, especially in terms of knowledge modeling and formalization. The following works on ontology development (Fernandez *et al.*, 1997; Vrandecic *et al.*, 2005; Gruninger and Fox, 1995; Uschold *et al.*, 1995; KBSI, 1994; Giunchiglia *et al.*, 2010) exemplify the complexity of the process.
- *Domain terminology* – one of the most significant steps of ontology construction is to identify and select the domain terminologies. The usual approach is to scan through the domain literature and also to consult with the domain experts to gather the domain terms. This is a cumbersome job and also not always easy to identify the required terms. It needs a thorough understanding of the domain and also the purpose, task and goal of the targeted ontology.
- *Degrees of formality* – generally ontologies are classified, from the perspective of degrees of formality, into two (Van Heijst *et al.*, 1997): lightweight and heavyweight. A lightweight ontology is "an ontology having thesaurus-like structure and is based on a minimal level of logic constructors" (Giunchiglia *et al.*, 2009). It supports simple taxonomic inferences. While a heavyweight ontology is an ontology which includes all the features of a lightweight ontology plus the additional restrictions and axioms that allow performing richer inferences with the underlying data (Corcho *et al.*, 2015). A heavyweight ontology supports complex semantic operations, but its design demands a deep understanding of the domain and sophisticated tools to reason and infer knowledge. On the other hand, a lightweight ontology supports limited semantic operations, but easy to design and implement.

The question is between the heavyweight and lightweight ontologies, which one should be preferred in LD web? In this context, we quote Hendler. According to him "A little semantics goes a long way" (Hendler, 2009). But still, it is not clear, does a lightweight ontology is sufficed to support the vision of the LD web? Can a lightweight ontology support in drawing a meaningful answer for a complex query? Or, do we need to find a middle path in between these two types, i.e., a middleweight ontology, which balances between a lightweight and heavyweight ontology?

- *Ontology reuse* - Ontology reuse is a real concern of the SW community (Obrst *et al.*, 2014; Dutta *et al.*, 2015). Since ontology is an expensive affair, the ideal situation would have been to "reuse" an existing ontology or a substantial part of it built for similar kinds of applications. However, it is hard to find a consensus among the ontology engineers in terms of knowledge modeling and representation. As a result, often we end up with creating the ontologies from scratch every time we build applications.

4. Linked Data: where are we?

4.1 What is Linked Data?

LD, in general, refers to the data published in accordance with principles designed to facilitate linkages among data sets, element sets and value vocabularies (Berners-Lee, 2006). It is about linking the web of data in a way so that both human being and machine can explore and make optimum use of available data on the web. According to Tim Berners-Lee, the vision of SW will come true by not just putting data on the web, but by making relationships between data. The relationships between data will facilitate both machine and human being, to explore and know more about a thing (a resource). The goal here is to evolve the web like a single global database to provide integrated access to data from a wide range of distributed and heterogeneous data sources.

LD uses the web technologies such as URI, HyperText Transfer Protocol URI (HTTP URI), RDF and SPARQL Query Language for RDF: W3C Recommendation (2008). URI for naming the things, HTTP URI, so that people can look up those names. Standards such as RDF and Simple Protocol and RDF Query Language (SPARQL) are to provide meaningful information when someone looks up for a URI. RDF together with its core technology URI facilitates the data linking and dereferencing (Berners-Lee, 2006).

4.2 The usefulness of Linked Data

LD can be better understood by exploring its significance from various aspects, for instance, data accessibility, federated search, data currency, contribution to science and research (Benefits of the Linked Data Approach, 2011) as discussed here:

- *Integrated access to data* – the fundamental strength of LD lies in its capacity of integrating the geographically sparse data and providing an integrated access to data. Through this, the navigation across resources becomes more sophisticated.
- *Data enrichment* – it refers to the enrichment of a knowledge base in terms of data volume and potentially the data quality. LD technology enables us to enrich the data in an easy way. The technology supports the data enrichment by linking the various data sets available on the web.

- *Data format* – LD method has brought a fundamental change in the way we share, retrieve and mix our data. All data published as LD on the web have a common and consistent data format, i.e., RDF. So, the data mixing has become relatively easy.
- *Decentralization* – LD technology provides decentralized platforms where data development, creation, curation and structuring are not centrally located.
- *Data sharing* – data sharing has become easy, which was never before. LD technologies and linking and publishing tools have made data sharing easy. For any organization, data sharing and publishing have become cost effective.
- *Data reuse* – the cost effectiveness of data sharing and publishing also has influenced and has increased the chances of data reusability.
- *Data maintenance and data currency* – LD technologies have also made data update easy. Data update at the source gets affected on run time to all the knowledge bases where the data are shared.

4.3 Linked Data applications

We provide here the glimpses of the real-world applications of LD. The applications are classified into two broad categories: general web applications and domain-specific applications.

4.3.1 General Applications

The applications those are of general kinds. For instance, LD browsers and search engines and review and rating systems.

Linked Data Browser

LD browsers are similar to the traditional browsers. The basic difference between these two is: in the traditional browser we navigate between HTML pages following the hyperlink links, whereas in LD browser we navigate between data and data sources following the links expressed as RDF triples. For instance, we start with a search on "Rabindranath Tagore" from a data set on "Poets in Bengal" maintained by Sahitya Academy and reach to a place "Kolkata" (where Tagore was born) and from Kolkata we reach to "Presidency College" that belongs to a data set on "Academic institutions" maintained by Government of West Bengal. So, LD enables us to start from a data set and traverse to another one following RDF's HTTP URI links rather than HTML links. The examples of LD browsers areMarbles (Becker and Bizer, 2009), Tabulator (Berners-Lee *et al.*, 2006), etc. (more can be found on: www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/SemWebClients)

Search engines

A search engine is a place where navigation starts. The LD browsers allow us to navigate information space, while search engines are often the place where navigation starts (Bizer *et al.*, 2009). Some of the notable search engines are Sig.ma (2011), FalconS (Cheng *et al.*, 2008), Swoogle (2007), Watson: exploring the Semantic Web (2010), etc. Figure 6 shows the results of a query in FalconS.

Objects **1 - 5** of **100** for your search **Tim Berners-Lee**

Tim Berners-Lee - Tag
- type: Tag
- preferred label: **Tim Berners-Lee**
- has broader: Grand Homme
- page: **tim_berners_lee**.html
- isDefinedBy: **tim_berners_lee**.rdf
- has related: Linking Open Data
- is tag of: http://www.w3.org/2005/Talks/0517-boit-tbl/
- has broader: Technical girls and guys
- has broader: SW guys (and girls)
- is broader of: Tabulator

http://www.semanlink.net/tag/**tim_berners_lee**

Tim Berners-Lee - Person
- type: Person
- label: **Tim Berners-Lee**
- name: **Tim Berners-Lee**
- is maker of: Policy Aware Content Reuse on the Web
- is Creator of: Policy Aware Content Reuse on the Web
- is _3 of: something that
- is author of: Policy Aware Content Reuse on the Web
- homepage: **Berners-Lee**
- is _4 of: editor_list
- is editor of: proceedings

http://data.semanticweb.org/person/**tim-berners-lee**

Figure 6. Search result in FalconS

4.3.2 Domain Specific Applications

Besides the above general applications of LD, there are many domain-specific applications and services exist. These applications are mostly built by mashing up data from various LD sources. Some of the significant applications are discussed as follows.

Revyu (http://revyu.com/)
Revyu is a live, publicly accessible generic reviewing and rating system. It allows reviewing and rating any named entity (Giunchiglia and Dutta, 2011), for instance, person, location, song, movie and event. The system is designed based on the LD principles and SW technologies, namely, RDF and SPARQL. One of the key design goals of Revyu system is to improve the user experiences by minimizing the burden on users and maximizing the reuse of external data sources by consuming the data available on the web. For instance, when we review a song, the system automatically retrieves additional information from DBPedia, where a match is found, about the song, say, the lyricists of the song. This reduces the job of a human being from re-entering the data that are already available on the web of data. On the other side, the system also makes sure that it also blossom the LD web by making links in RDF (Heath and Motta, 2008). So, we can say that Revyu system not only uses and exploits the existing LD resources but also contributes and adds data into the LD web. The data created in Revyu is open to the other systems to exploit further.

It is worth to mention here that Revyu does not use any ontology at the backend. According to the Revyu developers, the users need not classify the items they review, instead, they need to associate tags with the items. Some of the reasons behind the decision of not to use ontologies, as given in Heath and Motta (2007), are the lack of sufficiently comprehensive classification scheme for the review items; the users would be constrained to subscribe to a single classification scheme; and it is unfeasible to provide a usable interface to the non-specialists to classify items using arbitrary types discovered in ontologies.

RDF Book Mashup (http://wifo5-03.informatik.uni-mannheim.de/bizer/bookmashup/)
The RDF Book Mashup demonstrates how Web 2.0 data sources such as Amazon, Google and Yahoo can be integrated into the SW. Following the principles of LD, the

RDF Book Mashup makes information about books, their authors, reviews and online bookstores available on the SW. This information can be used by RDF browsers and crawlers, and other publishers of SW data and can set links to it.

GeoLinkedData.es (http://geo.linkeddata.es/)

GeoLinkedData is an open initiative of the Ontology Engineering Group. The goal is to publish and provide access to the Spanish geospatial data. The data set is prepared following the LD principles (Bizer *et al.*, 2009). The GeoLinkedData consisted of data collected and integrated from multiple data sources, namely, Numeric Cartographic Database, National Geographic Gazetteer, Conciso Gazetteer, National Atlas and EuroGlobalMap. Figure 7 presents GeoLinkedData visualization interface, which provides an integrated access to data.

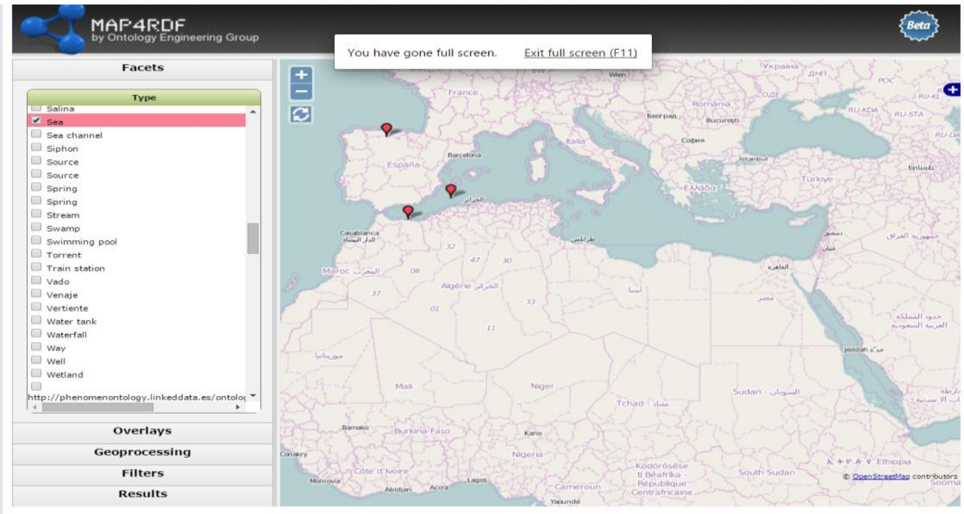


Figure 7. GeoLinked Data visualization interface

OpenAGRIS (www.agris.fao.org)

OpenAGRIS is a mashup web application. It links the International System for Agricultural Science and Technology (AGRIS) knowledge, a multilingual bibliographic database for agricultural science and technology, to the related web resources using the Linked Open Data methodology to provide access to information about a topic within the agriculture domain. At present OpenAGRIS (beta version) consisted of more than 60 million triples. AGROVOC, a multilingual thesaurus consisted of nearly 40,000 concepts in over 20 languages covering subject fields in agriculture and related areas, such as forestry and fisheries, food security, etc., is central to OpenAGRIS linking. AGROVOC is used for labeling AGRIS records and interlink to other thesauri (e.g. Eurovoc, NAL, DBPedia) for extracting more information about the concepts (Celli *et al.*, 2011).

From the above discussion, we can observe that LD is going to change the way present search systems work. In the LD web, searching information on a thing would be simpler and most of the time would be bounded to a single-page result. It is because all the data sources dealing with same/different aspects about a thing are linked. This will also essentially reduce the number of searches as we do not have to individually visit the multiple sites to find and gather information on a thing.

4.4 Linked Data: some truths and challenges

There is a viral growth of LD. Millions of triples are available on the web. For instance, DBpedia 2015-04 (Freudenberg *et al.*, 2015) release consists of 6.9 billion RDF triples, out of which 737 million were extracted from the English edition of Wikipedia, 3.76 billion were extracted from other language editions and about 2.4 billion are links to external data sets. Geonames (2016) consists of 162 million RDF triples. LODStats (2016) has reported 9,960 data sets adhering to the RDF available on the web. Nevertheless, irrespective of the spectaculars growth, LD also faces certain issues and challenges. Some of them are briefly discussed below:

- *Emphasis is on data linking and publication* – LD initiative has enabled easy publication, distribution and sharing of data including the linking with other data sets. The issue here is, the maximum emphasis is on publishing data in a structured format following the LD principles (Corcho *et al.*, 2015), but there is no or less attention on describing the data in terms of concept, property and relationships among the data sets. The publication of data in a structured format is not sufficient to realize the notion of LD. This will make Linked Data set as merely a collection of triples (Poveda-Villalon *et al.*, 2014).
- *Expressivity* – primarily the LDC, in its current form as stated above, is a collection of RDF triples and does not utilize the rich expressive power of ontology languages such as OWL and RDFS. As Jain, Hitzler, Yeh, Verma and Sheth (2010) have stated that the expressivity of LDC as a knowledge base is very shallow, and hence, it scarcely allows making any use of the underlying formal semantics through reasoning. This lack of expressivity also resists the reasoner from detecting the data inconsistency. For instance, there is an inconsistency related to the “population” size of “Bangalore” between DBPedia and Geonames. DBPedia shows 8.4 million, while Geonames shows 9.6 million. A reasoner could detect this inconsistency provided that the property of population is defined as functional. Because we know a functional property restricts to only one value for a property of a resource.
- *Ambiguity* – one of the main concerns of LD is its missing semantics. For instance, we consider a RDF triple <http://example.org/abu rdf:type http://example.org/bank>. We can make out from this triple, “abu” is a “bank,” but what is the term “bank” refers to? “Bank” may appear to be a financial institution, a river bank and so forth. The answer cannot be provided unambiguously unless the meaning of the concept “bank” is defined.
- *Data quality* - this is a very common and well-defined issue within the LD research. The data quality is usually judged from multiple dimensions (Zaveri *et al.*, 2012; Hogan *et al.*, 2012), for instance, data consistency (“means that a knowledge base is free of (logical/formal) contradictions with respect to particular knowledge representation and inference mechanisms”), data completeness (“refers to the degree to which all required information is present in a particular dataset”), data conciseness (“[…] refers to the case when the data does not contain redundant objects […]”) and so forth. There is an increasing amount of research undergoing in this area (Rula and Zaveri, 2014). Various quality assessment approaches are proposed in the state-of-the-art literature. Some of the approaches, especially the ontology-based approaches are discussed in Section 5.1.
- *Social trust* - LD is a community effort and this is the most positive and advantageous side of it. Because of the community participation, the mission of LD is appearing to be a success story. In fact, we have already started seeing various applications, as discussed above, based on LD. Nevertheless, the community-driven

approach also has its negative side as well. Some of the recent studies observed that LD also suffers from trustworthiness (Rowe and Butters, 2009), besides the data quality issues as discussed above. Sometimes data are manipulated and produced with the wrong intention (Bechhofer *et al.*, 2013). As an implication to these issues, the services designed based on LDC lack the social trust. LD with provenance may help to achieve the social trust.

- *Data reuse* - publishing data set on the LDC may not make the data set reusable by default. The data sets need to be described to make them identifiable, discoverable and selectable. We need to have metadata about the data itself (Berners-Lee, 2006).

5. Ontology and Linked Data: Made for Each Other

In this section, we explore the obvious and also the potential relatedness between an ontology and LD.

5.1 Ontology for Linked Data

In the following, we elaborate how an ontology can complement the Linked Data for developing a true Semantic Information Space and in totality the true Semantic Web.

Integration of semantics into data – publishing data with an ontology adds semantics into data. Here, the ontology can be seen as a layer on top of the data, which makes the data meaningful to both human being and software program. An ontology makes data amenable to interpretation and processing by software program. For instance, in the case of Figure 8, the data (below the dotted line) become more meaningful and easy to interpret in the presence of an ontology (above the dotted line consisting of classes and properties). In the presence of the ontology, we can say that both the resources Mauna Loa and Mount Vesuvius are volcanos (namespace references are omitted). In addition, we can also say from their class information that they are not the same types of volcanos. In the figure, the properties (written within the parenthesis) of the class Volcano are marked with a prefix a_. These properties are further get propagated into its subclasses and their instances. In the figure, the classes and instances are indicated with the solid and hollow circles, respectively.

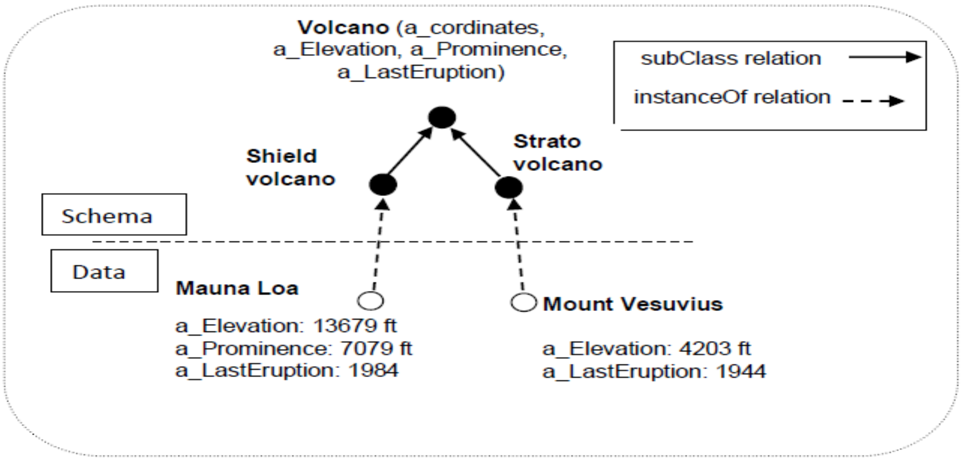


Figure 8. Semantic integration to data

Integration and alignment at the instance level and schema level - the use of an ontology in publishing LD helps in data integration and schema alignment. For instance, in the Figure 9, the resources Mount Vesuvius and Vesuvius belonging to two different data sets D1 and D2, respectively, are basically the same resource. The sameness is established based on their matching attribute values and is further confirmed by their class information, i.e., both of them are the type of Strato Volcano as indicated in the ontologies O1 and O2. Since Mount Vesuvius and Vesuvius are the same entities, we can link them using a semantic property, say, *owl:sameAs* (a property defined in OWL language (OWL Web Ontology Language Overview: W3C Recommendation, 2004). This linking enriches the data source D1 by adding an additional attribute *a_AgeOfRock* and its value to the resource Mount Vesuvius.

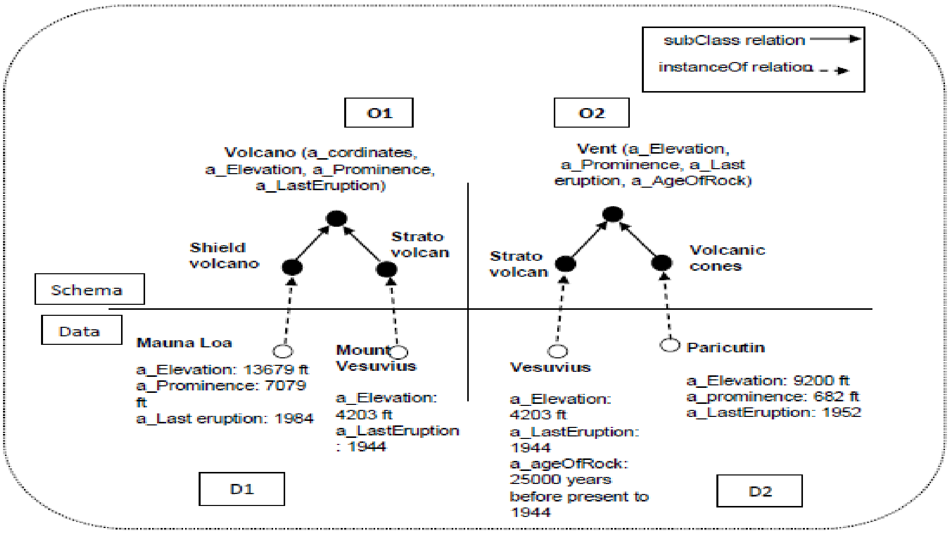


Figure 9. Integration and alignment at instance and schema level

Besides the above instance alignment, publishing LD with an ontology is also useful in schema alignment. For instance, in Figure 9, the root classes of both the ontologies O1 and O2 have two different names, namely, Volcano and Vent, respectively, but conceptually both refer to the same resource “a rupture on the crust of a planetary-mass object, such as Earth, that allows hot lava, volcanic ash, and gases to escape from a magma chamber below the surface.” Since both refer to the same resource, we can consider them as equivalent classes and hence can be aligned and linked through a semantic property *owl:equivalentClass* (a property defined in OWL language). The establishment of this linking increases the number of classes at the sub-class levels of both of the ontologies. Initially, in O1, we had two types of volcanos, namely, Strato volcano and Shield volcano, whereas, in O2, we had two types of volcanos, namely, Strato volcano and Volcanic cones. After the linking, in O1, one more volcano type, i.e., Volcanic cones will be added. Similarly, in O2, one more volcano type, i.e., Shield volcano will be added. The schema-level alignment also enhances the data sets by adding the corresponding data resources for the added classes.

The above example exemplifies the use of an ontology in data integration and schema alignment when the data sets are in the same language, i.e., English. However, as it can be easily understood from this example that even in the case of multilingual data sets, with the help of an ontology the data integration and schema alignment become easy.

Entity disambiguation and transparency in data linking - following the above discussions, we can also say that an ontology brings transparency in data linking. The data publication with the ontologies helps in disambiguating and linking the relevant resources across the data sets, for instance, Abu (a mountain) and Abu (a Person). Although the two resources have the same name, but in the presence of the ontologies, from their class and attribute definitions, we can easily disambiguate them.

Data modelling and publication – an ontology expresses the domain of interest at a high level of abstraction and exhibits the conceptual view of the domain. Hence, an ontology can be used to model and represent data of a domain of interest.

Ontology-based data access – many organizations face the problem of accessing the data irrespective of the availability of the powerful and efficient mechanism for accessing the data (Poggi *et al.*, 2008). Our argument is since an ontology exhibits the conceptual view of the domain of interest, it can be used as a formally defined sophisticated tool to provide access to data.

Inferencing new knowledge - ontology brings semantics into data and makes data amenable to infer implicit knowledge. For instance, from the following two axioms, as depicted in Figure 8, “Mauna Loa is an instance of Shield Volcano” and “Shield Volcano is a kind of Volcano,” the inference engine can conclude that “Mauna Loa is an instance of Volcano.” The reasoning and inferencing over data are possible only in presence of an ontology. Hence, we can say that an ontology can play a crucial role in making the LD useful to its full potential.

Ontology for LD data quality assessment – in LD web, the quality of data is a real concern to each of us (Zaveri *et al.*, 2013). The effectiveness of the applications of LD, for such as planning, development, decision and policy makings is highly dependent on the quality of data. Our argument is that an ontology can be applied in assessing and improving the LD quality. Toward this, we can mention some of the undergoing works. Zaveri *et al.* (2012) have elaborated how an ontology can be applied to determine the misuse of *owl:ObjectProperty* and *owl:DatatypeProperty*, or can be applied to detect the inconsistent values. They have discussed that the various data quality requirements can be encoded as data quality rules, and thus can be used to verify and determine the potential data quality issues. An ontology with axioms can help in detecting the data inconsistency in LD. Furber and Hepp (2011) have proposed a standard vocabulary to facilitate the knowledge representation required for monitoring the data quality, quality assessment, data cleansing and quality driven data retrieval in SW architectures.

5.2 Linked Data for Ontology

Similar to an ontology, LD and LDC also have a lot to offer to the ontologies. In this section, we explore the possible roles that LD can play in empowering the ontologies.

Data-driven ontology construction – at present, the majority of the ontology development process is based on a top-down approach, which starts from an abstraction of a domain and proceeds to a concrete level. From logic perspectives, the top-down approach corresponds to deductive reasoning (Grangel-Gonzalez *et al.*, 2015). In this approach, we start with known facts taken as premises and look for conclusions. With the availability of the LDC, there has been an evolving change observed in the ontology development approaches. Instead of following the top-down approach, there has been an increasing trend in applying the bottom-up approach. Unlike the top-down approach, the bottom-up approach (in Library and Information Science, this approach is popularly known as a last-link-upward approach (Ranganathan, 1997, p. 97)) starts with a set of grounded concepts and then analyzes and models the domain and builds the classification. The major advantage of the bottom-up approach is it is more efficient as the domain modeling is based on raw and evidential data and not mere theoretical conceptualization of a domain. As Janowicz and Hitzler (2013) correctly mentioned that "ontologies should be engineered based on the real data they are supposed to react and their axiomatization should be driven by the inference needs of typical queries." Grangel-Gonzalez *et al.* (2015) have advocated for the bottom-up approach which they named "vocabulary development by convention." They have elaborated a set of best practices and conventions for ontology construction using LD. Poveda-Villalon *et al.* (2014) have proposed a lightweight method for the data-driven ontology development rather than the competency question-based (Noy and McGuinness, 2001) development method.

Linked Data cloud as an enriched source of domain knowledge – as discussed above (Section 3.3), the primary challenges in ontology development are the discovery and acquisition of domain knowledge and terminologies. To extract a good amount of domain terminologies, an ontology engineer consults multiple resources. This is quite a cumbersome job. Here, the LD can be seen as an opportunity for the ontologists (an ontology developer who develops ontologies). The LDC can be used as an enriched source of domain knowledge. This is further evidenced in some of the recent works as follows. Thorsen and Pattuelli (2015) in their work on "Ontologies in the Time of Linked Data" have demonstrated the use of LDC in constructing a Linked Jazz ontology in the domain of the performing arts. They have concluded that "the ontology building process can be relatively simple in terms of data acquisition and modeling when compared to traditional practices." Garcia-Silva *et al.* (2014) have illustrated an ontology development approach in the financial domain extracting the terms from the Linked Open Data cloud.

Application driven ontology construction - LD can guide us to identify and select the domains for ontology construction. As per our experience, working on ontology development, selection of a right domain is always a complex task. Because ontology is a time-consuming process and an expensive affair, we cannot show lavishness in constructing an ontology for which we will not have an immediate use. LD can be used as a tool to foresee the domain requirements of the community.

Linked Data cloud for ontology alignment - ontology alignment (i.e. "a set of correspondences between two or more ontologies" (Euzenat and Shvaiko, 2007)) is a complex task. LDC consists of a vast amount of data. This data cloud can be used as a resource to disambiguate the word senses and align the ontologies. There are few works that have already been taken place on this, for instance, Parundekar *et al.* (2010) have proposed an automatic approach for finding alignments between the ontologies of geospatial data sources, namely, Geonames, BDPedia and LinkedGeodata (2016). Jain, Hitzler, Sheth, Verma and Yeh (2010) have developed a system, called BLOOMS for finding schema-level links between Linked Open Data sets (LOD). BLOOMS is based on the idea of bootstrapping information already present in the LOD cloud.

Ontology enrichment – more and more reuse of LD sources and the availability of dereferenceable links will enable the easier extension of the ontologies. Each time we find a new Linked Data set on the web for a given domain, we can cross check the data elements (the properties) and their existence in an ontology. In the case of their unavailability, we add them into the ontology. This will ensure the incremental extension and consequently the richness of an ontology.

6. Discussion and Conclusion

From the above discussion, we can observe that both ontology and LD have lots to offer each other. For instance, an ontology can contribute to LD in terms of making the data sets semantic compatible, support in reasoning and inferencing new knowledge and enhancing the data quality and acquiring the social trust. While LD, on the other hand, can revolutionize the way we construct the ontologies. LD can render support in identifying the domain requirements and terminologies, promote application-driven ontology development and support in data-driven ontology modeling. Hence, we can conclude that both ontology and LD are not here to compete with each other, rather they are to empower each other and ultimately to empower the web. Both of them together can support in achieving the vision of Semantic Data Web and making a true semantic information space.

The complementary features of both the technologies have also brought to light some research questions. For instance, how do we decide which ontology to be used for what kinds of data (as we know an ontology cannot fit all kinds of data even when the domain of the ontology and the data are the same); how to promote the reuse of ontologies in LD, and at what level; what skills are needed for ontology engineers to develop ontologies for LD; how ontologies should be built in LD era? In the current work, we could not resolve the issue of the kind of ontology (i.e. lightweight, heavyweight, middleweight) should be used in LD applications. Although many have favored the lightweight ontologies, we think the selection would depend on the complexity of the data and tasks in hand. Hence, it is important to investigate it further before reaching a general conclusion.

References

Adhikari, A., Singh, S., Dutta, A. and Dutta, B. (2015), “A novel information theoretic approach for finding semantic similarity in WordNet”, Proceedings of IEEE International Technical Conference, Macao, November 1-4, pp. 1-6.

Alonso, J.B. (2006), “Ontology-based software engineering support for autonomous systems”, available at: http://tierra.aslab.upm.es/documents/controlled/ASLAB-R-2006-06-v1-Draft-JB.pdf (accessed July 12, 2016).

Arenas, M., Grau, B.C., Kharlamov, E., Marciuska, S. and Zheleznyakov, D. (2014), “Faceted search over ontology-enhanced RDF data”, Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, New York, NY, pp. 939-948.

Basic Formal Ontology (2002), available at: http://purl.obolibrary.org/obo/bfo.owl (accessed June 28, 2016).

Bechhofer, S., Buchanb, I., Roured, D.D., Missiera, P., Ainsworthb, J., Bhagata, J., Couchb, P., Cruickshankc, D., Delderfieldb, M., Dunlopa, I. and Gamble, M. (2013), “Why linked data is not enough for scientists”, Future Generation Computer Systems, Vol. 29 No. 2, pp. 599-611.

Becker, C. and Bizer, C. (2009), “Marbles”, available at: http://mes.github.io/marbles/ (accessed May 5, 2016). Becker, K.G., Barnes, K.C., Bright, T.J. and Wang, S.A. (2004), “The genetic association database”, Nature Genetics, Vol. 36 No. 5, pp. 431-432.

Benefits of the Linked Data Approach (2011), available at: www.w3.org/2005/Incubator/lld/wiki/Benefits (accessed July 12, 2016).

Berners-Lee, T. (2006), “Linked data”, available at: www.w3.org/DesignIssues/LinkedData.html (accessed May 10, 2016).

Berners-Lee, T., Hendler, J.A. and Lassila, O. (2001), “The semantic web”, Scientific American, May, pp. 29-37.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A. and Sheets, D. (2006), “Tabulator: exploring and analyzing linked data on the semantic web”, Proceedings of the 3rd International Semantic Web User Interaction Workshop (SWUI 2006), Athens, Georgia, pp. 1-16.

Bertoa, M., Vallecillo, A. and Garcia, F. (2006), “An ontology for software measurement”, in Calero, C., Ruiz, F. and Piattini, M. (Eds), Ontologies for Software Engineering and Software Technology, Springer-Verlag, Berlin, pp. 175-196.

Bestbuy (2009), available at: www.bestbuy.com/ (accessed July 7, 2016).

BioPortal (2005), “Get annotations for biomedical text with concepts from ontologies”, available at: http://bioportal.bioontology.org/annotator (accessed June 27, 2016).

Bizer, C., Heath, T. and Berners-Lee, T. (2009), “Linked data – the story so far”, International Journal on Semantic Web and Information Systems, Vol. 5 No. 3, pp. 1-22.

Celli, F., Anibaldi, S., Folch, M., Jaques, Y. and Keizer, J. (2011), “OpenAGRIS: using bibliographical data for linking into the agricultural knowledge web”, Proceedings of SNLP-AOS-2011, Bangkok, available at: www.fao.org/3/a-an642e.pdf (accessed May 6, 2016).

Chemical Entities of Biological Interest (2017), available at: ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/ (accessed June 28, 2016).

Cheng, G., Ge, W. and Qu, Y. (2008), “Falcons: searching and browsing entities on the semantic web”, World Wide Web, Beijing, April 21-25, pp. 1101-1102.

Corcho, O., Poveda-Villalon, M. and Gomez-Perez, A. (2015), “Ontology engineering in the era of linked data”, Bulletin of the American Society for Information Science and Technology, Vol. 41 No. 4, pp. 13-17.

DBPedia (2016), “Downloads 2015-10”, available at: http://wiki.dbpedia.org/Downloads2015-10 (accessed June 27, 2016).

Doms, A. and Schroeder, M. (2005), “GoPubMed: exploring PubMed with the Gene ontology”, Nucleic Acids Research, Vol. 33, Web Server Issue, pp. 783-786.

Dutta, B. (2008), “Semantic web services: a study of existing technologies, tools and projects”, DESIDOC Journal of Library and Information Technology, Vol. 28 No. 3, pp. 47-55.

Dutta, B. (2014), “Symbiosis between an ontology and linked data”, Librarian, Vol. 21 No. 2, pp. 15-24.

Dutta, B., Nandini, D. and Shahi, G. (2015), “MOD: metadata for ontology description and publication”, Proceedings of DCMI International Conference on Dublin Core and Metadata Applications, Sao Paulo, September 1-4, pp. 1-9.

Euzenat, J. and Shvaiko, P. (2007), Ontology Matching, Springer-Verlag, New York, NY.

Fernandez, M., Gomez-Perez, A. and Juristo, N. (1997), “Methontology: from ontological art towards ontological engineering”, Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering, Stanford, CA, pp. 33-40.

Freebase (2009), available at: https://datahub.io/dataset/freebase (accessed May 5, 2016).

Freudenberg, M., Kontokostas, D. and Hellmann, S. (2015), “The DBPedia Data Set (2015-04)”, available at: http://wiki.dbpedia.org/dbpedia-data-set-2015-04 (accessed July 9, 2016).

Fu, G., Jones, C.B. and Abdelmoty, A.I. (2005), “Ontology-based spatial query expansion in information retrieval”, On the move to meaningful internet systems: CoopIS, DOA, and ODBASE, Springer, Berlin, pp. 1466–1482.

Furber, C. and Hepp, M. (2011), “Towards a vocabulary for data quality management in semantic web architectures”, Proceedings of LWDM, Uppsala, March 25, pp. 1-8.

Garcia-Silva, A., Garcia-Castro, L.J., Garcia, A. and Corcho, O. (2014), “Social tags and linked data for ontology development: a case study in the financial domain”, Proceedings of the 4th

International Conference on Web Intelligence, Mining and Semantics (WISM'14, Thessaloniki, Greece), ACM, New York, NY, June 2-4, pp. 1-10.

Gene Ontology (2016), available at: http://geneontology.org/ (accessed May 9, 2017).

Geonames (2017), available at: www.geonames.org/ (accessed May 9, 2017).

Geonames (2016), “Entry points into the GeoNames Semantic Web”, available at: www.geonames.org/ontology/documentation.html (accessed July 10, 2016).

Giunchiglia, F. and Dutta, B. (2011), “DERA: a faceted knowledge organization framework”, available at: http://eprints.biblio.unitn.it/archive/00002104/ (accessed July 9, 2016).

Giunchiglia, F., Dutta, B. and Maltese, V. (2009), “Faceted lightweight ontologies”, in Borgida, A., Chaudhri, V., Giorgini, P. and Yu, E. (Eds), Conceptual Modeling: Foundations and Applications, Vol. 5600, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Heidelberg and Berlin, pp. 36-51.

Giunchiglia, F., Maltese, V., Farazi, F. and Dutta, B. (2010), “GeoWordNet: a resource for geo-spatial applications”, in Aroyo, L., Antoniou, G., Hyvönen, E., Teije, A.T., Stuckenschmidt, H., Cabral, L. and Tudorache, T. (Eds), Semantic Web: Research and Applications, Vol. 6088, Lecture Notes in Computer Science (LNCS), Springer-Verlag, Berlin and Heidelberg, pp. 121-136.

GoodRelations: the Web vocabulary for e-commerce (2011), available at: www.heppnetz.de/ontologies/goodrelations/v1.html (accessed July 10, 2016).

Grangel-Gonzalez, I., Halilaj, L., Coskun, G. and Auer, S. (2015), “Towards vocabulary development by convention”, Proceedings of 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Lisbon, pp. 334-343.

Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Ebeling, C. and Schomburg, D. (2011), “The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources”, Nucleic Acids Research, Vol. 39, Database Issue, pp. D507-D513, doi: 10.1093/nar/gkq968.

Gruber, T.R. (1993), “A translation approach to portable ontologies”, Knowledge Acquisition, Vol. 5 No. 2, pp. 199-220.

Gruninger, M. and Fox, M. (1995), “Methodology for the design and evaluation of ontologies”, Proceedings of workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, pp. 1-10.

Halpin, H. and Presutti, V. (2009), “An ontology of resources for linked data”, WWW2009 Workshop on Linked Data On the Web (LDOW2009), April 20, Madrid, pp. 1-8, available at: http://events.linkeddata.org/ldow2009/papers/ldow2009_paper19.pdf (accessed July 10, 2016).

Heath, T. and Motta, E. (2007), “Revyu.com: a reviewing and rating site for the web of data”, Proceedings of 6th International Semantic Web Conference, Busan, pp. 1-8, available at: http://data.semanticweb.org/pdfs/iswc-aswc/2007/ISWC2007_SWC_Heath.pdf (accessed on July 10, 2016).

Heath, T. and Motta, E. (2008), “Revyu: linking reviews and ratings into the web of data”, Journal of Web Semantics, Vol. 6 No. 4, pp. 266-273.

Hendler, J.A. (2009), “A little semantics goes a long way”, available at: www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html (accessed July 9, 2016).

Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A. and Decker, S. (2012), “An empirical survey of linked data conformance”, Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 14, pp. 14-44, available at: www.websemanticsjournal.org/index.php/ps/issue/view/42

Jain, H., Thao, C. and Zhao, H. (2012), “Enhancing electronic medical record retrieval through semantic query expansion”, Information Systems and e-Business Management, Vol. 10 No. 2, pp. 165-181.

Jain, P., Hitzler, P., Sheth, A.P., Verma, K. and Yeh, P.Z. (2010), “Ontology alignment for linked open data”, 9th International Semantic Web Conference, Shanghai, November 7-11, pp. 402-417.

Jain, P., Hitzler, P., Yeh, P.Z., Verma, K. and Sheth, A.P. (2010), "Linked data is merely more data", in Brickley, D., Chaudhri, V.K., Halpin, H. and McGuinness, D. (Eds), Linked Data Meets Artificial Intelligence, Technical Report No. SS-10-07, AAAI Press, Menlo Park, CA, pp. 82-86.

Janowicz, K. and Hitzler, P. (2013), "Thoughts on the complex relation between linked data, semantic annotations, and ontologies", Proceedings of the 6th International workshop on Exploiting Semantic Annotations in Information Retrieval, San Francisco, CA, October 27-November 1, pp. 41-44.

KBSI (1994), "The IDEF5 ontology description capture method overview", KBSI Report, TX.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z.T., Han, B, Zhou, Y. andWishart, D.S. (2014), "DrugBank 4.0: shedding new light on drug metabolism", Nucleic Acids Research, Vol. 42 No. 1, pp. D1091-D1097.

LinkedGeoData (2016), "Adding a spatial dimension to the Web of Data", available at: http://linkedgeodata.org/About (accessed July 10, 2016).

Livingston, K.M., Bada, M., Baumgartner, W.A. and Hunter, L.E. (2015), "KaBOB: ontology-based semantic integration of biomedical databases", BMC Bioinformatics, Vol. 16 No. 126, pp. 1-21, doi: 10.1186/s12859-015-0559-3.

LODStats (2016), available at: http://lodstats.aksw.org/ (accessed July 9, 2016).

McGuinness, D.L. (1998), "Ontological issues for knowledge-enhanced search", in Guarino, N. (Ed.), Proceedings of 1st International Conference on Formal Ontology of Information Systems, IOS Press, Trento, pp. 302-316.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. and Schneider, L. (2003), "Dolce: a descriptive ontology for linguistic and cognitive engineering", WonderWeb Project, Deliverable D17 v2, pp. 75-105.

MusicBrainz (2000), available at: https://musicbrainz.org/ (accessed July 7, 2016).

Noy, N.F. (2004), "Semantic integration: a survey of ontology-based approaches", ACM SIGMOD Record, Vol. 33 No. 4, pp. 65-70.

Noy, N.F. and McGuinness, D.L. (2001), "Ontology development 101: a guide to creating your first ontology", available at: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf (accessed July 10, 2016).

Obrst, L., Gruninger, M., Baclawski, K., Bennett, M., Brickley, D., Berg-Cross, G., Hitzler, P., Janowicz, K., Kapp, C., Kutz, O., Lange, C., Levenchuk, A., Quattri, F., Rector, A., Schneider, T., Spero, S., Thessen, A., Vegetti, M., Vizedom, A., Westerinen, A., West, M. and Tim, P. (2014), "Semantic web and big data meets applied ontology", Applied Ontology, Vol. 9 No. 2, pp. 155-170.

OpenCyc for the Semantic Web (2012), available at: http://sw.opencyc.org/ (accessed June 27, 2016).

Ouksel, A.M. and Sheth, A. (1999), "Semantic interoperability in global information systems", ACM SIGMOD, Vol. 28 No. 1, pp. 5-12.

OWL Web Ontology Language Overview: W3C Recommendation (2004), Deborah L. McGuinness and Frank van Harmelen (Eds), February 10, available at: www.w3.org/TR/owl-features/ (accessed July 12, 2016).

Parundekar, R., Knoblock, C.A. and Ambite, J.L. (2010), "Aligning geospatial ontologies on the linked data web", Proceedings of the GIScience Workshop on Linked Spatiotemporal Data, Zurich, pp. 1-11.

Poggi, A., Lembo, D., Calvanese, D., Giacomo, G., De, Lenzerini, M. and Rosati, R. (2008), "Linking data to ontologies", in Spaccapietra, S. (Ed.), Journal on Data Semantics X, Lecture Notes in Computer Science, Vol. 4900, Springer, Berlin and Heidelberg, pp. 133-173.

Poveda-Villalon, M., Suárez-Figueroa, M.C. and Gómez-Pérez, A. (2014), "A reuse-based lightweight method for developing linked data ontologies and vocabularies", in Simperl, E., Cimiano, P., Polleres, A., Corcho, O. and Presutti, V. (Eds), The Semantic Web: Research and Applications, ESWC 2012. Lecture Notes in Computer Science, Vol. 7295, Springer, Berlin and Heidelberg, pp. 833-837.

Ranganathan, S. (1997), "Design of depth classification", in Neelameghan, A. (Ed.), S. R. Ranganathan's Postulates and Normative Principles: Applications in Specialized Databases Design, Indexing and Retrieval, SRELS, Bangalore, pp. 75-126.

Rowe, M. and Butters, J. (2009), "Assessing trust: contextual accountability", Proceedings of the Trust and Privacy on the Social Semantic Web Workshop, European Semantic Web Conference, Heraklion, June 1, pp. 1-12.

Rula, A. and Zaveri, A. (2014), "Methodology for assessment of linked data quality", 1st Workshop on Linked Data Quality, September 2, Leipzig, pp. 1-4.

Sanderson, M. (1994), "Word sense disambiguation and information retrieval", in Croft, W.B. and van Rijsbergen, C.J. (Eds), Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Springer, London, Dublin, July 3-6, pp. 142-151.

schema.org (2011), available at: http://schema.org/ (accessed June 22, 2016).

Sig.ma (2011), available at www.w3.org/2001/sw/wiki/Sig.ma (accessed May 5, 2016).

SPARQL Query Language for RDF: W3C Recommendation (2008), Prud'hommeaux, E. and Seaborne, A. (Eds), January, 15 available at: www.w3.org/TR/rdf-sparql-query/ (accessed July 12, 2016).

Studer, R., Benjamins, V.R. and Fensel, D. (1998), "Knowledge engineering: principles and methods", Data and Knowledge Engineering, Vol. 25 Nos 1/2, pp. 161-197.

Studer, R., Simperl, E. and Kampgen, B. (2011), "Linked data and ontologies", STI Semantic Summit, July 6-8, Riga.

Suggested Upper Merged Ontology (2016), available at: www.adampease.org/OP/ (accessed July 7, 2016).

Swoogle (2007), available at: http://swoogle.umbc.edu/ (accessed May 5, 2016).

Thorsen, H. and Pattuelli, C. (2015), "Ontologies in the time of linked data", NASKO-2015, Los Angeles, CA, June 18-19.

UniProt Gene Ontology Annotation (2016), available at: www.uniprot.org/ (accessed June 28, 2016). Uschold, M. (2008), "Ontology-driven information systems: past, present and future", Proceedings of the Fifth International Conference on Formal Ontology in Information Systems, Saarbrücken, October 31-November 3, pp. 3-18.

Uschold, M., King, M., Moralee, S. and Zorgios, Y. (1995), "The enterprise ontology", The Knowledge Engineering Review, Vol. 13 No. 1, pp. 31-89.

Van Heijst, G., Schreiber, A. and Wielinga, B. (1997), "Using explicit ontologies in KBS development", International Journal of Human-Computer Studies, Vol. 46 Nos 2/3, pp. 183-292.

Vrandecic, D., Pinto, S., Tempich, C. and Sure, Y. (2005), "The DELIGENT knowledge processes", Journal of Knowledge Management, Vol. 9 No. 5, pp. 85-96.

W3C (2015), "Linked data", available at: www.w3.org/standards/semanticweb/data (accessed July 22, 2016).

Watson: exploring the Semantic Web (2010), available at: http://watson.kmi.open.ac.uk/WatsonWUI/ (accessed June 27, 2016).

Weijian Xuan, M.D., Mirel, B., Jean Song, A., Brian, W., Stanley, J. and Meng, F. (2009), "Open biomedical ontology-based Medline exploration", BMC Bioinformatics, Vol. 10 Nos S5/S6, pp. 1-14.

Xenarios, I., Rice, D.W., Salwinski, L., Baron, M.K., Marcotte, E.M. and Eisenberg, D. (2000), "DIP: the database of interacting proteins", Nucleic Acids Research, Vol. 28 No. 1, pp. 289-291.

Zaveri, A., Kontokostas, D., Mohamed, A.S., Bühmann, L., Morsey, M., Auer, S. and Lehmann, J. (2013), "User-driven quality evaluation of DBpedia", Proceedings of the 9th International Conference on Semantic Systems, Graz, September 4-6, pp. 97-104.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. and Auer, S. (2012), "Quality assessment for linked data: a survey", Semantic Web – Interoperability, Usability, Applicability, Vol. 1 Nos 1/5, pp. 1-31.