ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 7(6), 745–764, June 2014* ISSN (Online) : 0974-5645

Linked Open Data: Towards the Realization of Semantic Web- A Review

Shah Khusro, Fouzia Jabeen*, Syed Rahman Mashwani and Iftikhar Alam

Department of Computer Science, University of Peshawar, Peshawar-25000, Pakistan; khusro@upesh.edu.pk, fouzia.jabeenphd@upesh.edu.pk, syed.rahman@upesh.edu.pk, iftikharalam@upesh.edu.pk

Abstract

This paper provides the reader from the base to the state of art in Linked Open Data (LOD), with issues and challenges. In addition, reader will be motivated by reading the projects analysed in the information space of five major computer science areas (Intelligence, Multimedia, Sensors, File System and Library), future trends and directions in LOD.

Keywords: Dereferencing, Linked Open Data, LOD Cloud, RDF, URI

1. Introduction

The Web contains tons of information, but the raw data itself not available rather only HTML documents constructed from data are displayed. The semantic web seeks to change the view of the internet regarding this problem in a number of ways. It persuades organizations, companies and individuals to bring out their data freely in an open standard format. It pushes and encourages the use of data already available on the Web. By doing so, an environment of giving and taking of data will be created. All the data available on the Web will be treated and researched as one database[34]. The aim is to share and reuse the existing data.

In order to understand the thought and worth of Linked Open Data (LOD), it is important to consider existing mechanisms for sharing and reusing data on the Web. A key aspect in the re-usability of data is the degree to which it is well structured. The more well-defined the structure of the data is, the more easily and effectively people can make tools to reliably process it for reuse. Web applications have some extent of structure. HTML is the language in which they are formed, and HTML, focuses on structuring textual documents rather than data. As data is mixed with the surrounding text, it is difficult for software applications to retrieve snippets of structured data from HTML pages. To deal with this issue, varieties of microformats were developed. Microformats tightly state how to embed data, so that applications can clearly dig out the data from the pages. However the major drawback of microformats is that they are limited to represent data about a small and restricted set of entities. And only small set of attributes that might be used to illustrate these entities. It is often not possible to convey relationships between entities. Therefore microformats are not appropriate for sharing arbitrary data on the Web.

A better and generic way to make structured data available on the Web is Web API. Web API offers simple query access to structured data over HTTP protocol. Examples of these API include Flickr API, Delicious API and Amazon Product Advertising API etc. The Programmable Web Website maintains a directory containing thousands of Web APIs. The advent of Web APIs has led to the development of applications called mashups. API gives an indisputable benefit of programmatic access to structured data, however every programmer who wants to retrieve data must have to understand the methods available to extract data using API, and then write custom code for retrieving data from each data

*Author for correspondence

source. Web APIs give result in structured data formats like XML and JSON, which have widespread support in most of the programming languages. However, they have some boundaries that can be understood if we compare it with HTML. HTML defines the anchor tag (<*aI*) with one of the valid and important attributes *href*. When used together, the anchor tag and href attribute point towards an outgoing link from the current document. Web user agents, such as browsers and search engine crawlers, are programmed to identify the worth of this combination. This connectivity between documents thus has formed the Web of documents. By contrast, data returned from most of the Web APIs do not have the equivalence of the HTML anchor tag and href attribute, to indicate links that should be followed for finding related data. As a result, data returned from Web APIs usually exist in the form of isolated fragments, with missing reliable onward links. Therefore, while Web APIs make data accessible on the Web, they do not position it truly on the Web, to make it linkable and discoverable.

Web Services are really no better than Web API. They are very difficult to develop and also consume much more resources (for example bandwidth and RAM). It involves the conversion (Request->XML->Data->Response->XML->Validation->Conversion->Data), that makes it very slow and resulting in lack of likeability and discoverability.

The capability for anyone to add new documents to the Web, and then for these documents to be automatically exposed by search engines to humans via browsers, have historically been key components for Web's massive growth. The same ideology of linking, and thus ease of discovery, can be applied and implemented on the data. LOD provides a technical solution to make such linkage practical.

The paper is organized in five sections. Section 2 gives answers to several questions that include how data on web can be related, how to create LD and incorporate it in LOD, graphical analysis of LOD, comparing traditional upper ontology and LOD approach. Section 3 explains the tools and applications of LOD with issues and challenges. In section 4 the specialized projects developed in the information space composed of Intelligence, Multimedia, File system, Sensors and Library have been analyzed. Section 5 concludes the paper and also states the future research directions in LOD.

2. Relating Data on Web

2.1 Open vs. Linked Data vs. Linked Open Data

Linked Data[68] (LD) is a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the semantic web using URIs and RDF.

LD and LOD are the two different terms. The difference is LOD is LD which is released under an open license. This does not impede its reuse for free. However, it is not necessary for LD to be open. The label "Linked Open Data" is widely used, but often to refer LD in general, rather than to LD that is explicitly published under an open license. It is therefore important to note that not all LD is open, and not all Open data is linked. So appropriate term (Open data, Linked Data, Linked Open Data) should be used, depending on the licensing terms of the data in question. Figure 1(a) (b) is the representation of the LOD (called LOD cloud) containing the datasets that have been published in the LD format and part of LOD cloud for the year 2007 and 2011 respectively. Click on any of the circles will take you to the home page of the dataset. The size of the circle in the figures depicts the

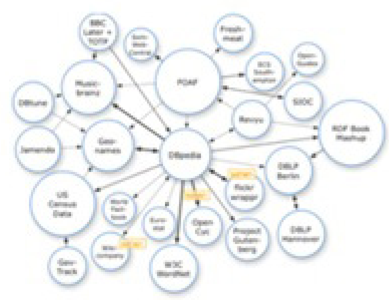

Figure 1. (a) Linked Open Data Cloud (LOD cloud)[65].

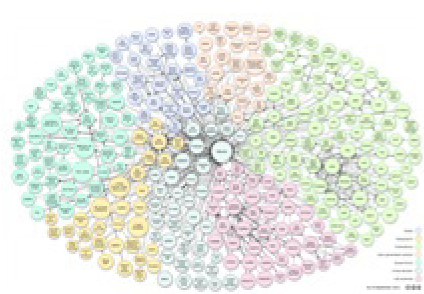

Figure 1. (b) Linked Open Data Cloud[67].

amount of data in the dataset. Arrows and the thickness of the lines represents the number of links and other associated information[68].

The first cloud diagram was created in 2007 with one dataset. Today there are more than 295 datasets[62]. Figure 2 gives an idea of increase in the number of the datasets included in LOD cloud. Contributions in LOD cloud are from the LOD community projects and involvement of various individuals and organizations[65].

2.2 Creating Well Formed Linked Data

Tim Berners-Lee gave a presentation on LD at the TED 2009 conference. In it, he restated the LD principles as three extremely simple rules. The design issues analogous to the stated principles are paraphrased in the following lines.

- ….Use *URIs* to recognize things.
- ….Use *HTTP URIs* so that these things can be referred to and looked up "*dereferenced*" by people and user agents.
- ….Provide useful information about the thing using standard formats such as *RDF/XML* when it's URI is dereferenced.
- ….links to other related URIs must be exposed to perk up the discovery of other related information on the web.

The above mentioned steps can be called rules because if we break them it will not destroy anything but data will not be interconnected. Further it will limit the reuse of the data[8].

2.2.1 Dereferencing: Traditional vs. Semantic Web

Dereferencing term is used in both traditional and semantic web. The question is how they differ? The difference in dereferencing is observed by realizing the fact that in traditional web URI is resolved by retrieving a document representation of the resource with status code "200 OK"[74]. URI is successfully resolved in case if resource is Information Resource. Non-Information Resource cannot be dereferenced directly. Instead of sending a representation of the resource, the server by using the "*HTTP response code 303 See Other*" sends the URI of an information resource which describes the non-information resource. With one more step, the client dereferences this new URI and gets the representation that will describe the original non-information resource.

In semantic web, by resolving a Real World Object URI (dereferencing URI), we cannot retrieve the object itself, but will be redirected to an information resource (a RDF document) which represents the URI. By parsing the RDF, the meaning of the URI will be known according to the RDF triples portrayal (dereferencing succeeded).

2.2.2 URI, URL and URN Distinct Terms

Confusion between the three related terms URI, URL and URN can be resolved by understanding that a URL is a URI and a URN is a URI, clear and simple. It's like trying to make clear that a Human is a Mammal but a Mammal is not always a Human.

There is no reasonable way to give examples that show the difference between the URL and URI. Tim Berners-Lee document regarding URI highlights one of the very important points called the URI opacity. The idea behind *URI Opacity* is that one should not look at the string to make any decision as to what is at the other end and just take URI as an identifier. All the examples given in Table 1 are valid URIs, even the presence of a file extension does not distinguish a URI from a URL. Furthermore, each URI is unique, means it cannot be assumed that "*http://www.abc/home*" is the same as "*http://www.abc/home.html*" just

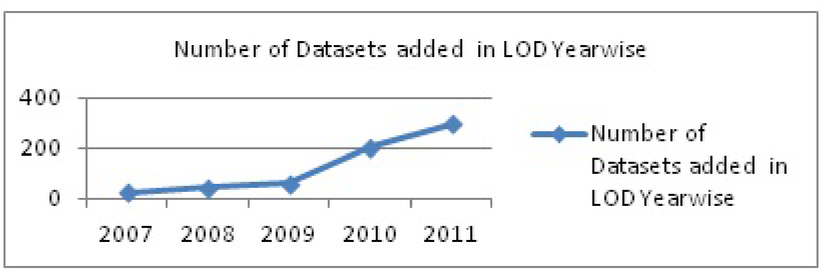

Figure 2. Increase of number of datasets in LOD Cloud.

by looking at the URI. These are both separate URIs that may or may not correspond to the same resource.

However, we can provide examples as given in Table 2 that enable us to differentiate URL from a URN. Observe all of the URNs are prefixed with "urn"[71].

2.3 To be Part of LOD

In order to add dataset in LOD, it is required to publish data according to the LD principles as discussed in section 2.2. After that verify it with following check list[75].

- ….There must be a resolvable *http://* (*or https://*) URIs.
- ….URIs must resolve to RDF (in any one of the accepted RDF formats that include RDFa, RDF/XML, Turtle, N-Triples).
- ….The dataset should be made up of at least 1000 triples.
- ….The dataset must be interlinked via RDF links to the datasets that are already in the LOD cloud diagram. This means, your dataset have to use URIs from the other datasets, or vice versa. We arbitrarily need at least 50 links.
- ….Access of the complete dataset must be achievable by either RDF crawling, or by an RDF dump, or by SPARQL endpoint.

After confirming that dataset follows above mentioned key points, add dataset to CKAN[75]. CKAN is an open registry of data and content package. Data Hub LOD Validator adds new dataset to Data Hub or edits existing dataset.

Table 1. Valid URI's[71]

1) http://www.domain.tld/somepath/file.php?mykey=somevalue
2) http://www.domain.tld/somepath/file.php" part is an URL?
3) www.abc/home
4) www.abc/home.html
5) mailto:someone@example.com
6) https://github.com/afs/TDB-BDB.git
7) file:///home/someuser/somefile.txt

Table 2. Examples of URNs[71]

1) urn:mpeg:mpeg7:schema:2001urn:isbn:0451450523
2) urn:sha1:YNCKHTQCWBTRNJIV4WNAE52SJUQCZO5C
3) urn:uuid:6e8bc430-9c3a-11d9-9669-0800200c9a66

2.4 LOD Motivation and Use

Let's have look at the scenarios that motivated the development of LOD. We will discuss scenarios from two points of views firstly scenarios common to all the areas and second in the information space of five major computer science areas (File system, Intelligence, Multimedia, Sensors and Library).

- ….A person A wants to consume data from x different sources which are heterogeneous in nature. Current information integration techniques are ineffective and time consuming[11]. LOD provides a brilliant approach to publish and use data on the Web and make the Web to be a global data space which can be both machine and computer consumable. Figure 3 depicts how LOD concept made web as single global data space. If people can develop interlinkage between different datasets, they will be able to enrich datasets with relevant and related information from the interlinked datasets (e.g. if a webpage is about the city Paris in France, it would be valuable to add related data about this city from Wikipedia). Data for inferring new information from already present information, is an added motivational factor of using LOD, for example in pharmaceutical applications or IBM Watson[20].
- ….Data in LOD sources could be related to desktop in many scenarios. But most of the desktop applications are unaware of URIs, RDF triples and therefore cannot directly consume data in LOD. Currently people explore or consume this data by switching between desktop and Web contexts or environments. They download and export related data to application specific format in order to be used in desktop environment. So, if these two different disconnected worlds are bridged together then desktop applications could be able to directly consume related data in LOD[57].

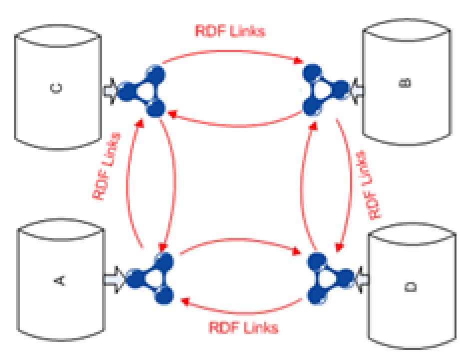

Figure 3. Linked data as single global data Space.

A large portion of personal and organizations' data is kept in file systems. Data in file system is disconnected from semantically related data on another file system or Web of data. And it is difficult to integrate them because of their current identifier. So, by exposing file system according to LD principles, it becomes easy to integrate file system objects to other (personal, organizational internal or global) data sources. This would also enable us to semantically annotate file system objects, as in current modern file systems there is not a proper mechanism for platform independent and arbitrary annotations of files[59].

- ….In April 1980 United States conducted an operation with the name "Eagle Claw". Its objective was to rescue the 52 Americans held captive at the United States Embassy at Tehran[98]. Although like sandstorm or Haboob the operation was failed. Besides the inadequacy of navigation system of two helicopters and the hydraulic system of one helicopter (that were taking part), another factor for failure was lack of open source intelligence. The personnel involved in the planning of the "Operation Eagle Claw" had very limited geographic and terrain information, also they had no or very less information about the attitude of the local people, and their information about the weather on the day of operation was not sufficient. The kinds of information mentioned above came under the category of *Open Source Intelligence*. This intelligence information is very important in conducting the operations successfully.

The question arises that what is the connection between the LOD and Open Source Intelligence? The answer is that LOD offers a facility to acquire the Open Source Intelligence. Next question is how it provides the Open Source Intelligence? The answer is by providing information present in its datasets which are semantically linked with other datasets (like Weather reports, News, Geographic and Terrain information etc). This information plays a vital role in conducting the operation successfully. For example weather report enables the planning crew to identify the right equipment, weapon, and logistic support. The news reports can help the crew about the attitude of the local people etc.

- ….LOD based video annotation system provides the facility to user to browse and annotate whole video, scene, event, and objects of the videos in addition to interlinked related videos, scenes, events and objects of the videos that are present on different data sources. Suppose a politician discuss the issue on human rights in a specific scene, event or in a video and on the other side he discuss the same issue on the other video that is present on the other data source, LOD can interlink the related scene, events, objects or videos with each other.
- ….Sensor data can be annotated with contextual information by linking with LOD cloud. A driver can drive faster if he follows the routes suggested by GPS navigation system of his car. The route suggestions will be based on composite set of data accessed from LOD datasets such as information about the hilly surrounding area (from Geography LOD datasets), information about nearby road works (from Government LOD datasets), and information about ongoing social events in the locale (from Media LOD datasets).
- ….Dr X, a Ph. D scholar wants to seek the information about the novel and innovative work on which he can do further research. His key requirements are as follows :
  - …The scholar searches for unique work and in process also make relationship with his novel ideas and the existing ideas in the area.
  - ….Different tasks related to literature search and analysis like to locate the current status of the scholar's ideas in the recent literature and to which extent the work about the idea is described.
  - ….The scholarly hierarchy and relationship chain of the idea.
  - …..How other domains and the current domain is influenced by this idea i.e. someone else's work on this idea to generalize or specify it for some other area.
  - ….What is the relationship of aspects of an idea to different other ideas and as a whole relationship graph, its significance, perspective or view point, interrelationship or comparative importance of the idea. That is, has someone from different schools of thought worked on the similar topic with different view point?[61].
  - ….The current work is the follow up of the existing theories, ideas, policies, strategies or plans or it is contradictory i.e. any evidence against the idea. That is working on the same idea but in contradicting way.

So the said scholar wants to discover such platform which can give the handy services, information of the resources, interlinked semantic relationship among the resources, the rich data resources. His/her work will be easy by adopting the LOD services and datasets.

2.5 Graphical Analysis of LOD Cloud

The LOD cloud visualization shown in Figure 1(a)(b) is based on metadata collected and accumulated by contributors to the CKAN directory[67]. Many researches have been carried out on graphical analysis of LODcloud. A few prominent conclusions and insights are discussed in this section.

- ….Linking Open Data Graph-A graph version of LOD cloud[42] as shown in Figure 4, emphasizes and represents the ratings of datasets. Datasets with high average rating are represented in green colour, and those with low average rating are coloured red. Intensity of the colour displays the number of received ratings that is strong colour indicates many ratings, while a near-white colour indicates few ratings, and unrated datasets are white.
- ….Empirical Cloud-This is a graph of *owl:sameAs* links (between hosts) in the Billion Triple Challenge data[26]. In order to make the graph viewable it was restricted to hosts that share ten or more links. It is possible to zoom or pan on portions of the graph, and hover over a node to see the hostname. Node colour represent the host types like green represent org, yellow for com as shown in Figure 5.
- ….Lumpy structure of LOD- Christophe Guéret[39] analyzed the network structure of LOD and characterizes it as lumpy structure. He argued that even the LOD gives impression of high interconnection but experimentally he deduce that, LOD is not one cloud, but actually they are three clouds. Internally these clouds are highly interconnected while between these clouds interconnection is sparse. One cloud is related to bio or life sciences data, second is academic bibliographic material and the third one is all the rest, which connects with other two having DBPedia as its hub as shown in Figure 6.
- ….LOD Cloud analysed with Gephi– Figure 7 shows the outlook of LOD cloud after clustering[43].
- …. Graph Analysis of LOD-Marko A. Rodriguez[56] analysed the network structure of an early 2009 version of LOD cloud diagram. This visual representation was manually changed into a directed graph as shown in Figure 8 and was further analysed. Web of data maintains publicly accessible interrelated data. In the relational database world (closed world), hardly ever

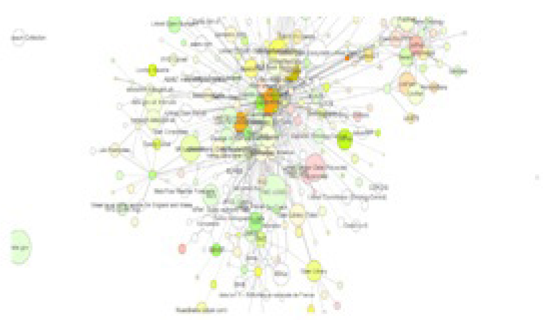

Figure 4. Linking Open Data graph[42].

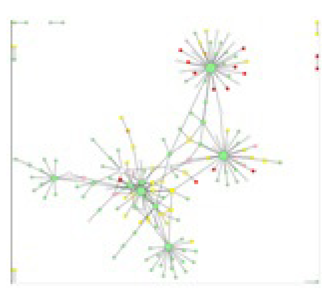

Figure 5. Empirical cloud[26].

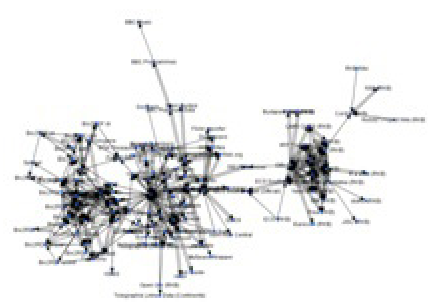

Figure 6. Lumpy structure of LOD[39].

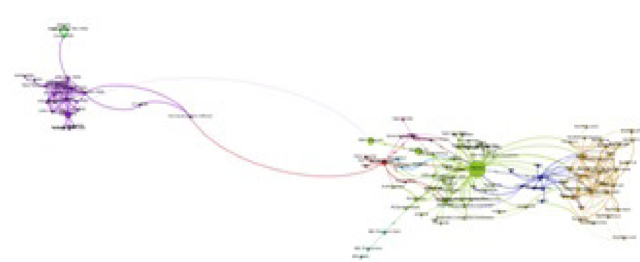

Figure 7. LOD Cloud analysed with Gephi[43].

Figure 8. Graph analysis of the Linked Data Cloud[56].

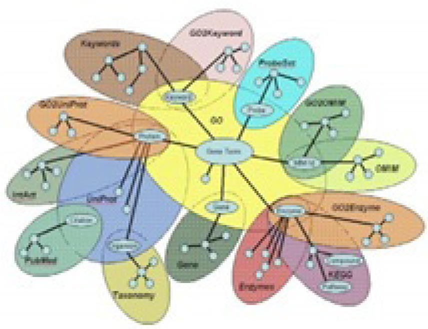

Figure 9. Aggregation of bioinformatics data using semantic web technology[63].

are database ports made publicly available for harvesting or relational schemas published for reuse.

- ….Aggregation of Bioinformatics data using semantic web technology-Stephens et al.[63] has developed an analogous diagram as shown in Figure 9 for LOD cloud. Diagram represents interconnection between several bioinformatics datasets. This diagram was repeatedly used by Tim Berners-Lee in his presentations.

2.6 LOD and Traditional Ontology based Approach

Standard upper ontology working group is an IEEE-sanctioned working group of experts from fields of engineering, philosophy and information science. They have proposed Upper Merged Ontology (SUMO) as a starter document. SUOWG provide definitions for general-purpose terms. SUMO was formed by merging publicly obtainable ontological content into a single, comprehensive and cohesive structure. The sources that had been included in SUMO are ontologies developed by ITBM-CNR, John Sowa's upper level ontology, and various mereotopological theories[49].

SUMO had been mapped to WordNet (a structured lexicon of English meanings). The purpose of such integration is to promote use of SUMO in natural language processing tasks such as sense disambiguation, summary generation etc. Moreover this integration also tests the coverage of ontology. With creation of mappings between SUMO and WordNet, gaps in conceptual space has been identified, primarily of the type when most specific concept in SUMO could be at best mapped to a broad (meaning wise) term in WordNet. This issue required creation of new, more specific concepts in SUMO.

Aside from developing the SUMO and creating the mappings from SUMO to WordNet, domain ontologies are also created and aligned with the SUMO. These domain ontologies inherit the broad conceptual distinctions of SUMO and specify the concepts and axiomatic content of a particular domain[35]. Therefore SUMO acts as a foundation for more specific domain ontologies.

Due to broad nature of SUMO, the goal to construct a single, consistent and comprehensive ontology isn't easy or perhaps not achievable. In this case the best approximation is to make clear the representational choices and compartmentalize them in consistent and independent packages and then where possible state mappings between corresponding packages such as in LOD[49].

LOD datasets are interlinked well on the instance level, but they are very loosely connected on the schema level[35]. Since upper level ontology captures diverse domains at a fairly abstract level, integrating LOD with it results in better understanding. Figure 10 represents

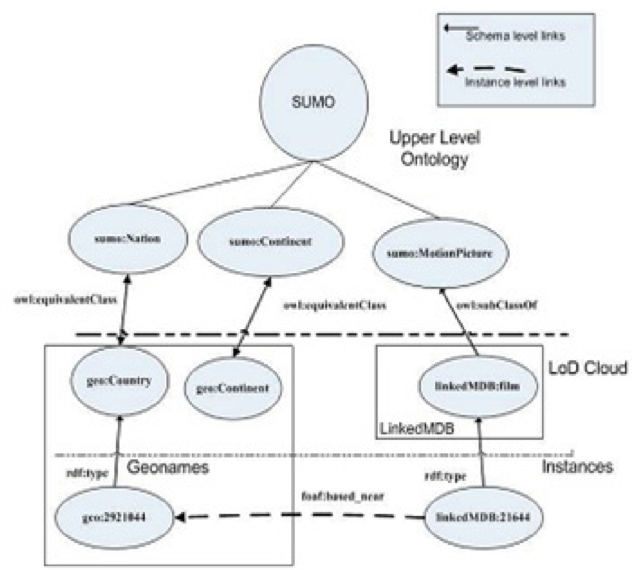

Figure 10. LOD integration with SUMO[36].

the idea. However integrations need to act as a smooth bridge between abstraction of upper level ontology and instantiations present in LOD cloud. Another effort in this direction is to utilize another well known upper level ontology called Cyc for providing structural backbone to LOD cloud via UMBEL. UMBEL contains schema level links to 21 different LOD datasets, and is a much needed step in this direction.

2.7 LOD Datasets

According to statistics currently the LOD cloud contains almost more than two hundred and ninety datasets. Datasets belongs to different categories. Which include datasets of media, publications, government and many others. The categories are indicated by different colours in Figure 1(b). Among all datasets the most linked and utilized dataset is DBPedia.

2.7.1 DBPedia Dataset

DBpedia serves as a nucleus of LOD cloud which has links to other several location related datasets such as GeoName, Flicker, Revyu, US Census, CIA Factbook, YAGO, FOAF, and EuroStatetc. DBpedia dataset has been extracted from Wikipedia. DBpedia 3.7 data set contains a vast collection of things, links and images etc. The contents of dataset includes more than“*3.64 million things, of which 1.83 million are classified in a consistent ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 video games, 169,000 organizations, 183,000 species and 5,400 diseases. The DBpedia data set features labels and abstracts for 3.64 million things in up to 97 different languages; 2,724,000 links to images and 6,300,000 links to external web pages; 6,200,000 external links into other RDF datasets, and 740,000 Wikipedia categories*”.

The dataset is composed of 1 billion RDF triples out of which 385 million were retrieved from the English edition of Wikipedia and approximately 665 million were extracted from other language editions and have links to external datasets[81].

In order to facilitate DBpedia users to find out more related information, the DBpedia knowledge base is interlinked with a number of other data sources. The knowledge base contains approximately 4.9 million outgoing RDF links pointing at corresponding information about DBpedia entities, as well as meta-information about media items representing an entity[14].

Besides the datasets, DBPedia provides several services including *Lookup Service*(Lookup index and OPenLink,Virtusoso), *Query Builders*(for which sparkle endpoints are provided including OPenLinkiSPARQL, SNORQL, Demo Query Virtuso), DBPedia *spotlight services*(which are not only developed for DBPedia but are utilized in many researches).

Rest of the LOD datasets along with their categorization aregiven in Table 3. The outline of most of the datasets along with their size(in terms of number of triples it contains) and URLI’s is given[76].

3. LOD Tools and Applications

Broadly the tools and applications developed for LD and LOD can be classified as either publishing (tools and applications) or consuming (tools and application). Publishing tools and applications contribute to the LOD cloud growth and consuming tools and applications aim

Table 3. Datasets in LOD along with their Categories

Category	Datasets
Registry Service	CKAN[75]
Bibliography	DBLP , Lipris, Open Library Project[66, 51]
Geographical	Geonames
Entity Information	FOAF, YAGO, DBPedia, Revyu.com, Loticoare
Statistical	Reiese
Referenced Based	UMBEL
Entertainment	LinkedMDB
BBC	BBC music, BBC Programs
Media	Event Media, EUROPEANA
Research	OpenPSI
Library	VIAF(Virtual International Authority File) ,LCSH(Library of Congress Subject Headings), DDC(Dewey Decimal Classification)[73,41,21]
Intelligence	Crime Reports
News Reports	London Gazette, Ontos News Portal, New York Times
Maps, Terrain Information and Weather	Ordnance Survey Linked Data, EI Viajero’s tourism, GeoNames, LinkedGeoData, Yahoo Geoplanetare
Sensor Linked Sensor Data	Linked Sensor Data ,Linked observation data[53,52]
Cultural Heritage	Europeana[18]

to consume and utilize the LOD datasets. Table 4 and Table 5 give overview of tools and applications related to publishing and consuming respectively.

3.1 LOD Publishing Tools and Applications

- ….From Relational To LOD-With the growth of the semantic web, the existing standard of relational data for data storage becomes obsolete[29]. Obtaining such data in a new, satisfactory format is impracticable using manual method. It is therefore necessary to create automated processes that can translate relational data into LD.D2R Server is a tool for publishing relational databases as LD. The OpenLink Virtuoso server82 facilitates serving RDF data via LD interface and a SPARQL endpoint[12]. RDF data can be either stored directly in Virtuoso or can be created based on mapping from non-RDF relational databases. Google Refine[83] is a powerful tool utilized for messy data. It cleans the data, converts from one format into another and extends it with web services.

In addition to the tools that covert relational data to LOD, tools are there to convert from other formats to LOD. ConverterToRdf[82] and RDFizers[83] convert information currently represented in formats such as CSV, Microsoft Excel, or BibTEX as LD. SparqPlug[17] is a service that enables the extraction of LD from legacy HTML documents on the Web that do not contain RDF data. After conversion to RDF, stores data in RDF repository.

- ….RDF Repositories-A list of RDF repositories to store RDF triples or LDis maintained in the ESW Wiki[77]. Example includes Talis Platform, 4store, D2RQ and D2R Server, Dojo Data, Franz Inc'sAllegroGraph, Intellidimension's RDF Gateway, Kowari, Puby etc.
- ….Linked Data Management System-In contrast to content management system like Joomla, drupal and wordpress which are designed to handle mostly the

Table 4. Major LOD Publishing Tools and Applications

Categories	Tools
Publish Linked Open Data from relational databases[5,75]	D2R Server, Triplify, Virtuoso Universal Server , RDBToOnto, R2O, D2RQ, B2OWL,web2py
Publish Link Open Data From Other formats like CVS, Excel[11]	ConverterToRdf and RDFizers, SparqPlug
RDF Repositories[74]	4store[94] ,D2RQ and D2R Server[95], Dojo Data, Franz Inc's AllegroGraph[96], Intellidimension's RDF Gateway , Kowari , Mulgara , OpenLink Virtuoso , Oracle Spatial 11g , OWLIM , Pubby , RDFStore , SemWeb for .NET ,Sesame , SDB, SWI-Prolog Semantic Web Server ,TalisPlatform, Tucana Suite, YARS, 3Store, bigdata, djubby
Linked Open Data Management System	Callimachus, Paget
Linked Open Data Wrappers	OAI2LOD Server ,SIOC Exporters[97]
RDF Editor/RDF Validator	Hyena,Graphl,Vapour: Linked Data Validator W3C's RDF Validator

Table 5. Major LOD Consuming Tools and Applications

LOD Browser	Tabulator Browser (MIT, USA),Marbles (FU Berlin, DE) OpenLink RDF Browser (OpenLink, UK), Zitgist RDF Browser (Zitgist, USA), Humboldt (HP Labs, UK),Disco Hyperdata Browser (FU Berlin, DE), Fenfire (DERI, Irland), Longwell (faceted browser)
LOD Visualization Tools	Exhibit,Timeline, Many eyes,Open platform for visualization
LOD Search Engines	Falcons, Sindice,Swoogle , SWSE
Client Libraries/ Datasets	CKAN ,DBpedia, DBLP Bibliography, GeoNames,Revyu, riese,UMBEL, Sensorpedia, FOAF,OpenPSI,VIAF

unstructured text, there are LD management systems. Example includes Callimachus[84], Mulgara or OWLIM, AliBaba and Paget[85].

- ….Linked Data wrappers-many institutions allow access to their metadata repositories via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). However, this protocol has two considerable drawbacks; first it does not make its resources accessible via dereferencable URIs, and secondly it provides only controlled and choosy access to metadata[31]. The OAI2LOD[86] Server handles these drawbacks by republishing metadata originating from an OAI-PMH endpoint according to the principles of LD.
- ….Linked Data/RDF Editor and RDF Validator-There are number of tools that are utilized by the users for collaborative editing and visualization of RDF graphs[74]. Example include Graphl (a tool for collaborative editing and visualization of RDF graphs) and Vapour (aLD Validator) etc.

3.2 LOD Consuming Tools and Applications

3.2.1 LOD Browsers

RDF links enable users to navigate between different datasets. Following this direction *Linked data browsers* are developed for exploring or navigation, analyzing and visualizing LOD. With the growing popularity of LOD, discovering in this large information space gives many challenges.

3.2.1.1 Challenges for LOD Browsers

The utilization of Web of data mainly depends on usability of browser interfaces for different users of the Web. Shneiderman et al.[60] classified users into *Lay user*, *Technical user* and *Domain User*. The classification is based on level of understanding of the technology and domain expertise. Summary of the challenges identified by Alahmari, Thom, Magee, & Wong[1] for LOD browsers that caters the needs of all type of users are explained as under.

- ….Linked Data Exploration- As the Web of data connects huge data from the real world domains and other Web resources, two questions arise. First how a browser will present this huge range of data to the user in a well presentable and highly interactive format. Secondly how it will handle resource linkage between real world domain dataset and general web data?
- ….Navigation- Navigation technique is quite different in LOD as compared to the traditional web browsing. Traditional browser use un-typed links to surf between different pages, while LOD browser use typed links to navigate between different RDF resources in the form URIs. The major challenge is that how browser will respond in allowing forward and backward navigation techniques to the users along with the support of context navigation.
- ….Interactivity- The utilization of LOD cloud mainly lies in its interactive feature of sub selection and faceted browsing. However achieving high user interaction features is hardly difficult due to use of different languages of the semantic web (RDF, OWL and SPAQL). Considerable research is devoted to develop such browsers which can facilitate users in finding information in huge sets of LOD and to reduce cognitive load on the users.

3.2.1.2 Key Indicators for Evaluation of LOD Browsers

Analysis of major LOD browsers is given in Table 6. The evaluation indicators adopted by Nikolov, Uren, Motta, Roeck[48] and Niles I, Pease A[49] are used to compare different LOD browsers. It include data conversion, 5-star schema, data overview, presentation, detail on demand, scalability support, querying, filtering, presentation templates, entry point, non domain specific, publication, edit underlying data, data reuse, navigation and plug in for HTTP browsers.

Disco, Marble, ODE are mainly text oriented browsers yielded powerful results. However Fenfire, Tabulator browsers' power lies in their visualization, data manipulation and in additional feature of plug in availability with traditional web browsers (like FireFox etc). Sparrax and Factes are faceted browsers. They provide various options of sub selection. Sparrax user make use of SPAQRL end point based Freebase Parallax. However issue of slow response can't be neglected in case of huge datasets.

3.2.2 Search Engines

LOD/RDF centric search engines crawls the web of LD. The name used for these search engines is semantic search engines. The most popular search engines includes Falcons, SWSE(Semantic Web Search Engine), Sindice, Swoogle, and Watson.

Table 6. Comparison of LOD Browsers

Indicator	Dipper[23]	Disco[24]	Marble[46]	Sig.ma[69]	URI burner[70]	Fenfire[32]	Tabulator[9]	ODE[50]	Sparrax[44]	Factes[44]
Visual presentation	-	-	-	-	-	X	X	X	-	-
RDF graph view	-	-	-	-	-	X	-	X	-	-
Visual overview	-	-	-	-	-	-	-	-	-	-
Detail on demand	-	-	-	-	-	X	X	X	-	-
Highlight links in data	-	-	X	-	-	X	X	X	-	-
Support for scalability	-	-	-	X	-	X	-	X	-	-
Query (formal syntax)	-	-	X	-	-	-	X	X	-	-
Query (forms / keyword)	-	-	-	-	X	-	X	X	-	-
Filtering	-	-	-	X	-	X	-	X	-	-
History	-	-	-	X	-	-	X	X	-	-
Presentation Templates	X	-	X	X	-	-	-	X	-	-
Keyword / entry point				X	X	-	-	X		
Non-domain specific		X	X	X	X	X	X	X		
Faceted search / browse	-	-	-	-	-	-	-	-	X	X
Edit underlying data	-	X	-	-	-	X	X	-		
Reusable output	X	-	-	X	X	X	-	-		
Target – Lay-users	-	-	-	-	-	X	-	X		
Target – Tech-users	X	X	X	X	X	X	X	X		
Produce 5 star	-	-	-	5	5	-	5	5		
Consume 5 star	5	4	5	5	5	4	5	5	5	5
Plugin available	-	-	-	-	-	-	X	X	-	-
Facets view	-	-	-	X	-	-	-	X	X	X
Forward navigation	-	X	X	-	X	-	X	X	-	-
Backward navigation	-	X	X	-	X	-	X	X	-	-
Export RDF / JSON	X	-	-	X	X	-	-	X	-	-
Navigating Global Linked Data	-	-	X	X	X	X	C	X	-	-
Navigating Local Linked Data	X	-	X	X	X	X	X	X	-	-

Yahoo and Googlesearch engines also started to make use of structured data on the web to enhance the search results. Yahoo data can be accessed through BOSS API. It also uses data within SearchMonkey[87] to make search results more precise. Google uses crawled RDF data in order to enhance search results snippets for products, reviews, and people (i.e. entities)[12]. But still not successful and put forward the need to develop new search engines.

3.2.2.1 Evaluating LOD/RDF Centric Search Engines

Different LOD/RDF centric search engines have different architecture, but the purpose is the same i.e. search for RDF related documents. Table 7 is a feature wise comparison of the major LOD/RDF centric search engines. Some of the semantic search engines provide human oriented interface, while others provide services for semantic web applications. The semantic search engines normally crawls and provides results in the form of RDF type of documents[22].

LOD/RDF centric search engines are evaluated according to the evaluation indicators adopted form[22] are given in Table 8. The indicators include ontologies, types of document treated, clustering of results and number of triples (statistics is taken directly from source web site).

3.2.2.2 Natural Language Interfaces for Semantic Search Engines

The Natural Language Interfaces (NLIs) provide support for natural language queries to facilitate end users that lack knowledge of OWL, RDF or SPARQL. For example Kaufmann and Bernstein[37] discussed four types of query interfaces i.e. Ginseng, NLP-Reduce, Querix, and Semantic Crystal (for the semantic web). These NLIs converts natural language query to some formal query structure like SPARQL. The natural language query may be based on keyword and some complex English sentence. After parsing natural language query is converted to SPARQL queries, which is then executed using Jena. Table 9 is derived from Kaufmann and Bernstein[37], shows three different natural language interfaces for the semantic search engines along with their features.

Table 7. Feature Comparison LOD/RDF Centric Search Engines

Name	Design	Return type/ Results	Crawling/Type of documents	Remarks
SWSE	Human oriented	Links + exploits underlying structure of the data + summary of the entity + summary of the entity select from the list	RDF/XML + Normal HTML + RSS feeds converted into RDF/ XML.	SWSE allows keyword queries and responds with a ranked list of result snippets. The results refer to entities not a document.
Falcons	Human oriented	Links + exploits underlying structure of the data + summary of the entity + summary of the entity select from the list	RDF/XML	About 30 million Semantic Web entities have been Indexed
Sindice	Application oriented	RDF /XML/JSON/HTML/ Plain Text	RDF	Continuous crawling but lake of full query support.
Swoogle	Application oriented	Ontologies (OWL + RDF files).	Ontologies on web(RDF) + (OWL)+ RDF embedded in HTML	Swoogle currently has indexed nearly 1.3M Semantic Web documents which contain almost 240M triples. Swoogle uses page rank as developed by Google.
Watson	Application oriented	RDF/XML	RDF documents	Watson support keyword search for ontologies + SPARQL queries.

Table 8. Evaluating LOD/RDF Centric Search Engines

Names	Ontologies	Types of document Treated	Clustering of Results	Number of triples
SWSE	Implicit through RDF	RDF related documents	Yes	Indexing a crawl of 1 billion facts (May 2010) + 1 billion inferred facts
Falcons	Implicit through RDF	RDF related documents	Yes	N/A
Sindice	Implicit through RDF	RDF related documents	N/A	708.26 Million documents
Swoogle	Implicit through RDF	RDF related documents	N/A	1,142,290,057
Watson	Implicit through RDF	RDF related documents	N/A	N/A

Table 9. Natural language interfaces for semantic search engines

NLIs (Natural Language Interfaces)	Support	Underlying technologies
NLP-Reduce	Keyword + Full sentence	SPARQL + Jena + Pellet Reasoner + WordNet + OWL-based knowledge base.
Qurix	Full sentence query	SPARQL + Jena + WordNet + OWL -based knowledge base
Ginseng	Full sentence query	SPARQL + Jena + OWL-based knowledge base

3.2.2.3 *Searching and Retrieving in LOD: Challenges and Opportunities*

LOD/RDF centric search engines still have the problem of low precision and high recall. Therefore it is required to increase its precision and lower its recall.

Another problem is entering *wrong queries or domain specific queries*. As a user has knowledge of specific area and he may enter the queries according to his knowledge. That's why accurate query is also a problem in both syntactic and semantic search engines[45].

Semantic computing[16] recognized and put forward challenges in answering natural language queries on LOD cloud that includes mapping natural language expressions to RDF vocabulary, handling complex categories, aggregation function and comparisons, and temporal reasoning.

4. Domain Specific Applications and Projects

In this section we will analyse major projects based on LOD in the information space of Mobile, Multimedia, File system, Intelligence, Sensors and Library along with challenges. Table 10 provides listing of these projects.

- ….DBPedia Mobile- At present semantic web has grown tremendously and holds a massive amount of location related information. DBpedia mobile is a location-aware client and linked data browser, that is designed to run on a variety of platforms including iPhone[88] and Android[89]. DBpedia mobile allows users to search, and post information related to their physical vicinity using their mobile phones and standard web browsers. DBpedia mobile uses information from DBpedia and enable users particularly tourist exploring cities[3].

Table 10. LOD based domain specific projects

Domain	Project Names
Mobile Based	DBPedia Mobile
Intelligence	N-DEx, Army Knowledge Online ,Data Cloud for Afghan War, Cazoodle
File System	TripFS,LODFS
Multimedia	LUCERO, KMI,NoTube, Yumma, EUROPEANA, Synote,SEmTube
Sensors	SEmSOS, SensorMesher
Library	Fast(Faceted Application of Subject Terminology), Amsterdam Museum LOD

DBpedia mobile is not only restricted to access DBpedia dataset but can also access any dataset that will interlink to DBpedia or in other words any dataset that is reachable from DBpedia. DBpedia mobile application is composed of map view and Fresenl based linked data browser. Standard phone web browser is used to access DBpedia mobile, where supplementary launcher applications are used to initialize DBpedia mobile using the user's current locations (that can be retrieved from built-in or externally connected GPS receivers)[6,7]. Two views that are summary view and photo view provides summary and detail view of selection respectively. DBpedia mobile also enable users to publish information related to current geographic location (such as GPS coordinates, photos, and reviews) using DBpedia or other interlinked data sources in the semantic web by its *Content Creation Panel* feature. DBpedia mobile can be viewed as a client-server application, where queries, storage, data retrieval along with other tasks are performed over server-side Marbles engine.

- ….File System has Gone Linked-A very limited work has been carried out to link file system with other file systems and LD or enabling desktop applications to access LOD sets. File system has gone linked and linked data towards file system, describe two different aspects of file system and LD. The former describes linking of file to other information objects and the later describes the representation of LD as virtual file system.

- ….TripFS[58,59] is an attempt by Bernhard Schandl to link file system contents to LD. The TripFS service is implemented in Java using Jena semantic web framework.
- ….Bernhard Schandl[69] presents LD as a virtual file system to bridge two information spaces. This enables desktop applications to read, browse and navigate linked datasets as if they were present in file system.

Despite researchers' contributions the area still need further research to ensure the privacy and security of user's information, reliability and consistency of links, and links within files. Heuristics are needed to establish links to paragraphs within a document file, to regions in a picture file, and to temporal duration of a video file.

- ….LOD in Intelligence-Many new intelligence gathering and decision support systems are now using the concept of LD. LD is till now answering all the questions. For example U.S. government issued an executive order[64] which direct the government departments to share their information with each other in order to avoid future 9/11 types of attacks. The LD is used here for data sharing[80]. US army even gave a task to a private firm to build a private army cloud for the soldiers, stationed in Afghanistan[10].
  - ….Cazoodle- Project was started for apartment hunting few years back. This project was started by few graduate students of university of Illinois Urbana Champaign whose interest was to find a cheap apartment. However the U.S. Army started to use it in a different way. According to the CTO of Cazoodle "GovindKabra" the Cazoodle is using publicly available different data sources for example "Flickr" and "OpenStreetMap" datasets to build a detailed map and guidebook of Afghanistan for the US soldiers[72]. The goal of the project is to describe in detail about the towns and cities which includes everything about a place like names, locations (with list of coordinates of schools, mosques, banks hotels and other landmarks) and population etc. Cazoodle uses openly available LD for creating detailed map.
  - ….Data Cloud for Afghan War- During April 2009 U.S. army started a program. A big data cloud for the troops deployed in Afghanistan to provide the latest intelligence information to the troops. The project of private cloud was started after convincing U.S. army with the versatility of the linked data that the cloud can easily fulfil their requirements and solves many issues related to the intelligence analysis, processing and production. Now a days the size of Afghanistan private data cloud reaches to petabytes of the data. In order to find the data in the cloud, predictive models are used[10]. Army Knowledge Online[19], N-DEx (National Data Exchange" of Law Enforcement)[28] are the other popular projects in the area of intelligence based on LOD.

To end with LD has a bright future in the intelligence gathering tools however there are some challenges which must be addressed like the privacy issue, data fusion[13,38,40,48], metadata mapping.

- ….LOD in Multimedia-Now a days multimedia data is becoming the primary content of the web and a lot of video sharing web applications have been developed such as youtube[90] and myspace[91] etc. These applications allow users to easily share, search, bookmark and attach annotations to multimedia objects such as audios, videos and images. But browsing specific regions or fragments of the multimedia object is still a difficult task. For this purpose a lot of web applications have been developed such as SemTube Video Annotation Tool, SyNote, YUMA Media Annotation Framework, KMI Annotation Tool[92], LUCERO, EUROPEANA Connect[93] and NoTube etc. The purpose of these tools is to interlink the annotation of multimedia resources across different repositories to achieve better indexing and searching. Analysis of these tools based on features identified in Table 11 is given in Table 12.
- ….LOD in Sensors- Using linked sensor data is relatively a novel idea, having great potential but not attracted the attention of application developers due to some of its complexities. Anyhow, researchers have developed some prototypes to leverage the effectiveness of linked sensor data and encourage other developers to join the area.
- …..Joshua Pschorr et al.[55] has presented an idea of semantic sensor network middleware, leveraging the power of both semantic web and existing datasets found

Table 11. Features of Video Annotation Tools/Projects

Annotation depiction	(1) HTTP-derefrenceble RDF document, (2) Linked Data, (3) Linked Open Data, (4) embedded in content representation
Annotation target object type	(5) web documents, (6) multimedia objects, (7) multimedia and web documents
Vocabularies used	(8) RDF/RDFS, (9) Media fragment URI, (10) OAC(Open annotation Collaborative), (11) Open Archives Initiative Object reuse and Exchange (OAI-ORE), (12) Schema.org, (13) LEMO, (14) FOAF(friend of A friend), (15) Dublin Core, (16)Timeline, (17) SKOS(simple knowledge organization system), (18) W3C Media ontology, (19) Bibliography ontology(Bibbo), (20) Course and AIISo Ontology, (21) Creative commons Rights, (22) Expression Vocabulary and Nice Tag Ontology, (23) Sioc Ontology, (24) WP1, (25) WP2, (26) WP3, (27) WP4, (28) WP5, (29) WP6, (30) WP7a, (31) WP7b, (32) WP7c
Flexibility	(33) Yes, (34) No
Annotation type	(35) Text, (36) Drawing tools, (37) public, (38) private
Definition languages	(39) RDF/RDFS, (40) OWL
Media fragment identification	(41) Xpointer, (42) Media fragment URI 1.0, (43) MPEG-7 fragment URI, (44) MPEG-21 fragment URI, (45) N/A

Table 12. Feature analysis of Video Annotation Tools/Projects

Features		Annotation Depiction	Annotation target object	Vocabularies	Flexibility	Annotation Type	Definition Languages	Media Fragment URI
Projects & Tools	EUROPEANA Connect	3	7	10,13	33	35,36,37,38	39	41,42
	SemTube	2	6	8,10	33	35,36	39	41
	YUMA	3	7	10,13	33	35,36,37,38	39	41,42
	KMI	3	6	14,15,16,17	33	35	39,40	41,42
	LUCERO	3	7	14,17,18,19, 20,21,22,23	33	35,36	39	41,42
	NoTube	3	7	24,25,26,27, 28,29,30, 31,32	33	35,37,38	39	41

on the LOD (e.g. GeoNames and LinkedGeoData etc) for effective discovery of sensors on the web using named-locations. The author claims that to use rich, and location-based semantics for sensor discovery, sensors descriptions and observation needs to be annotated with useful metadata. The proposed idea extends existing SWE framework by integrating semantic web technologies and constructing a Semantic Sensor Observations Service (SemSOS). SemSOS is a set of methods having potential of accessing ontological knowledgebase to support queries with high-level features such as named-locations etc. Harshal Patni et al.[52] has assumed that linking Linked Sensor Data dataset to geographical names provided by GeoNames dataset can be advantageous for answering sensor discovery queries using named-locations. An application with simple map-based GUI has been built for finding nearby sensors using named-location. A user is only required to enter location name in the text box, the application will automatically build and execute SPARQL query over the Linked Sensor Data dataset on LOD and renders all of the sensors available nearby the given location on a map.

- ….SensorMasher - Danh Le-Phuoc et al.[54], has presented a system called “SensorMasher” which integrates sensors data available on LOD into mashups. SensorMasher composes and enables non-technology oriented users to derive new sensors data sources by fusing existing sensors data from multiple sources. Payam Bernaghi et al.[4] has also demonstrated a mashup application using Google Maps API to show the LOD data usage and integration from multiple sources. A user has to only provide location attribute of a resource. The application then extracts geographical coordinates as well as other related attributes of the resource from the LOD and show available sensors along with their properties through Google Map application.

- ….LOD in Library- The current digital libraries' standards and practices are only limited to the library domain established by an Integrated Library System industry. LOD principles and best practices will motivate the organizations to strengthen preservation, linking of resources with their description and publishing of data. The LOD is the first step for openly linking different cultural heritage resources and their description, so the visibility of the organizations will be increased.

By using LOD practices, tools and applications, different vendors and developers can have freedom of selecting non-library standard practices for publishing, dissemination, sharing, reusing the rich library data and to be available with the global scope[74].

- ….FAST (Faceted Application of Subject Terminology) - is the application of the Library of Congress subject headings dataset (LCSH), for making the use, application and understanding more easily. Because the Library of Congress value vocabularies are complex structured and syntactically much difficult to understand and use. For this purpose the LCSH subject headings' dataset is used in the FAST project. FAST make the LCSH rich vocabulary and schema very simple and easy for indexing the terms and subject headings[27]. It is the controlled vocabulary driven from the LCSH subject headings for the names, subjects, events, chronology, places and genre of the subject or any other resource[47].
- ….SwetoDblp- ontology of computer science publications is the shallow web ontology application for processing large amount of data for discovering and analysing the resources based on the Bibliography of DBLP[2].

Amsterdam museum LOD(using the OAI model's ORE and OAI-PMH protocols and dataset), Europeana Data Model (EDM), Dublin Core dataset (for aggregation of different heritages and cultural records from variety of sources)[15], The 20th Century Press Archives of the German National Library of Economics (ZBW) are the some of the other important examples.

In conclusion it is obvious that every area is strongly influenced by the concept of LOD that results in new dimension of exploration in every domain.

5. Conclusion and Future Directions

In this paper review of LOD from beginning to state of art is presented. The need, motivational factors, development tools and applications are discussed along with issues and challenges. Furthermore the ongoing projects in the information space of Library, Sensors, Multimedia, File system and intelligence are reviewed. Given below are the key challenges in LOD framework that need to be addressed via further research in this domain.

Despite of researchers' contributions, the areas and trends identified in this section are still not mature enough and needs further research. One of the major challenges in the evolution of LOD is its limitedness in multilingualism. The web of data shows the possibility for being extended to a truly multilingual web. Vocabularies and data are available in a language-independent fashion. And related language-dependent (linguistic) information (supporting the access across languages) can be stored separately[30].There is a strong separation among, on the one hand, terminology, lexical and language resources, and on the other, the technologies used for linked data. This gap makes it hard to exploit any potential synergy among approaches. This also makes it difficult to take advantage of the opportunities offered for linked data by multilingual web technologies. Removing silos and integrating these technologies is therefore a significant goal[78]. In this sense, the multilingual web of data can be realized in our view as a layer of services and resources on top of the existing linked data infrastructure adding i) linguistic information for data and vocabularies in different languages, ii) mappings between data with labels in different languages, and iii) services to dynamically retrieve and traverse LD across different languages[30]. Recent initiative taken by W3C in this direction is the Multilingual Web-LT project[74]. Multilingual Web project focused on the intersection between LOD and multilingual technologies.

Establishing benchmarks for LD processing with cloud computing offerings is another challenge in future. Processing huge volumes of LD needs sophisticated methods and tools. In the recent years main focus was on systems based on relational databases and custom-built systems for LD processing. Cloud computing assistance such as SimpleDB or BigQuery, and cloud-enabled NoSQL systems (including Cassandra or CouchDB) as well as frameworks such as Hadoop offer attractive alternatives along with great promises concerning performance, scalability and elasticity[33].

Future browsers should be made highly interactive with faceted browsing to give better response time. They must be able to face the challenge of data discovery that is fetching data from diverse datasets (across structure and unstructured datasets).

LOD or RDF centric search system future dimensions consists of five areas, also identified[45]. These five areas are input style, expressiveness, query execution, response time and result presentations. Input style area should provide dual query formulation. One is for view-based interfaces which expose the structure of the ontology in graphical shape, and second, natural language interfaces (because of its easiness and quick input). The user should be allowed to input complex queries consisting of logical operators like 'AND', 'OR'. For example "What are the rivers that pass through Arizona and pass through California? It should be noted that the response time of the semantic search systems are slow as compared to key-word search engines. Some semantic search systems provide intermediate, or partially complete results, which although provides quick responses, but continuously updating user screen creates confusion. So the delay time must be overcome and should be closer to utilized time of keyword search engines. The results presentation of the semantic search engines must be presented in an accessible and attractive manner. Results management such as sorting, filtering, provenance and trustworthiness of results must be established.

In real applications, having full information about a query object may not be practical so it may not be achievable to mention exact query criteria. For example, we may know that a famous politician was born on February 12 and died on April 15, but we have no idea about his precise birth and death years. In this scenario, we have to carry out a query with wildcards. Secondly in some applications, RDF repositories are not static. For example, Yago and DBpedia datasets are continually growing to take in the newly extracted knowledge from Wikipedia. Future work can be done on developing techniques for supporting SPARQL queries with wild cards and for managing large constantly changing RDF datasets in a scalable manner. Existing RDF storage systems, such as Jena, Yars2 and Sesame 2.0, cannot work well in large RDF datasets (such as Yago dataset)[25]. SW-store, RDF-3x, x-RDF-3x and Hexastore are designed to address scalability, however, they can only support exact SPARQL queries, since they change all literals (in RDF triples) by ids using a mapping dictionary.

6. References

1. Alahmari F, Thom J, Magee L, Wong W. Evaluating semantic browsers for consuming linked data. Proceedings of the Twenty-Third Australasian Database Conference; 2012. p. 89–98.
2. Aleman-Meza B, Hakimpour F. SwetoDblp ontology of Computer Science publications. Web Semantics: Science. 2007;5(3):151–155.
3. Auer S, Bizer C, Kobilarov G, Lehmann J. Dbpedia: A nucleus for a web of open data. The semantic web; 2007. Available from: http://link.springer.com/chapter/10.1007/978-3-540-76298-0_52
4. Barnaghi P. Publishing linked sensor data. 3rd International Workshop on Semantic Sensor Networks (SSN), Organised in conjunction with the International Semantic Web Conference (ISWC); 2010.
5. Baron C, Pierro M. Publishing Linked Data using web2py; 2010. Available from: http://web2py.com/semantic/static/semantic.pdf
6. Becker C, Bizer C. DBpedia mobile-a location-aware semantic web client. Proceedings of the Semantic Web Challenge; 2008. p. 1–8.
7. Becker C, Bizer C. DBpedia Mobile: A location-enabled linked data browser. Workshop on Linked Data on the Web (LDOW); 2008. p. 6–7.
8. Berners-Lee T. Linked data–design issues. Available from: http://www.w3.org/DesignIssues/LinkedData.html
9. Berners-Lee T, Chen Y, Chilton L, Connolly D, Dhanaraj R, Hollenbach J, Sheets D. Tabulator: Exploring and analyzing linked data on the semantic web. Proceedings of the 3rd International Semantic Web User Interaction; 2006.
10. Big Data Cloud Delivers Military Intelligence to U.S. Army in Afghanistan. Available from: http://www.datanami.com/datanami/2012-02-06/big_data_cloud_delivers_military_intelligence_to_u.s._army_in_afghanistan.html?featured=top

11. Bizer, Chris, Cyganiak R, Heath T. How to publish Linked Data on the Web; 2007. Available from: http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/
12. Bizer, Christian. The emerging web of linked data. Intelligent Systems, IEEE. 2009; 25(5):87–92.
13. Bizer, Christian, Heath T, Berners-Lee T. Linked data-the story so far. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems; 2009. Available from: http://www.igi-global.com/article/linked-data-story-far/37496
14. Bizer, Christian, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S. DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web. 2009; 7(3):154–65.
15. De BV, Wielemaker J, Gent V, Oosterbroek M. Amsterdam Museum Linked Open Data, 0(0), 1–5. Available from: http://www.semantic-web-journal.net/sites/default/files/swj293.pdf
16. Citec CU. BCogS 2011. Available from: http://www.cit-ec.de/sites/www.cit-ec.de/files/slides_cunger.pdf
17. Coetzee P, Heath T, Motta E. SparqPlug: Generating linked data from legacy HTML, SPARQL and the DOM. Linked Data on the Web (LDOW); 2008. Available from: http://ra.ethz.ch/CDstore/www2008/events.linkeddata.org/ldow2008/papers/05-coetzee-heath-sparqplug.pdf
18. Cultural Heritage Dataset. Available from: http://datahub.io/dataset/europeana-lod
19. DARPA - Army Knowledge Online Project. Available from: http://dig.csail.mit.edu/2010/DARPA-AKO/
20. Data L. Towards a basic profile for linked data. 2011; 1–15. Available from: http://open-services.net/pub/Main/TowardsABasicProfileForLinkedData/Towards_a_Basic_Profile_for_Linked_Data_for_LEDP_Workshop.pdf
21. Dewey Decimal Classification (DDC) Datset. Available from: http://datahub.io/dataset/dewey_decimal_classification
22. Dietze H, Schroeder M. GoWeb: a semantic search engine for the life science web. BMC bioinformatics. 2009; 10 Suppl 1:S7.
23. Dipper Browser-Talis Platform. Available from: http://notes.3kbo.com/talis
24. Disco Hyperdata Browser. Available from: http://wifo5-03.informatik.uni-mannheim.de/bizer/ng4j/disco/
25. Elbedweihy K. Evaluating semantic search systems to identify future directions of research. Proceedings of the Second International Workshop on Evaluation of Semantic Technologies (IWEST). 2012. p. 25–36.
26. Empirical Cloud. Available from: http://inkdroid.org/empirical-cloud/
27. FAST (Faceted Application of Subject Terminology). Available from: http://www.oclc.org/research/activities/fast.html?urlm=159754
28. FBI — N-DEx Overview. Available from: http://www.fbi.gov/about-us/cjis/n-dex/ndex_overview
29. Froment C. Translating relational databases into Linked Open Data; 2012.
30. Gracia J, Montiel-ponsoda E, Cimiano P, Gómez-pérez A, Buitelaar P. Challenges for the Multilingual Web of Data.
31. Haslhofer B, Schandl B. The OAI2LOD server: exposing OAI-PMH metadata as linked data. LODWS. 2008. Available from: http://eprints.cs.univie.ac.at/284
32. Hastrup T, Cyganiak R, Bojars U. Browsing linked data with Fenfire. Linked Data on the Web Workshop. 2008. Available from: http://ir.library.nuigalway.ie/xmlui/handle/10379/537
33. Hausenblas M, Grossman R, Harth A, Cudre-Mauroux P. Large-scale linked data processing: cloud computing to the rescue? Available from: http://webofdata.files.wordpress.com/2012/03/closer12-processing-lod.pdf
34. Introducing linked data and the semantic web. Available from: http://www.linkeddatatools.com/semantic-web-basics
35. Jain P, Hitzler P, Sheth A, Verma K, Yeh P. Ontology alignment for linked open data. Ontology Alignment for Linked Open Data. ISWC 2010, Part I, LNCS 6496. Springer-Verlag Berlin Heidelberg. p. 402–17.
36. Jain P, Hitzler P, Yeh P, Verma K, Sheth A. Linked data is merely more data. Linked data meets artificial intelligence. [Technical Report]. 2010. p. 82–86. Available from: http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1130/1454
37. Kaufmann E, Bernstein A. How useful are natural language interfaces to the semantic web for casual end-users? Proceedings of the 6th international semantic web conference and 2nd Asian conference on Asian semantic web Springer-Verlag; 2007. p. 281–94).
38. Lambert DA. An exegesis of data fusion. Soft Computing in Measurement and Information Acquisition. 2003; 127.
39. LarKC. Lumpy structure of LOD. Available from: http://blog.larkc.eu/?p=1941
40. Le-Phuoc D, Polleres A, Hauswirth M, Tummarello G, Morbidoni G. Rapid prototyping of semantic mash-ups through semantic web pipes. Proceedings of the 18th international conference on World wide web; 2009. p. 581–90).
41. Library of Congress Subject Headings (LCSH) Dataset. Available from: http://datahub.io/dataset/lcsh
42. Linked Open Data Graph. Available from: http://inkdroid.org/lod-graph/

43. LOD cloud analysed with gephi. Available from: http://twitpic.com/17qj1h/full
44. Logan A. Exploiting UK government linked data; 2010. Available from: http://homepages.abdn.ac.uk/advaith/pages/teaching/abdn.only/CS5915/information/CS5577-Logan A J.pdf
45. Madhu G, Govardhan A, Rajinikanth TKV. Intelligent semantic web search engines: a brief survey. International Journal of Web & Semantic Technology. 2011; 2(1):34–42.
46. Marble. Available from: http://www5.wiwiss.fu-berlin.de/marbles
47. Neill ETO. Description of the FAST (Faceted Application of Subject Terminology) Dataset. Semantic Web Journal: Special Call for Linked Dataset Descriptions. 2012; 1–7. Available from: http://www.semantic-web-journal.net/sites/default/files/swj296.pdf
48. Nikolov A, Uren V, Motta E, De Roeck A. Integration of semantically annotated data by the KnoFuss architecture. Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management; 2008. p. 265–74.
49. Niles I, Pease A. Towards a standard upper ontology. Formal Ontology in Information Systems. New York, New York, USA: ACM Press; 2001. doi:10.1145/505168.505170
50. ODE Browser. Available from: http://ode.openlinksw.com/
51. Open Library Project. Available from: http://openlibrary.org/about
52. Patni H, Henson C, Sheth A. Linked sensor data. International Symposium on Collaborative Technologies and Systems. IEEE. 2010; 362–70. doi:10.1109/CTS.2010.5478492
53. Patni H, Sahoo S. Provenance aware linked sensor data. 2nd Workshop on Trust and Privacy on the Social and Semantic Web. 2010. Available from: http://trac.assembla.com/soray/export/393/user/Marcell/readings/Linked Data/Provenance Aware Linked Sensor Data.pdf
54. Phuoc D, Hauswirth M. Linked open data in sensor data mashups. Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN09) in conjunction with ISWC; 2009. p. 1–16. Available from: http://ir.library.nuigalway.ie/xmlui/handle/10379/1113
55. Pschorr J, Henson C. Sensor discovery on linked data.2010. In Proceedings of the 7th. Available from: http://trac.assembla.com/soray/export/230/user/Marcell/readings/Linked Data/Sensor Discovery on Linked Data.pdf
56. Rodriguez M. A graph analysis of the linked data cloud.2009. arXiv preprint arXiv:0903.0194[cs.CY]. Mar 1–7. Available from: http://arxiv.org/abs/0903.0194
57. Schandl B. Representing linked data as virtual file systems. 2nd International Workshop on Linked Data on the Web (LDOW 2009).
58. Schandl B. TripFS Exposing file systems as linked data. Linked Open Data Triplification Challenge; 2009. Available from: http://triplify.org/files/challenge_2009/tripfs.pdf
59. Schandl B, Popitsch N. Lifting file systems into the linked data cloud with TripFS. 3rd International Workshop on Linked Data on the Web (LDOW2010).
60. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the user interface: strategies for effective human- computer interaction. 5th ed. Addison Wesley; 2009.
61. Shum S, Motta E, Domingue J. ScholOnto: an ontology-based digital library server for research documents and discourse. Int J Digit Libr. 2000; 3(3):1–29. Available from: http://link.springer.com/article/10.1007/s007990000034
62. State of the LOD Cloud. Available from: http://lod-cloud.net/state/
63. Stephens S, LaVigna D, DiLascio M, Luciano J. Aggregation of bioinformatics data using Semantic Web technology. Web Semantics: Science, Services and Agents on the World Wide Web. 2006; 4(3):216–21.
64. Strengthening the sharing of terrorism information to protect americans. Available from: http://nodis3.gsfc.nasa.gov/displayEO.cfm?id=EO_13356_
65. The growth of linked data. Available from: http://readwrite.com/2011/01/18/the-concept-of-linked-data#awesm=~ogs2e1ybeyWc5k
66. The LIBRIS project dataset. Available from: http://datahub.io/dataset/libris
67. The Linking Open Data cloud diagram. Available from: http://lod-cloud.net/
68. Tools Linked Data - Connect Distributed Data across the Web. Available from: http://linkeddata.org/tools
69. Tummarello G, Cyganiak R. Sig. ma: Live views on the web of data. Web Semantics: Science, Services and Agents on the World Wide Web. 2010. 8(4).
70. URIBurner. Available from: http://linkeddata.uriburner.com
71. URL vs. URI vs. URN: The confusion continues. Available from: http://damnhandy.com/2011/01/18/url-vs-uri-vs-urn-the-confusion-continues/
72. Using public data to fight a war. Available from: http://readwrite.com/2011/02/07/fighting-a-war-with-a-search-e#awesm=~ogy30APUs1eL0g
73. VIAF Datasets. Available from: http://datahub.io/dataset/viaf
74. W3C. Architecture of the World Wide Web. Available from: http://www.w3.org/TR/webarch/

75. W3C. CKAN- LinkingOpenData DataSets. Available from: http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKANmetainformation
76. W3C. DataSetRDFDumps. Available from: http://www.w3.org/wiki/DataSetRDFDumps
77. W3C. SemanticWebTools. Available from: http://www.w3.org/wiki/SemanticWebTools#head-805c63479c854babe4657d5184de605910f6d3e2
78. W3C. (2010). The multilingual web – linked open data and multilingualweb-LT requirements. Available from: http://www.multilingualweb.eu/program
79. W3C Library Linked Data Incubator Group. Available from: http://www.w3.org/2005/Incubator/lld/
80. Waterman KK, Wang S. Prototyping fusion center information sharing; implementing policy reasoning over cross-jurisdictional data transactions occurring in a decentralized environment. IEEE International Conference on Technologies for Homeland Security (HST); 2010. IEEE. p. 63–69.
81. Available from: http://blog.dbpedia.org/category/dataset-releases/
82. Available from: http://www.openlinksw.com/dataspace/dav/wiki/Main/VOSRDF
83. Available from: http://code.google.com/p/google-refine/
84. Available from: http://callimachusproject.org/index.xhtml?view
85. Available from: http://code.google.com/p/paget/
86. Available from: http://sioc-project.org/exporters
87. Available from: http://en.wikipedia.org/wiki/Yahoo!_SearchMonkey
88. Available from: http://www.apple.com/
89. Available from: http://www.android.com/
90. Available from: http://www.youtube.com/
91. Available from: http://www.myspace.com/
92. Available from: http://annomation.open.ac.uk/annomation/annotate
93. Available from: http://dme.ait.ac.at/annotation/
94. Available from: http://4store.org/
95. Available from: http://d2rq.org/
96. Available from: http://www.franz.com/agraph/allegrograph/
97. Available from: http://sioc-project.org/exporters
98. Available from: http://en.wikipedia.org/wiki/Operation_Eagle_Claw