International Journal on Semantic Web and Information Systems
Volume 14 • Issue 4 • October-December 2018

On the Current State of Linked Open Data: Issues, Challenges, and Future Directions

Nosheen Fayyaz, University of Peshawar, Peshawar, Pakistan
Irfan Ullah, University of Peshawar, Peshawar, Pakistan
Shah Khusro, University of Peshawar, Peshawar, Pakistan

ABSTRACT

This article describes how Linked Open Data (LOD), under the umbrella of the Semantic Web, integrates the openly-published semantic information making it easily understandable and consumable by humans and machines. Currently, researchers have applied the principles of LOD in several domains including e-government, media, publications, geography, and life sciences. Besides the fast pace of research, the field is still an emerging one, where researchers face several prominent challenges and issues that need to resolve to exploit LOD to its fullest. In this article, the authors have identified challenges, issues, and research opportunities in the publishing, management, linking, and consumption of LOD. The research work presented here will grab the attention of researchers and may aid to the current state-of-the-art in this area.

KEYWORDS

Big Data, Information Retrieval, Linked Data, Linked Open Data, Ontology, Semantic Web

INTRODUCTION

The World Wide Web is growing at a fast pace with lots of data available, but finding relevant information efficiently and timely is still a challenging job. Therefore, a layer of semantics is added to the current Web to make the machines understand the web documents (Berners-Lee, Hendler, & Lassila, 2001). The Semantic Web not only puts data on the Web but also establishes links among data so that humans and machines can explore the web of data rather than the web of documents and find other related data (Berners-Lee, 2006; Bizer, Heath, & Berners-Lee, 2009). This makes the Web more meaningful with the emergence of new paradigms such as Linked Open Data (LOD). Linked Open Data, an application of Semantic Web, can be defined as the Linked Data that is released under the open license and can be reused for free. Tim Berners-Lee presented a set of four basic rules for publishing, connecting and consuming the structured data on the Web (Berners-Lee, 2006). This Linked Data should have some general features including openness, modularity, scalability, and connectivity (Berners-Lee, 2009). These features make LOD as a source of knowledge discovery with several applications in multiple domains including e-government, health, education, and research.

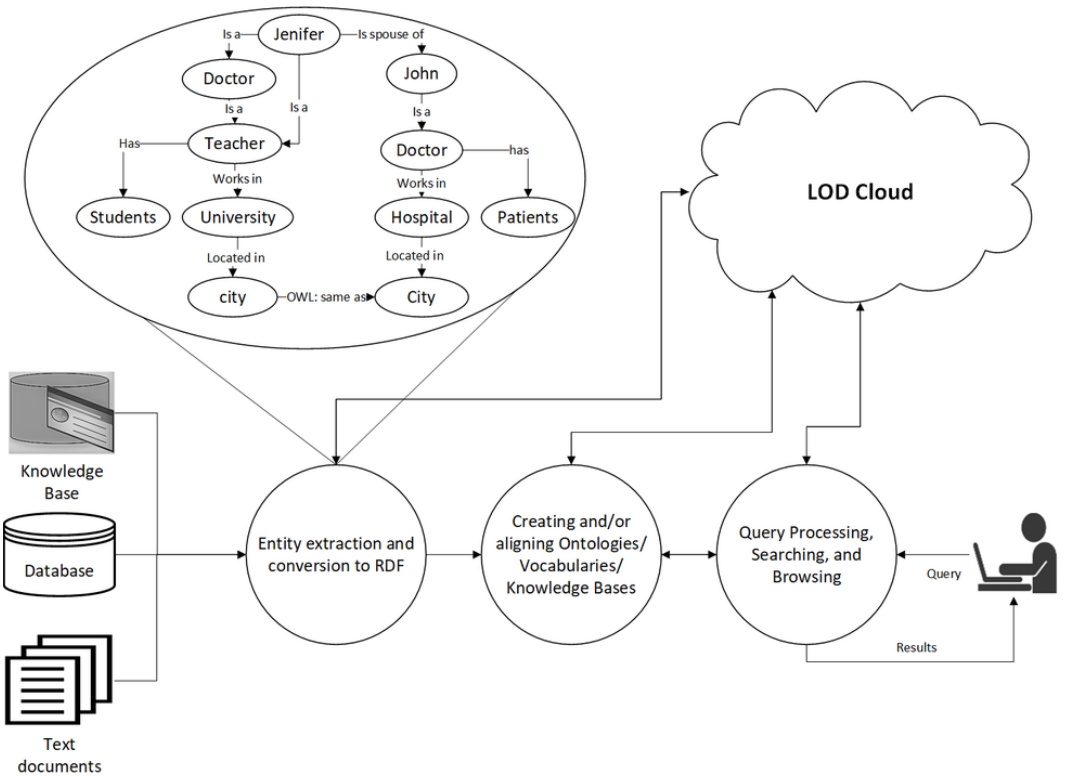
Publishing data on LOD requires five basic steps including data preparation, dataset selection, plugging into the LOD cloud, its announcement, and identifying social contract ("Best Practices for Publishing Linked Data," 2014; "Cookbook for Open Government Linked Data," 2011; Hausenblas, 2009; Peinl, 2016). Figure 1 graphically illustrates these steps. First, the data is converted to triples. Second, the desired datasets (which could enrich the content) are selected (using the "follow your

DOI: 10.4018/IJSWIS.2018100106

nose" principle). Third, the dataset is plugged into the LOD cloud for integrating data seamlessly to the chosen datasets to allow users and machines to consume it. Fourth, the data is published on a web server for its use and reuse. Finally, the implicit social contract with the consumers is recognized, which includes the responsibility of data availability, data freshness, accuracy, and maintenance ("Best Practices for Publishing Linked Data," 2014; "Cookbook for Open Government Linked Data," 2011; Hausenblas, 2009). The details of tools for the data extraction, transformation, storage, visualization and quality in publishing data on LOD can be found in the relevant literature (Bikakis, Tsinaraki, Gioldasis, Stavrakantonakis, & Christodoulakis, 2013; Butt, Haller, & Xie, 2015; Ferrara, Meo, Fiumara, & Baumgartner, 2014; Gangemi, 2013; Purohit et al., 2016).

Several review articles and survey papers have been published on the various aspects of the state-of-the-art in Linked Open Data and its applications. (Purohit et al., 2016) identified the best practices for publishing scientific research Linked Data and explored several tools for data extraction, transformation, storage, visualization and quality. Similarly, (Mouzakitis et al., 2017) explored the state-of-the-art in Linked Data technologies especially tools and identified their shortcomings and common barriers in their usage. They also proposed a framework to maintain and promote the interlinking capabilities of Linked Data so that novice users may use it (Mouzakitis et al., 2017). (Emani, Cullot, & Nicolle, 2015) focused big Linked Data and identified several issues regarding publishing, processing and linking with LOD and non-LOD resources commonly known as "Billion Triple Challenge" as well as the efficient use of semantics in integrating and designing database management systems (Emani et al., 2015). (Nentwig, Hartung, Ngonga Ngomo, & Rahm, 2017) compare multiple link discovery frameworks based on effectiveness (precision and accuracy), efficiency (speed and scalability), and a set of evaluation criteria including linking specifications,

Figure 1. A typical LOD environment



matching methods, runtime optimization, input formats, type of links, supported GUI and the availability of the software (Nentwig et al., 2017). (Bikakis & Sellis, 2016) explored several challenges in exploring and visualizing big Linked Data. (Zaveri et al., 2016) conducted a systematic literature review on LOD and identified, classified and compared 21 quality assessment methodologies using 23 quality dimensions and presented quality measures for data to be accessible, intrinsic, contextual, dynamic, trustworthy and representable. (Ristoski & Paulheim, 2016) researched on the applications of Data Mining and Knowledge Discovery techniques in Semantic Web. A detailed description of methodology and benchmarks of cross-media and some of their issues and challenges can be found in (Y. Peng, Huang, & Zhao, 2017). A comprehensive review of LOD-based semantic video annotation systems and their evaluation can be found in (Khan, Khusro, & Ullah, 2017; Sikos, 2017). (Safarov, Meijer, & Grimmelikhuijsen, 2017) conducted a systematic literature review on the use of open government data and its effects, moderating conditions like quality and legal issues, and users as well as applying it in multiple scenarios such as hackathons, data analysis, and as a research tool.

The list of these review and survey articles confirm that LOD is used in multiple domains and researchers are taking a keen interest in its adoption, linkage, consumption, and quality. However, these articles individually focus on specific aspects of a specific problem and ignore the others, where finding a general and holistic overview of multiple aspects/phases of LOD including publishing, linking and consumption is difficult. In addition, challenges and issues covered in these articles are also limited to a specific domain or phase. Currently, to the best of our knowledge, no one has attempted to give an overall picture of the Linked Open Data issues and challenges and the future research opportunities. The research presented in this article identifies some of the prominent challenges and issues in publishing, linking and consuming LOD by reviewing more than 100 primary research articles from several well-reputed journals and conferences. The most relevant, related and updated papers were then selected and included in the article. The rest of the paper identifies issues and challenges in LOD, which is followed by research opportunities and then conclusion. References are furnished at the end.

ISSUES AND CHALLENGES IN LINKED OPEN DATA

This Section presents some of the prominent challenges and issues in publishing, linking and consuming Linked Open Data. To do this, we critically reviewed the state-of-the-art literature (2013-2017) along with some review articles (Bikakis & Sellis, 2016; Emani et al., 2015; Khan et al., 2017; Mouzakitis et al., 2017; Nentwig et al., 2017; Y. Peng et al., 2017; Purohit et al., 2016; Ristoski & Paulheim, 2016; Safarov et al., 2017; Sikos, 2017; Sikos & Powers, 2015; Zaveri et al., 2016). The following sections briefly report on these challenges and issues.

PUBLISHING AND MANAGING HUGE DATASETS

The introduction of new datasets continuously adds to the massive size of LOD cloud. These datasets are published using the existing tools, which make them reusable in the form of RDF (Mouzakitis et al., 2017). However, this approach leads to the Billion Triple Challenge dataset, where each entity is available as a triple with descriptions from a single vocabulary and linked to other vocabularies or knowledge bases. The Billion Triple Challenge gives rise to specific issues including inconsistencies and noise, ontology hijacking, accessibilities and dereferenceability of URIs, and syntax errors (Emani et al., 2015). The RDF conversion tools are difficult to use and lack in the graphical user interface, which not only requires an expert's time but also challenging to understand for non-technical users. In addition, few tools support format heterogeneity. Therefore, additional tools are required to convert data from different formats to RDF (Mouzakitis et al., 2017). Most of these datasets wholly or partially lack in the associated metadata, which could otherwise be very helpful towards their discovery. Moreover, the alignment of entities and alignment of vocabularies need sophisticated algorithms, which is a big challenge (Ding, Peristeras, & Hausenblas, 2012). In addition, data ownership, its accuracy, handling

individual's data, and organizing big data are challenges to be considered (Kaisler, Armour, Espinosa, & Money, 2013). Furthermore, issues exist in using proprietary/non-proprietary vocabularies, finding/ripping relevant datasets, and ranking schema using its quality and usefulness (Bizer, Boncz, Brodie, & Erling, 2012), data ownership, privacy and legal problems (Zuiderwijk, Janssen, & Susha, 2016).

Data from several application domains have been made available as LOD, where the nature and type of challenges and issues change from domain to domain and covering them all lies outside the main scope of the paper. Therefore, we take e-Government as an example domain and explore some challenges and issues in publishing its data as linked open data on LOD cloud by following several steps including specification, modeling, generation, publication, and exploitation (Villazón-Terrazas, Vilches-Blázquez, Corcho, & Gómez-Pérez, 2011). Issues in following these steps include obtaining the open license of data, developing metadata standards, handling temporal, geographical and methodological gaps in datasets, designing user interfaces to present data from different datasets in a normalized form, resolving co-references, and ensuring data transparency (Shadbolt et al., 2012). This open data can be decentralized to integrate real-world entities including public agencies, public services and laws, each responsible for managing their own liked data (Kalampokis, Tambouris, & Tarabanis, 2013). The data needs to be made transparent for public consumption along with ensuring its privacy of the users (Janssen & van den Hoven, 2015; Thompson, Ravindran, & Nicosia, 2015). In conclusion, before publishing its data on LOD cloud, it is necessary for the e-government research to review its objectives, institutes, funders, and beneficiaries (Heeks & Bailur, 2007). For example, the usability of linked government data is a challenge when it comes to drawing useful inferences for the benefit of users and in policy-making (Weerakkody, Irani, Kapoor, Sivarajah, & Dwivedi, 2017). In addition, the government and public should be well-prepared in coping with uncertainties regarding data and its usage including data security, privacy, misuse, and dependencies among data from multiple organizations (Klievink, Romijn, Cunningham, & de Bruijn, 2017). Moreover, data security and privacy, in most cases like bank transactions and medical records, etc., need predefined strategies for sharing information including access control or certification and anonymization. However, developing secured certification is challenging whereas anonymization may introduce new challenges of data analysis due to data uncertainties (Wang, Xu, Fujita, & Liu, 2016). Finally, government agencies consider their data as a source of power and discourage their data to open their data to other departments and agencies (Huang, Lai, & Zhou, 2017).

LINKING DATASETS ON LINKED OPEN DATA

Link Discovery in Multiple Domains

Consuming data from heterogeneous LOD sources in a truly integrated environment is a daunting task with several challenges including co-reference, ontology mapping, aggregation from distributed sources, resource discovery, and query spanning over multiple datasets (Millard, Glaser, Salvadores, & Shadbolt, 2010). Several solutions have been proposed in the form of frameworks including SILK, LIMES, which are discussed and evaluated in detail in (Nentwig et al., 2017). Link discovery frameworks and tools should have two essential features including the effectiveness (precision and accuracy) and efficiency (speed and scalability) in combination with parallel processing, blocking and filtering may also improve their efficiency. The utilization of LOD can be improved by the usage or exploitation of the existing links or background knowledge (data sources and dictionaries) of datasets (Nentwig et al., 2017).

According to (Ristoski & Paulheim, 2016), few approaches use links among datasets and therefore, are rarely able to exploit the endless possibilities with the full knowledge of the Semantic Web. The approaches that leverage data from only one dataset, therefore, stay below what is possible with LOD. The reason behind this limitation is the non-trivial task of developing "schema-agnostic" applications, although machine-interpretable schemas are available. In addition, scalability challenges

lie in developing solutions that could exploit the whole LOD as background knowledge by following links autonomously. Furthermore, Research is needed to discover whether this limited integration is due to limited awareness, lack of suitable and reusable ontologies or their imperfect fitness to the practical problems (Ristoski & Paulheim, 2016).

Matching Ontologies

Ontology matching is the process of identifying the semantically related entities from different ontologies and the correspondence among them (Shvaiko & Euzenat, 2013). There are two common types of matchers: the element-based and structure-based. The former uses similarity, e.g., n-grams and edit distance in string matching. The latter is more sophisticated and uses the context of resources and background knowledge (Nentwig et al., 2017). During this process of matching and mapping the ontologies, several challenges and issues arise like the need of reliable, mature, fast, and efficient ontology matching methods and techniques to automate the alignment especially in the case of large datasets. In addition, ontology matching should be context-dependent to maximize recall. However, care should be taken to avoid incorrect matches and low precision (Shvaiko & Euzenat, 2013). Also, the mapping process could be improved with the help of semi-automatic machine learning, semantic mapping, holistic ontology matching for complex relations and practical applications (Otero-Cerdeira, Rodríguez-Martínez, & Gómez-Rodríguez, 2015).

Several issues have been investigated in (Zhang, Gentile, Blomqvist, Augenstein, & Ciravegna, 2017), e.g., many of the LOD ontologies are incomplete and noisy and are not well-structured and well defined. The alignment of heterogeneous relationships is not properly addressed because of synonymy and polysemy, lack of structural information about relations, inconsistency in the meaning and usage of schemata. The interoperability of LOD-based applications is also affected due to rare links at schema level LOD datasets (as most of the research focused the data level links in LOD datasets) (Zhang et al., 2017).

Handling Linked Data Streams

The data streams from several sources including sensors (e.g., GPS, and RFID) and social media applications are diverse, dynamic and frequently change. This nature of streaming datasets makes them difficult to integrate with static and less frequently changing datasets on the LOD. To keep LOD datasets updated, the data coming from these data streams must be handled in real-time to answer user search queries with updated results (Le-Phuoc, Dao-Tran, Parreira, & Hauswirth, 2011).

The challenging issues are to extract the RDF from unstructured (textual) data streams (Gerber et al., 2013) to compute the efficiency, manage uncertainty, define provenance of reasoning, provide native support for representation and reasoning over temporal knowledge, and draw inferences from the vast linked streaming data (Ye et al., 2015). Along with that, tracking the provenance of data, investigation of the matching ontology and the re-publishing of data (already published as plain text) using the semantics of Linked data uses several ontologies to get aligned to grasp the actual concepts are also challenging (Yu & Liu, 2015). Another issue is to handle incomplete and delayed information, and the skewed distribution of classes (unbalance) in the streaming datasets (Krempl et al., 2014).

Linking Social Media to LOD

The Social Web has connected users allowing them to generate and share contents in the form of short messages. The additional semantic information about this content can be obtained by integrating social media with LOD (Cano, Varga, Rowe, Ciravegna, & He, 2013). Some general issues of social media are the social identification of the user, secure data communication, evaluation and testing of the privacy-preserving services of real-data, efficient methods for the fusion of multimedia objects with social media applications (Bello-Orgaz, Jung, & Camacho, 2016). Whereas linking the social Web with LOD brings additional challenges and issues including ensuring data accuracy, handling unorganized data from several social media applications, (ensuring the privacy of) data about users

and deciding data ownership. The "Web is evolving from data-centric to knowledge-centric (Kaisler et al., 2013)". Therefore, working on data without taking care of its metadata would not give us useful insights and the required knowledge. Linking the Social Web to the LOD can be challenging due to a multitude of technologies, considering the cultural aspects of users, their security and privacy, trust, and freedom of speech (Dini, 2016).

The social media user wants to access, participate, and remain anonymous. Their desire to be everywhere but nowhere helps governments in policy making by using the user-generated content (Zavattaro & Sementelli, 2014). However, this concept of omnipresence introduces new issues of privacy, which may lead to asymmetry of information. Therefore, more focused strategies are required (Zavattaro & Sementelli, 2014). The social media content can be used to gather information, e.g., news regarding emergency, events, disasters, etc., which needs information filtering systems. Such systems face two challenges, the lack of context and dynamic vocabulary. The LOD knowledge bases can provide the required content, but the issue is the infrequent updating of these knowledge bases, which needs further research inquiry (Sheth & Kapanipathi, 2016).

CONSUMING LINKED OPEN DATA

A user consumes the linked data from multiple sources; this data arrives over HTTP, type checked and then loaded into the local triple store (Ciobanu, Horne, & Sassone, 2015). A user can access LOD either through the LOD browsers (e.g., Tabulator, Disco, Link Sailor, and Marbles) or LOD search engines including Sigma and Falcons. LOD browsers explore the Web of Data using RDF links, whereas the search engines crawl, integrate and search linked data from multiple relevant data sources. Sigma and Falcons provide keyword-based, interactive search services to provide links, by using the structure of data and filtering results by classes, getting summarized views, and showing related links (Heath & Bizer, 2011). Semantic Web Search Engine (SWSE) contains traditional web search engine-like features including crawling, indexing, browsing, user interface and works on RDFs along with scalability and robustness to cope with the heterogeneity of large data sets especially ones with noisy and conflicting data (Hogan et al., 2011).

Despite the availability of several LOD browsers and search engines, as discussed, users face several challenges and issues in consumption. These include: users must understand the working of Semantic Web including RDF and SPARQL and the literals with complex data types for spatial and temporal purposes (Scheider et al., 2015). One possible solution, according to (Scheider et al., 2015) is developing a mechanism for an interactive and explorative visual query construction but comes with several issues including vocabulary explanation, filtering of space-time (windows), and linking them to other displays. In addition, on-the-fly link traversal and federated querying may probably suffer from scalability issues as benefiting from LOD requires discovering new data sources and making them available to the applications promptly. This issue can be mitigated with widespread crawling and caching (Bizer et al., 2009; Oliveira, Delgado, & Assaife, 2017). The distributed and inconsistent data gives problems in consuming data like co-reference (duplicate identifiers), ontology mapping, resource discovery, and querying multiple datasets (Millard et al., 2010). In addition, discovering and updating data from a triple store needs an error-free and simple scripting language for programming background processing and type-checking (Ciobanu et al., 2015).

Assessing the Quality of Linked Open Data

The linked data varies in quality, which may affect its consumption. It is, therefore, necessary to assess this data for its quality before consuming it. The quality of linked data measures its usefulness and utility regarding data conformance and quality of datasets. The conformance checks whether the RDF, generated for data, follows the basic principles of LOD, and reusable by looking at the properties of labels, class definitions and links for data discovery. Data providers are responsible for ensuring conformance (Hogan et al., 2012). The data may not always adhere to the similar scheme,

i.e., its triples may not precisely follow the definitions in associated RDF schema or ontology (Tonon, Catasta, Demartini, & Cudré-Mauroux, 2015). This affects adversely the quality of the results obtained from LOD. A dataset is of high-quality if it is free from missing or erroneous data, syntax errors, triplification errors, misleading equivalence (owl:sameAs) links, unavailable SPARQL endpoints (Hitzler & Janowicz, 2013), and problems due to mapping from semi-structured data (Dimou et al., 2015).

Most often, datasets are developed by mapping the semi-structured data to RDF triple. This mapping may violate the dataset schema due to either error in the original data sources or the incorrect usage of the schema, therefore, affecting the quality of resulting datasets. Capturing these violations at the dataset level is complicated and time-consuming, and therefore, it is necessary to detect and correct them during mapping (Dimou et al., 2015). Either the publishers or the authors of the LOD cloud diagram (Abele, McCrae, Buitelaar, Jentzsch, & Cyganiak, 2017) classify the data into different topical domains based on its characteristic features. The classification is prone to errors, and therefore, machine learning techniques could be used to automate it and reduce these errors (Meusel, Spahiu, Bizer, & Paulheim, 2015).

The community members may have different perspectives while creating datasets and therefore the data models they adopt leading to quality issues including missing and erroneous data, syntax errors, triplification errors, misleading equivalence links (owl:sameAs), and unavailable SPARQL endpoints (Hitzler & Janowicz, 2013). To mitigate these issues, (Mendes, Mühleisen, & Bizer, 2012) devised the Sieve framework, which is integrated into Linked Data Integration Framework (LDIF) for accessing data, mapping schemas, resolving identities, assessing the quality and fusion of data, and resolving conflicts in data integration process. According to (Zaveri et al., 2016), the linked data should be accessible (available and licensed); intrinsic (correct, compact and consistent); trustworthy; dynamic (dataset dynamicity, i.e., freshness and change ratio); contextual (relevant and complete); and representable. (Radulovic, Mihindukulasooriya, García-Castro, & Gómez-Pérez, 2017) used these parameters and developed a comprehensive hierarchical quality assessment model that classifies as well as defines new quality measures (along with related formulas) to extend the ISO 25012 data quality model to infer the evaluation results of the datasets and use them for interoperability (Radulovic et al., 2017). These generalized quality features of LOD require several low-level considerations that span from the publication of datasets to the consumption and requires the involvement of the researchers, practitioners, and organizations.

Visualizing Linked Open Data

Linked and Open Data should be visualized to the user to increase its consumption (Dadzie & Rowe, 2011; Sabol et al., 2014). Several frameworks exist for such a visualization (Mouzakitis et al., 2017). The Linked Data Visualization Model (LDVM) framework (Klímek, Helmich, & Necaský, 2015), which follows visualization pipelines consisting of four phases: collecting source data, creating analytical abstraction, creating visualization abstraction and generating views. It supports creating reusable components and showing the content of a given dataset and its relations to other datasets. The implementation of LDVM in multiple facets can be beneficial like in distributed environment, Linked Data exploration and providing a user-friendly environment where one can look for the library of visualizers and analyzers (Klímek et al., 2015). LinkDaViz is an automatic visualization tool that recommends and generates visualizations to the user depending on the input data. It uses recommendation algorithm to generate results from a specific subset of data. However, it should be extended to deal with large datasets and accurately recommend the visualization by using machine learning algorithms (Thellmann, Galkin, Orlandi, & Auer, 2015). QueryVOWL allows users to place a visual query using the graphical elements of the VOWL (Visual Notation for OWL Ontologies), which is mapped to SPARQL (Haag, Lohmann, Siek, & Ertl, 2015). The QueryVOWL lacks specific advanced features like filtering the properties or introducing quantifiers. A flexible, expressive, and

easy-to-use visual query language is needed to enable a non-technical user to place a graphical query (Haag et al., 2015).

The massive size of most of the contemporary datasets has been a challenge in exploration and visualization of Linked Data. The non-availability of query and API endpoints for search & retrieval and handling streaming data in dynamic settings involve the processing overhead. Efficient and scalable data structures and algorithms are required to deal with the crucial issues of storage, management, access, presentation and interaction (e.g., search, zoom, pan, and drill-down, etc.) over large and dynamic datasets as well as to operate efficiently on machines with limited memory and computational resources. Users must be able to easily explore large datasets regarding performance, analysis, presentation, and interaction while considering their diversity of preferences and requirements (Bikakis & Sellis, 2016).

Despite the ordinary users, an expert of SPARQL also needs some manual work (look up in documentation). The reasons include the unavailability of the required options and classes (Mouzakitis et al., 2017), limited or incomplete evaluation of the visualization techniques, mapping and synchronization among multiple visualizations, and issues in visualization and annotation of dynamic graphs (Beck, Burch, Diehl, & Weiskopf, 2017). Several issues exist in visualizing social data; efforts are required in the collection, storage, and analysis of data especially the spatiotemporal data about moving people and objects at a large scale (Bello-Orgaz et al., 2016). In conclusion, there is much more to solve in LOD visualization as user interfaces are still at their infancy, research efforts are required to exploit Linked Open Data to its fullest (Dimou et al., 2017).

LOD is also consumed in information filtering especially in making content-based recommendations. LOD-based recommenders mostly target the data serendipity problem, i.e., finding unusual, unknown but interesting items for the user (Peska & Vojtas, 2015). These recommenders combine open data from multiple domains in suggesting related resources. The knowledge extracted from the links in LOD is used to enrich the semantics of the objects and can find out the similarity among them using similarity measures. A proper methodology is required to exploit fully and appropriately the knowledge of multiple domains in developing recommenders (Yang & Hsu, 2015). Additionally, LOD should be considered in alleviating cold-start, diversity, novelty and serendipity issues in recommender systems. The visualizations should work around user needs and their emotions, for which different techniques, design guidelines, and user interfaces should be considered, evaluated, and compared. The user interfaces should be adaptive to different devices, display and interactive technologies (He, Parra, & Verbert, 2016). New algorithms along with evaluation metrics should be developed to help mitigate the challenge of diversity in recommendation systems (Kunaver & Porl, 2017).

A user may find it challenging to use SPARQL queries to find answers to their question. Therefore, researchers propose to consume LOD in Question & Answering (Q&A) systems. Several solutions (Shekarpour, Ngomo, & Auer, 2013; Yahya et al., 2012) have been devised to accept user queries in natural language and transform them to SPARQL to search against single or multiple datasets to find answers. The natural language questions can be written for a specific RDF repository (Zou et al., 2014). The more complex question may require answers from both structured and unstructured data. To do this, HAWK (Hybrid Question Answering using Linked Data), a hybrid entity-based search solution can be used, which searches the linked and textual data sources to generate SPARQL queries from the predicate-argument structures. However, it requires an optimal ranking solution to map this predicate-argument tree to possible interpretations for query generation (Usbeck, Ngomo, Bühmann, & Unger, 2015). The LOD-based Q&A systems face several challenges including understanding questions by developing semantic parsers with deep learning capabilities to separate semantic and syntactic properties; resolving heterogeneity of datasets; bringing interoperability among components; and ensuring data quality (Shekarpour et al., 2016). Other challenges include the lexical gaps (the differences in the question's vocabulary and knowledge base), ambiguity in natural language, multilingualism, complex queries, distributed knowledge, procedural queries (as

the existing knowledge bases have no procedural knowledge), and adding the user temporal or spatial information to clarify the search query (Höffner et al., 2016).

To document a resource in multiple languages, RDF(S) or SKOS labels are used, which makes the Linked Data language-independent (Gracia et al., 2012). DBpedia uses an ontology to extract the multilingual knowledge of Wikipedia and maps it to language specifications to form a dataset, which is then linked to more than 30 datasets in the LOD cloud (Lehmann et al., 2015). BabelNet, a large multilingual encyclopedic dictionary and ontology, is also now on Linked Data following the LEMON (Lexicon Model for Ontology) model (Ehrmann et al., 2014). However, the challenge is to provide language-specific information to the user from this language-independent Web of Data, which can be achieved by adding a layer on top of Linked Data having multiple services and resources. This may include the linguistic information in multiple languages, its mappings, and a proper mechanism for the access and traversal of the Linked Data across multiple languages (Ehrmann et al., 2014; Gracia et al., 2012). However, the lexical gap further widens by including multiple languages due to the existence of a "cross-lingual gap" (Hakimov, Jebbara, & Cimiano, 2017). To cope with this issue, the AMUSE software could be trained for any language, which needs no language-specific knowledge or heuristics. However, it was tested and evaluated only with Question Answering over Linked Data (QALD-6) train dataset for German, English, and Spanish languages, and therefore, needs to be tested and evaluated with other datasets and languages (Hakimov et al., 2017).

Multimedia content is easy to produce but somewhat challenging to find and reuse on the Web. The Linked Data principles should be applied to fragments of multimedia resources, URIs should be assigned for addressing them by following four dimensions such as time, space, track and name (Khan et al., 2017). The currently available systems for video annotations have complex interfaces, consume many resources and have limited usability for the searching, indexing, and browsing of specific scene or theme. These difficulties are primarily due to limited use of semantics in video annotations and incompleteness in domain level ontologies which will enable the users to connect the related scenes, objects and themes (Khusro, Khan, & Ullah, 2016).

There are several tools used for semantic multimedia annotation like Annomation[1], Annotorius[2], ImageSnippets[3], OpenVideoAnnotation[4], ConnectMe, and LinkedTV[5]. According to (Nixon & Troncy, 2014), these tools suffer from several issues including the automatic creation of media fragments, classification of media fragments, relations among media fragments, copyrights, reuse and remix of media, and finally a consensus should be developed on data formats, data models, vocabularies, reuse of fragments, retrieval, and presentation of multimedia. The existing LOD-based video annotation systems lack in support for annotating specific scenes, objects, events, and themes and limited in linking and searching related scenes, objects, themes and events available on different data sources (Khan et al., 2017).

The automatic processing of text-based metadata annotation has some limitations, especially with the manually added tags (e.g., in YouTube), requiring structured and machine-readable annotations. The interoperability and same information representation issues are also required to be addressed (Sikos & Powers, 2015). In addition, the complex user interfaces, limited use of Semantic Web technologies and limited availability of datasets are also challenging issues that hinder the usage of LOD-based video annotation systems (Khan et al., 2017). Video indexing and retrieval using ontology mostly use concept detection, which can be made more accurate by analyzing the relation between detected concepts. This analysis can be based on computing the concept co-occurrence, visual descriptors and hybrid semantic similarity, which provide the context for classifying video content (Sikos, 2017). Moreover, users may be interested in retrieving more than one type of multimedia content including image, audio or video with flexible requirements, e.g., using an image file to retrieve its related audio file.

For cross-media retrieval, not only a few datasets are publically available but they also suffer from several shortcomings including small size, the rationale behind categorization, and a limited number of media types. To construct high-quality datasets, several aspects including the type and number of

categories, the number of media types in the dataset, and the size of the dataset are used. In addition, efficient and effective methods are required to improve the accuracy of cross-media retrieval, handle the data type heterogeneity, and development of more practical applications (Peng et al., 2017).

RESEARCH OPPORTUNITIES

Linked Open Data can be applied in any sphere of life; however, its applicability can be beneficial when we are able to cope with certain prominent issues and challenges discussed in Section 3. Based on the findings in Section 3, this Section identifies some research opportunities in the publishing and managing datasets, their linking with other datasets on LOD, and their consumption. To do this, we critically reviewed the state-of-the-art literature (2013-2017) along with some review articles (Bikakis & Sellis, 2016; Emani et al., 2015; Khan et al., 2017; Mouzakitis et al., 2017; Nentwig et al., 2017; Y. Peng et al., 2017; Purohit et al., 2016; Ristoski & Paulheim, 2016; Safarov et al., 2017; Sikos, 2017; Sikos & Powers, 2015; Zaveri et al., 2016). Researchers and practitioners in Linked and Open Data may find this Section more useful. The identified research opportunities are listed below:

- The domain of LOD has future potentials like the usability of open data to infer insights from datasets and considering it in decision-making, but the results are sometimes poor than the expected, because the open data may not be linked or structured properly. Due to this issue, a common citizen may be unable to draw meaningful conclusions from government data (Weerakkody et al., 2017). In addition, open government data has several utilities including innovation, data analysis, decision making, and research; still, a gap exists between literature and its utilization, which needs further research to identify the effects of this data in concrete terms rather than making estimates. The relation of a user with the type of utilization and its effects and their moderating conditions should also be considered (Safarov et al., 2017);
- The data on the LOD cloud is growing bigger and bigger, where the insufficient conceptual information about these datasets makes it challenging to find the relevant datasets from this big data (Jain, Hitzler, Yeh, Verma, & Sheth, 2010). Therefore, along with handling big data on LOD, ontology, entity and vocabulary alignments should be focused (Ding et al., 2012). Other issues regarding big data include the collection of data about individuals, the accuracy and ownership of data (Kaisler et al., 2013), the complexity of tools for converting data to RDF, and developing user-friendly graphical user interfaces (Mouzakitis et al., 2017);
- Applying LOD principles on government's data has made the official data available to general public, which introduces several opportunities. These include the open license of data, metadata standards, presenting the information in a user-friendly manner, co-reference resolution, transparency, and privacy of open government data (Shadbolt et al., 2012);
- The e-Government domain has been exploited almost to its fullest by researchers, whereas other domains present many research opportunities regarding their publishing, linking, management, and consumption on LOD. These domains include multimedia, health and Life Sciences, Education, Geography, Library and Information Science, Social Media and user-generated content, and cross-domains;
- Data Mining and Knowledge Discovery techniques are rarely integrated into ontologies or expressive schema and reasoning. Therefore, it is needed to investigate this limited awareness and lack of such ontologies that could be reusable, suitable, and fit into this practical domain (Ristoski & Paulheim, 2016);
- Linked streaming data is used for water resource management, flood control, emergency, and can be used for more similar purposes like geo-streaming data. Several services regarding the linked streaming data can be provided to users as Linked API to show the provenance of data, the data processing modules, data aggregation and interpolation, etc. (Yu & Liu, 2015);

- The extraction of RDF from unstructured and heterogeneous data streams and their integration with other datasets should be further worked on so that continuous and simultaneous user queries can be handled in real-time (Gerber et al., 2013; Le-Phuoc et al., 2011). A quick and efficient mapping language for the conversion of streams to RDF is needed with caching and preprocessing capabilities to expedite the creation of the links along with some storage capabilities (Llanes, Casanova, & Lemus, 2017);
- The impact and analysis of social influence on social data is a research interest, and studies are conducted for its evaluation metrics and models, casual relationships, differentiating the positive, negative and controversial influences, and evaluating the influence of heterogeneous social networks (S. Peng, Wang, & Xie, 2017). Therefore, linked social media should also be considered for social influence analysis, e.g., the identification of an influential individual in the linked datasets who is changing the perception of people;
- According to (Nentwig et al., 2017), about 44% of the datasets on LOD cloud are not linked (although available in RDF) to other datasets. This happens due to problems in link discovery as manually establishing links among datasets is a daunting and tedious job. This can be resolved by developing automatic link discovery methods with two desirable features, i.e., effectiveness and efficiency, something that has already been tried up to some extent in the link discovery frameworks. These frameworks still use the property-based and structure-based matching techniques and do not exploit the LOD by their existing links and background knowledge like dictionaries on the LOD cloud (Nentwig et al., 2017);
- To publish and link different datasets is comparatively easy but to query and access data from these datasets especially from cross-domains in a rational manner needs efforts to resolve specific issues like co-reference, ontology mapping, resource discovery, and querying multiple datasets (Millard et al., 2010). Sometimes, a user queries multiple datasets containing knowledge in different languages and expects the results in a specific language, but the current solutions are unable to map these datasets and generate the results in the target language. In this regard, several research opportunities exist in mapping datasets, providing language-specific information to the users along with developing additional services (Ehrmann et al., 2014);
- There are several research opportunities in LOD consumption. Querying LOD in SPARQL makes it difficult for an ordinary user to write the query. Therefore, either the system should ask the user in natural language and automatically and accurately convert it to SPARQL (Yahya et al., 2012) or an alternative query language or a visual query language should be developed that is more expressive and easy to use (Scheider et al., 2015). Along with that, some other issues that need the researchers' attention are the lexical gaps, ambiguity, multilingualism, complex queries, procedural queries, distributed knowledge, temporal or spatial information of the user to suggest the language for response to the user (Höffner et al., 2016);
- The outputs of LOD query are in technical formats, and their consumption is problematic for non-experts. Therefore, an easy-to-understand graphical output should be developed (Sabol et al., 2014) for some common scenarios. These scenarios include finding some schools and hospitals in a particular area, number of relief camps in areas affected due to natural disasters, pointing out a population's living standards and so forth;
- Users need user-friendly and robust recommender systems that give best choices to them. Therefore, we need to develop LOD-based recommenders for overcoming the problems like cold-start, scattered data, and content analysis (Noia, Mirizzi, Ostuni, Romito, & Zanker, 2012). Good recommenders can save the time of students in digging out the textual or multimedia content (Fernández, Llaves, & Corcho, 2014). In this regard, the online multimedia annotations still have space for research regarding the automatic creation and classification of media fragments and their linkage, copyrights, reuse and remix of media, interoperability, retrieval, and the presentation of fragments (Nixon & Troncy, 2014; Sikos & Powers, 2015);

- The geographic data is also available on the LOD cloud but has some inaccuracies in error estimations, shortened coordinates or un-captured areas, etc., (Ahlers, 2013). (Moura, Davis, & Fonseca, 2017) tried to integrate Linked Data to Gazetteer like GeoNames and pointed out that 95% of GeoName entities are not linked with other datasets of Linked Data. This lacking in the actual integration of sources is due to missing links, duplications and problems with integration techniques. Despite the well-structured and official data of GeoNames, its data is distorted, duplicated and has mismatched relationships. Apart from this, the volunteered datasets like DBpedia and Freebase have no enough documentation. There are issues in their predicates which can be resolved by clustering and classification techniques. Another gazetteer, LinkedgeoData follows no LOD principles, and therefore the need is to work out practical solutions for these gazetteers to exploit the capabilities of LOD to its fullest (Moura et al., 2017);
- Library and Information Science presents several outstanding research opportunities including the conversion of semi-structured metadata to RDF, schema matching, automated indexing using machine learning, indexing multimedia data, distributed data management, and tracking the provenance of data to become a trustworthy source to the user (Latif, Scherp, & Tochtermann, 2016);
- Social media has connected people from different cultures, civilizations, and territories and opened new avenues of research in the use of sophisticated technologies and dealing with issues of privacy, security, and freedom of speech of users and their trust on the social media web applications (Dini, 2016). In addition, the freely available data of users on these applications can be used for constructive purposes including, e.g., identifying marketing trends, analyzing the behavior of people of a particular region and their attitude towards other regions of the world, and using in counter-terrorism;
- The quality of the generated response from the LOD depends upon the quality of data published on the Web. Therefore, an automatic conformance mechanism should be there to ensure that Linked Data principles have been followed and that the RDF triples are the correct mappings of the semi-structured data along with its adherence to its ontology (Tonon et al., 2015). This means that the quality issues including missing and faulty data, misleading equivalence links, syntax, and the unavailability of SPARQL endpoints etc., are some of the research avenues that need research attention (Hitzler & Janowicz, 2013);
- Mature ontology mapping techniques are required for automatic alignment and matching with the context, (Shvaiko & Euzenat, 2013), finding complex relations, using machine learning techniques in mapping, improving the precision of automatic matching, semantic mappings and bringing parallelism in the mapping processes (Otero-Cerdeira et al., 2015);
- Using Linked Data as the foundation, several applications associated with the Semantic Web vision such as intelligent agents, etc., could be developed (Bizer et al., 2009) to assist the busy users in daily activities. In addition, several application domains including education, digital libraries, the publishing industry and health can significantly benefit from semantics-aware and LOD-based web applications;
- The datasets are available for few domains including education, e-Commerce, and news, etc., and other domains such as politics etc., are yet to be covered (Khan et al., 2017). By appropriately developing new datasets and enriching the existing ones, users would be able to annotate and link videos based on related objects, scenes, events, and themes etc., enabling them to search, browsing and share videos in LOD-based video annotation systems in a user-friendly manner (Khan et al., 2017).

The list of research opportunities presented in this Section is not limited and can be further extended by doing further research in the area. However, by handling the research challenges presented in Section 3 and considering the research opportunities presented in this Section would make LOD and its application domains more useful and beneficial. These open issues invite researchers and

industry practitioners to come forward and play their vital role in mitigating these issues and challenges as well as further work on the research opportunities presented in this Section and the ones that lie beyond the current work.

CONCLUSION

Linked Open Data, under the umbrella of Semantic Web, deals with integrating the openly published semantic information making it easily understandable and consumable by humans and machines. Currently, researchers have applied the principles of LOD in several domains including e-government, media, publications, geography, life sciences, user-generated content, and others. However, besides the fast pace of research and development in LOD, it is still emerging where researchers may face prominent issues and challenges. Moving in this direction, we have tried to identify and report challenges, issues in publishing, linking, managing and consuming Linked Open Data. To mitigate these challenges and issues, we also tried to identify the potential areas of research, which hopefully, will open new avenues of research and development in the field of Semantic Web in general and Linked Open Data in specific.

One limitation of the present work is that we explored state of the art literature presented in English language only and were unable to cover other languages. In addition, due to limitations on the size of the paper, we covered only some of the challenges, issues and research opportunities and there is much more to discover. For example, LOD has been extensively studied from e-government perspective whereas it should be studied when applied in other domains. As a future work, we plan to study these domains in detail. We hope that researchers will find this article as a short introductory guide towards realizing and exploiting LOD to its fullest.

REFERENCES

Abele, A., McCrae, J. P., Buitelaar, P., Jentzsch, A., & Cyganiak, R. (2017). The Linking Open Data cloud diagram 2017. Retrieved March 08, 2018, from http://lod-cloud.net/

Ahlers, D. (2013). Assessment of the accuracy of GeoNames gazetteer data. *Paper presented at the Proceedings of the 7th Workshop on Geographic Information Retrieval*, Orlando, FL. doi:10.1145/2533888.2533938

Beck, F., Burch, M., Diehl, S., & Weiskopf, D. (2017). A Taxonomy and Survey of Dynamic Graph Visualization. *Computer Graphics Forum*, *36*(1), 133–159. doi:10.1111/cgf.12791

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, 45–59. doi:10.1016/j.inffus.2015.08.005

Berners-Lee, T. (2006, June 18, 2009). Design Issues: Linked Data. Retrieved March 08, 2018, from https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T. (2009, June 30, 2009). Design Issues: Putting Government Data Online. Retrieved March 08, 2018, from https://www.w3.org/DesignIssues/GovData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 28–37. doi:10.1038/scientificamerican0501-34 PMID:11341160

Best Practices for Publishing Linked Data. (2014). *W3C Working Group Note.* Retrieved 24 July 2016, from http://www.w3.org/TR/2014/NOTE-ld-bp-20140109/

Bikakis, N., & Sellis, T. (2016, March 15-18). Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art. *Paper presented at the The EDBT/ICDT 2016 Joint Conference*, Bordeaux, France.

Bikakis, N., Tsinaraki, C., Gioldasis, N., Stavrakantonakis, I., & Christodoulakis, S. (2013). The XML and Semantic Web Worlds: Technologies, Interoperability and Integration: A Survey of the State of the Art. In I. E. Anagnostopoulos, M. Bieliková, P. Mylonas, & N. Tsapatsoulis (Eds.), *Semantic Hyper/Multimedia Adaptation: Schemes and Applications* (pp. 319–360). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-28977-4_12

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: Four perspectives--four challenges. *SIGMOD Record*, *40*(4), 56–60. doi:10.1145/2094114.2094129

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22. doi:10.4018/jswis.2009081901

Butt, A. S., Haller, A., & Xie, L. (2015). A Taxonomy of Semantic Web Data Retrieval Techniques. *Paper presented at the 8th International Conference on Knowledge Capture*, Palisades, NY. doi:10.1145/2815833.2815846

Cano, A. E., Varga, A., Rowe, M., Ciravegna, F., & He, Y. (2013). Harnessing linked knowledge sources for topic classification in social media. *Paper presented at the 24th ACM Conference on Hypertext and Social Media*, Paris, France. doi:10.1145/2481492.2481497

Ciobanu, G., Horne, R., & Sassone, V. (2015). Minimal type inference for Linked Data consumers. *Journal of Logical and Algebraic Methods in Programming*, *84*(4), 485–504. doi:10.1016/j.jlamp.2014.12.005

Cookbook for Open Government Linked Data. (2011, March 15). *Linked Data Cookbook*. Retrieved 25 July 2016, from https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Dadzie, A.-S., & Rowe, M. (2011). Approaches to visualising linked data: A survey. *Semantic Web*, *2*(2), 89–124.

Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., & Van de Walle, R. (2015). *Assessing and Refining Mappingsto RDF to Improve Dataset Quality The Semantic Web-ISWC 2015* (pp. 133–149). Springer.

Dimou, A., Vahdati, S., Di Iorio, A., Lange, C., Verborgh, R., & Mannens, E. (2017). Challenges as enablers for high quality linked data: Insights from the semantic publishing challenge. *PeerJ Computer Science*, *3*, e105. doi:10.7717/peerj-cs.105

Ding, L., Peristeras, V., & Hausenblas, M. (2012). Linked open government data [Guest editors' introduction]. *IEEE Intelligent Systems*, *27*(3), 11–15. doi:10.1109/MIS.2012.56

Dini, A. A. (2016). The Current State of Social Media Research for eParticipation in Developing Countries: A Literature Review. *Paper presented at the 49th Hawaii International Conference on System Sciences (HICSS)*. doi:10.1109/HICSS.2016.339

Ehrmann, M., Cecconi, F., Vannella, D., McCrae, J. P., Cimiano, P., & Navigli, R. (2014). Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0. *Paper presented at the proceedings of Ninth International Conference on Language Resources and Evaluation (LREC-14)*.

Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable Big Data. *Computer Science Review*, *17*(C), 70–81. doi:10.1016/j.cosrev.2015.05.002

Fernández, J. D., Llaves, A., & Corcho, O. (2014). Efficient RDF Interchange (ERI) Format for RDF Data Streams. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, & C. Goble (Eds.), *The Semantic Web – ISWC 2014: 13th International Semantic Web Conference*, *Riva del Garda, Italy*, *October 19-23*. *Proceedings, Part II* (pp. 244-259). Cham: Springer International Publishing. doi:10.1007/978-3-319-11915-1_16

Ferrara, E., Meo, P. D., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques. *Knowledge-Based Systems*, *70*(C), 301–323. doi:10.1016/j.knosys.2014.07.007

Gangemi, A. (2013). A Comparison of Knowledge Extraction Tools for the Semantic Web. In P. Cimiano, O. Corcho, V. Presutti, L. Hollink, & S. Rudolph (Eds.), *The Semantic Web: Semantics and Big Data: 10th International Conference, ESWC 2013*, *Montpellier, France*, *May 26-30*. *Proceedings* (pp. 351-366). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-38288-8_24

Gerber, D., Hellmann, S., Bühmann, L., Soru, T., Usbeck, R., & Ngomo, A.-C. N. (2013). *Real-time rdf extraction from unstructured data streams. In The Semantic Web–ISWC 2013* (pp. 135–150). Springer.

Gracia, J., Montiel-Ponsoda, E., Cimiano, P., Gómez-Pérez, A., Buitelaar, P., & McCrae, J. (2012). Challenges for the multilingual web of data. *Journal of Web Semantics*, *11*, 63–71. doi:10.1016/j.websem.2011.09.001

Haag, F., Lohmann, S., Siek, S., & Ertl, T. (2015). Visual querying of linked data with QueryVOWL. *Paper presented at the Joint Proceedings of SumPre 2015 and HSWI 2014-15*. CEUR-WS.org, Portoroz, Slovenia (2015).

Hakimov, S., Jebbara, S., & Cimiano, P. (2017). *AMUSE: Multilingual Semantic Parsing for Question Answering over Linked Data. Paper presented at the Proceedings of the 16th International Semantic Web Conference (ISWC 2017)*. doi:10.1007/978-3-319-68288-4_20

Hausenblas, M. (2009). Exploiting linked data to build web applications. *IEEE Internet Computing*, *13*(4), 80–85. doi:10.1109/MIC.2009.79

He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems. *Expert Systems with Applications*, *56*(C), 9–27. doi:10.1016/j.eswa.2016.02.013

Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, *1*(1), 1–136. doi:10.2200/S00334ED1V01Y201102WBE001

Heeks, R., & Bailur, S. (2007). Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice. *Government Information Quarterly*, *24*(2), 243–265. doi:10.1016/j.giq.2006.06.005

Hitzler, P., & Janowicz, K. (2013). Linked Data, Big Data, and the 4th Paradigm. *Semantic Web*, *4*(3), 233–235.

Höffner, K., Walter, S., Marx, E., Usbeck, R., Lehmann, J., & Ngonga Ngomo, A.-C. (2016). Survey on challenges of Question Answering in the Semantic Web.

Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., & Decker, S. (2011). Searching and browsing Linked Data with SWSE: The Semantic Web Search Engine. *Journal of Web Semantics*, *9*(4), 365–401. doi:10.1016/j.websem.2011.06.004

Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A., & Decker, S. (2012). An empirical survey of Linked Data conformance. *Journal of Web Semantics*, *14*(Suppl. C), 14–44. doi:10.1016/j.websem.2012.02.001

Huang, R., Lai, T., & Zhou, L. (2017). Proposing a framework of barriers to opening government data in China: A critical literature review. *Library Hi Tech*, *35*(3), 421–438. doi:10.1108/LHT-01-2017-0014

Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., & Sheth, A. P. (2010). *Linked Data Is Merely More Data*. *Paper presented at the AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*, Palo Alto, California.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, *32*(4), 363–368. doi:10.1016/j.giq.2015.11.007

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). *Big data: Issues and challenges moving forward.* Paper presented at the 2013 46th Hawaii International Conference on System Sciences (HICSS). doi:10.1109/HICSS.2013.645

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). *On publishing linked open government data. Paper presented at the 17th Panhellenic Conference on Informatics*. doi:10.1145/2491845.2491869

Khan, M., Khusro, S., & Ullah, I. (2017). On Linked Open Data (LOD)-based Semantic Video Annotation Systems. Pakistan Academy of Sciences, 54(1), 1-12.

Khusro, S., Khan, M., & Ullah, I. (2016). Collaborative Video Annotation Based on Ontological Themes, Temporal Duration and Pointing Regions. *Paper presented at the Proceedings of the 10th International Conference on Informatics and Systems*, Giza, Egypt. doi:10.1145/2908446.2908471

Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, *19*(2), 267–283. doi:10.1007/s10796-016-9686-2

Klímek, J., Helmich, J., & Necaský, M. (2015). Use cases for linked data visualization model. *Paper presented at the Workshop on Linked Data on the Web (LDOW)*.

Krempl, G., Žliobaite, I., Brzeziński, D., Hüllermeier, E., Last, M., Lemaire, V., & Stefanowski, J. et al. (2014). Open challenges for data stream mining research. *SIGKDD Explorations*, *16*(1). doi:10.1145/2674026.2674028

Kunaver, M., & Porl, T. (2017). Diversity in recommender systems A survey. *Knowledge-Based Systems*, *123*(C), 154–162. doi:10.1016/j.knosys.2017.02.009

Latif, A., Scherp, A., & Tochtermann, K. (2016). LOD for Library Science: Benefits of Applying Linked Open Data in the Digital Library Setting. *KI - Künstliche Intelligenz, 30*(2), 149-157. doi:10.1007/s13218-015-0420-x

Le-Phuoc, D., Dao-Tran, M., Parreira, J. X., & Hauswirth, M. (2011). A native and adaptive approach for unified processing of linked streams and linked data. *Paper presented at the Proceedings of the 10th international conference on The semantic web*. doi:10.1007/978-3-642-25073-6_24

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., & Auer, S. et al. (2015). DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, *6*(2), 167–195.

Llanes, K. R., Casanova, M. A., & Lemus, N. M. (2017). From Sensor Data Streams to Linked Streaming Data: A survey of main approaches. *Journal of Information and Data Management*, *7*(2), 130–140.

Mendes, P. N., Mühleisen, H., & Bizer, C. (2012). Sieve: linked data quality assessment and fusion. *Paper presented at the 2012 Joint EDBT/ICDT Workshops*. doi:10.1145/2320765.2320803

Meusel, R., Spahiu, B., Bizer, C., & Paulheim, H. (2015). Towards automatic topical classification of lod datasets. *Paper presented at the 24th International Conference on World Wide Web, LDOW Workshop*.

Millard, I. C., Glaser, H., Salvadores, M., & Shadbolt, N. (2010). Consuming multiple linked data sources: Challenges and Experiences. *Paper presented at the First International Conference on Consuming Linked Data*.

Moura, T. H., Davis, C. A. Jr, & Fonseca, F. T. (2017). Reference data enhancement for geographic information retrieval using linked data. *Transactions in GIS*, *21*(4), 683–700. doi:10.1111/tgis.12238

Mouzakitis, S., Papaspyros, D., Petychakis, M., Koussouris, S., Zafeiropoulos, A., Fotopoulou, E., & Psarras, J. et al. (2017). Challenges and opportunities in renovating public sector information by enabling linked data and analytics. *Information Systems Frontiers*, *19*(2), 321–336. doi:10.1007/s10796-016-9687-1

Nentwig, M., Hartung, M., Ngonga Ngomo, A.-C., & Rahm, E. (2017). A survey of current link discovery frameworks. *Semantic Web*, *8*(3), 419–436. doi:10.3233/SW-150210

Nixon, L., & Troncy, R. (2014). Survey of semantic media annotation tools for the web: towards new media applications with linked media. *Paper presented at the European Semantic Web Conference*. doi:10.1007/978-3-319-11955-7_9

Noia, T. D., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012). Linked open data to support content-based recommender systems. *Paper presented at the Proceedings of the 8th International Conference on Semantic Systems*, Graz, Austria.

Oliveira, J., Delgado, C., & Assaife, A. C. (2017). A recommendation approach for consuming linked open data. *Expert Systems with Applications*, *72*(Suppl. C), 407–420. doi:10.1016/j.eswa.2016.10.037

Otero-Cerdeira, L., Rodríguez-Martínez, F. J., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, *42*(2), 949–971. doi:10.1016/j.eswa.2014.08.032

Peinl, R. (2016). Semantic Web: State of the Art and Adoption in Corporations. *KI-Künstliche Intelligenz*.

Peng, S., Wang, G., & Xie, D. (2017). Social influence analysis in social networking big data: Opportunities and challenges. *IEEE Network*, *31*(1), 11–17. doi:10.1109/MNET.2016.1500104NM

Peng, Y., Huang, X., & Zhao, Y. (2017). An Overview of Cross-media Retrieval: Concepts, Methodologies, Benchmarks and Challenges. *IEEE Transactions on Circuits and Systems for Video Technology*, 1. doi:10.1109/TCSVT.2017.2705068

Peska, L., & Vojtas, P. (2015). Using Linked Open Data in Recommender Systems. *Paper presented at the 5th International Conference on Web Intelligence, Mining and Semantics*. doi:10.1145/2797115.2797128

Purohit, S., Smith, W., Chappell, A., West, P., Lee, B., Stephan, E., & Fox, P. (2016). Effective Tooling for Linked Data Publishing in Scientific Research. *Paper presented at the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*. doi:10.1109/ICSC.2016.87

Radulovic, F., Mihindukulasooriya, N., García-Castro, R., & Gómez-Pérez, A. (2017). A comprehensive quality model for linked data. *Semantic Web*.

Ristoski, P., & Paulheim, H. (2016). Semantic Web in data mining and knowledge discovery: A comprehensive survey. *Journal of Web Semantics*, *36*. doi:10.1016/j.websem.2016.01.001

Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B., & Granitzer, M. (2014). *Discovery and visual analysis of linked data for humans. In The Semantic Web–ISWC 2014* (pp. 309–324). Springer.

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, *22*(1). doi:10.3233/IP-160012

Scheider, S., Degbelo, A., Lemmens, R., van Elzakker, C., Zimmerhof, P., Kostic, N., & Banhatti, G. et al. (2015). Exploratory querying of SPARQL endpoints in space and time. *Semantic Web*.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., & schraefel, . (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, *27*(3), 16–24. doi:10.1109/MIS.2012.23

Shekarpour, S., Endris, K. M., Kumar, A. J., Lukovnikov, D., Singh, K., Thakkar, H., & Lange, C. (2016). Question Answering on Linked Data: Challenges and Future Directions. *Paper presented at the Proceedings of the 25th International Conference Companion on World Wide Web*, Montréal, Québec, Canada. doi:10.1145/2872518.2890571

Shekarpour, S., Ngomo, A.-C. N., & Auer, S. (2013). Question Answering on Interlinked Data. *Paper presented at the 22nd International Conference on World Wide Web*, Rio de Janeiro, Brazil. doi:10.1145/2488388.2488488

Sheth, A., & Kapanipathi, P. (2016). Semantic filtering for social data. *IEEE Internet Computing*, *20*(4), 74–78. doi:10.1109/MIC.2016.86

Shvaiko, P., & Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, *25*(1), 158–176. doi:10.1109/TKDE.2011.253

Sikos, L. F. (2017). RDF-powered semantic video annotation tools with concept mapping to Linked Data for next-generation video indexing: A comprehensive review. *Multimedia Tools and Applications*, *76*(12), 14437–14460. doi:10.1007/s11042-016-3705-7

Sikos, L. F., & Powers, D. M. (2015). Knowledge-driven video information retrieval with LOD: from semi-structured to structured video metadata. *Paper presented at the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval*. doi:10.1145/2810133.2810141

Thellmann, K., Galkin, M., Orlandi, F., & Auer, S. (2015). LinkDaViz – Automatic Binding of Linked Data to Visualizations. In M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, & S. Staab (Eds.), *The Semantic Web - ISWC 2015: 14th International Semantic Web Conference*, *Bethlehem, PA*, *October 11-15*, *Proceedings, Part I* (pp. 147-162). Cham: Springer International Publishing. doi:10.1007/978-3-319-25007-6_9

Thompson, N., Ravindran, R., & Nicosia, S. (2015). Government data does not mean data governance: Lessons learned from a public sector application audit. *Government Information Quarterly*, *32*(3), 316–322. doi:10.1016/j.giq.2015.05.001

Tonon, A., Catasta, M., Demartini, G., & Cudré-Mauroux, P. (2015). Fixing the Domain and Range of Properties in Linked Data by Context Disambiguation. *Paper presented at the WWW 2015 Workshop: Linked Data on the Web (LDOW 2015)*.

Usbeck, R., Ngomo, A.-C. N., Bühmann, L., & Unger, C. (2015). *HAWK–Hybrid Question Answering Using Linked Data. In The Semantic Web. Latest Advances and New Domains* (pp. 353–368). Springer. doi:10.1007/978-3-319-18818-8_22

Villazón-Terrazas, B., Vilches-Blázquez, L. M., Corcho, O., & Gómez-Pérez, A. (2011). *Methodological guidelines for publishing government linked data. In Linking government data* (pp. 27–49). Springer. doi:10.1007/978-1-4614-1767-5_2

Wang, H., Xu, Z., Fujita, H., & Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences*, *367*, 747–765. doi:10.1016/j.ins.2016.07.007

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: An empirical view from the Citizen's perspective. *Information Systems Frontiers*, *19*(2), 285–300. doi:10.1007/s10796-016-9679-1

Yahya, M., Berberich, K., Elbassuoni, S., Ramanath, M., Tresp, V., & Weikum, G. (2012). Natural language questions for the web of data. *Paper presented at the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Yang, H.-C., & Hsu, C.-C. (2015). Semantic Recommendation Using Linked Open Data. *Paper presented at the ASE Big Data & Social Informatics 2015*.

Ye, J., Dasiopoulou, S., Stevenson, G., Meditskos, G., Kontopoulos, E., Kompatsiaris, I., & Dobson, S. (2015). Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive and Mobile Computing*, *23*, 1–25. doi:10.1016/j.pmcj.2014.12.009

Yu, L., & Liu, Y. (2015). Using Linked Data in a heterogeneous Sensor Web: Challenges, experiments and lessons learned. *International Journal of Digital Earth*, *8*(1), 17–37. doi:10.1080/17538947.2013.839007

Zavattaro, S. M., & Sementelli, A. J. (2014). A critical examination of social media adoption in government: Introducing omnipresence. *Government Information Quarterly*, *31*(2), 257–264. doi:10.1016/j.giq.2013.10.007

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, *7*(1), 63–93. doi:10.3233/SW-150175

Zhang, Z., Gentile, A. L., Blomqvist, E., Augenstein, I., & Ciravegna, F. (2017). An unsupervised data-driven method to discover equivalent relations in large Linked Datasets. *Semantic Web*, *8*(2), 197–223. doi:10.3233/SW-150193

Zou, L., Huang, R., Wang, H., Yu, J. X., He, W., & Zhao, D. (2014). Natural language question answering over RDF: a graph data driven approach. *Paper presented at the 2014 ACM SIGMOD international conference on Management of data*. doi:10.1145/2588555.2610525

Zuiderwijk, A., Janssen, M., & Susha, I. (2016). Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 116–146. doi:10.1080/10919392.2015.1125180

ENDNOTES

1 http://annomation.open.ac.uk/
2 http://annotorious.github.io/
3 http://www.imagesnippets.com/
4 http://www.openvideoannotation.org/
5 http://linkedtv.eu/ontology

Nosheen Fayyaz is Lecturer in Department of Computer Science at University of Peshawar, Pakistan. She has completed her Master degree in 2000 from Department of Computer Science, University of Peshawar and MS in Computer Science from Institute of Business and Management Sciences, The University of Agriculture Peshawar in 2010. She has been teaching for 15 years. Currently she is enrolled as a PhD scholar in University of Peshawar. Her research interests are Web Semantics, Web Engineering, Cross Domains, Ontologies and Linked Open Data.

Irfan Ullah has received his MS(CS) degree in Web Engineering from the Department of Computer Science, University of Peshawar, Pakistan. He is currently pursuing his PhD in Computer Science specializing in the area of Web Semantics at Department of Computer Science University of Peshawar. He is currently working as Assistant Professor in the Department of Computer Science, Shaheed Benazir Bhutto University, Sheringal, Pakistan. He has more than 6 years of research experience and is the author of several research papers published in national and international journals and conferences. His research interests include Semantic Web, Linked Open Data, Information Retrieval, Web Engineering, Ontology Engineering, and Digital Libraries.

Shah Khusro is working as Professor of Computer Science and Head of Information and Web Semantics Research Group at the Department of Computer Science, University of Peshawar, Peshawar, Pakistan. He has done his PhD from the Institute of Software Technology & Interactive Systems, Vienna University of Technology, Vienna, Austria under the supervision of Prof. Dr. A Min Tjoa. Earlier he did his M.Sc. from the Department of Computer Science, University of Peshawar with a Gold Medal. He is a member of different academic bodies of several universities in the region. He has attended/organized several international conferences. He is working on some real-world projects in eGovernance, Health, and for Blind and Elderly People. His research interests include Web Semantics, Web Engineering, Information Retrieval, Web Mining, Search Engines, Augmented Reality, Mobile based Systems for People with Special Needs, etc.