Security Aspects in Web of Data Based on Trust Principles. A brief of Literature Review

Jhon Francined Herrera-Cubides[1], Paulo Alonso Gaona-García[1], Carlos Montenegro-Marín[1], Diego Cataño[1] and Rubén González-Crespo[2]

[1]Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.
[2]Universidad Internacional de la Rioja, Logroño, España.

***Abstract*:** Within scientific community, there is a certain consensus to define "Big Data" as a global set, through a complex integration that embraces several dimensions from using of research data, Open Data, Linked Data, Social Network Data, etc. These data are scattered in different sources, which suppose a mix that respond to diverse philosophies, great diversity of structures, different denominations, etc. Its management faces great technological and methodological challenges: The discovery and selection of data, its extraction and final processing, preservation, visualization, access possibility, greater or lesser structuring, between other aspects, which allow showing a huge domain of study at the level of analysis and implementation in different knowledge domains. However, given the data availability and its possible opening: What problems do the data opening face? This paper shows a literature review about these security aspects.

***Keywords*:** Security Aspects, Literature Review, Principle of Trust, Linked Open Data**.**

1. Introduction

According to [6], it seems clear that personal information, national security information and data affected by intellectual property issues must be protected and might be out the scope of data opening. Even though, in an ecosystem of open data and probably linked data, total protection is hard to reach. Successive recombination of data might show sensitive information, so it is difficult to data owners to foresee the consequences of their publications. In addition, once data is open, possibilities of closing data are lesser. That is way, you should use what some authors call "negative freedom", which is the establishment of clear rules of what must and must not be opened. Even so, it is possible to access to this data information as long as the information can be disintegrated [18].

On the one hand, organizations take advantage of Open Data in order to generate new business opportunities and supporting research. However, the unfair use of these data by other organizations, thus generating the non-opening of the data. Additionally, the lack of data transparency may compromise the supplier company image, generating a serious reputation crisis if the consumer catalogs the opening of the information as of low quality [57].

According to these referents, next literature review focus on made an analysis on actual state about security mechanisms principles applied to Linked Open Data. More specific in security solutions under the data Trust Principles, focused mainly on issues of quality, authentication and access control. To achieve that, in section 2 it is described a theoretical background where it is explained the LOD approach, the related works where particularly are mentioned studies about accessibility, origin, quality and its related concepts of data reliability and integrity. In section 3 the criteria for searching and the methodology with which the systematic review on the research topic was carried out are presented. Section 4 presents the results of the methodology and the criteria used to classify the found information, in addition, it is included a discussion section about the research questions. Finally, the conclusions and the future work are described.

2. Background

In order to contextualize the literature review, this paper will address aspects related with the panorama offered by Link Open Data, its principles and its security approaches applied to the Semantic Web.

2.1 Linked Open Data Approach

Semantic Web technologies aim to simplify the distribution, sharing and exploitation of information and knowledge, across multiple distributed actors on the Web. As with all technologies that manipulate information, there are privacy and security implications, and data policies (e.g., licenses and regulations) that may apply to both data and software artifacts. Additionally, semantic web technologies could contribute to the more intelligent and flexible handling of privacy, security and policy issues, through supporting information integration and sense-making [53]. From this approach, the vision of what Semantic Web supposes is shifted from strong problems such as logical data inference to simpler but fundamental problems for practical Web development, such as data exchange and integration. To achieve this, LOD community has resumed the Web Semantic proposed standards, and they have been adapted them to the needs of a more generic Web development, for instance, offering simple socialization of Resource Description Framework – RDF formats [88] – based on JSON [80] or HTML [58], with the aim at being a realistic option for Web developers. At the same time, LOD community has proposed new standards like WebId [40] [52], in order to solve other basic problems in data integration in Web applications such as authentication.

Either adapting Semantic Web technologies or proposing new standards, LOD community has always tried to follow REST principles of Web Architecture, widely accepted among for Web developers, including practices like content negotiation [41] or clarifying the difference between information resources and non-information resources [13]. From LOD point of view, these proposals are just the extension of REST architectural principles at data interchange, reusing the work done by the Semantic Web community, instead of proposing new solutions from scratch. [30].

The WWW suffer from a similar problem, owing to most HTML web pages are not completely readable. Nonetheless, there are significant differences. Firstly, a big effort in development has done Web browsers more robust against misuse of HTML. Secondly, traditional Web of Document is aimed at human reader. Even if web pages were designed free of errors, human agent could process some of the page contents. However, in Semantic Web, agents have less intelligence than the human. For that reason, even a small syntax error breaks its browsing and processing capabilities [8].

To mitigate this problem, Semantic Web community has formulated a collection of guides about best practices documents. It also has had a strong approach in education through courses, manuals, summer courses and tutorials. The main problem of these approaches is that all of them lead the human data and consumer editor. As a result, the success of these approaches depends fundamentally on the willingness, ability and capability of a number of human beings to do the right thing [8].

2.2 Access Control Models and Standard

Access control refers to the model, which is used to guide the access control process. The decision to grant or deny access is based on two distinct processes, authentication and authorization. Authentication involves the verification of credentials (you are who you say you are). Whereas, authorization is the process of granting or denying access to system resources based on credentials. Some access control models, and relevant standardization efforts are described below [74]-[45]-[51]-[42]-[100].

- ACL - Access Control List, specifies the level of permission granted to a user of an application.
- MAC - Mandatory Access Control, this security policy is centrally controlled by a security policy administrator.
- DAC - Discretionary Access Control, the controls are discretionary in the sense that a subject with a certain access permission is capable of passing that permission (perhaps indirectly) on to any other subject.
- RBAC - Role Based Access Control, C provides a valuable level of abstraction to promote security administration at a business enterprise level rather than at the user identity level
- VBAC - View Based Access Control, is a complementary access control model which grants access to sets of entities, logically structured as views.
- ABAC - Attribute Based Access Control, is a general framework which combines the benefits of DAC, MAC and RBAC and goes beyond their limitations. The model is based on generic attributes which are used to capture identities and access control lists for DAC, clearances and classifications for MAC and roles for RBAC.
- CBAC - Context Based Access Control, uses properties, pertaining to users, resources and the environment, to grant/deny access to resources.

Regarding the access control standards, the main ones are presented below.

- XACML - The eXtensible Access Control Markup Language, is used to represent attribute-based access control policies.
- WebID - Web Identity and Discovery, is a mechanism used to uniquely identify and authenticate a person, company, organization or other entity, by means of a Uniform Resource Identifier (URI).
- WAC – Web Access Control, demonstrates how together WebID and access control policies specified using the WAC vocabulary, can be used to enforce distributed access control.
- P3P - Platform for Privacy Preferences, enables websites to express their privacy preferences in a machine-readable format.
- ODRL - The Open Digital Rights Language, is used to define rights to or to limit access to digital resources.

According to [34], there has been a large amount of previous research on the security of the Semantic Web, yet none of it looks at the security properties of the Semantic Web infrastructure itself. In general, there have been three streams of research on security on the Semantic Web:

- Semantic Web policy languages for access control,
- Semantic Web ontologies for cybersecurity, and
- Problems with privacy in publishing Semantic Web data.

Researches such [24], [83], [53] and [22] have shown those streams.

2.3 Security and Principle of Trust in Semantic Web

2.3.1. Principle of Trust

Taking into account that Trust is based on experience, and by tracking and propagating trust among web sources in a similar fashion to the way trust is created and maintained in a human community, trust can be established on the Semantic Web [69]. Some researches offer different general definitions of Trust. [3], and [29] offer definitions from about Trust:

- "[Trust is] a subjective expectation an agent has about another's future behavior based on the history of their encounters", from (Mui et al., 2002) cited by [3], it refers to past encounters, and may be thought of by some as "reputation-based" trust.
- "[Trust is] the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context", from (Grandison and Sloman, 2000) cited by [3], it introduces context and is unique in referring to the "competence" to act (instead of actions, themselves).
- "Trust of a party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period within a specified context (in relation to service X).", from (Olmedilla et al., 2005) cited by [3], it applies to many cases in this survey, and it refers to actions and not competence like the previous definition.
- A data consumer C is the endpoint, an individual or some specific system, which consumes the data. The trustworthiness of a piece of data d for a data consumer C is the measurable belief of C to represent the reliability of d within a specific context [29].

According with [82], from computing point of view, trust is modeled after human relationships, for that reason it is strongly associated with security. So, generically the concept of trust (and security) may be applied to other domains, for example, a party may "trust" another party to deliver secure quality service, in which case trust becomes a measure of the "security" of service-availability. In other to identify the

differences between Security, Privacy and Trust, these concepts are shown in Table 1.

Table 1. Differences between Security, Privacy and Trust. Source: [2]

Attributes	Meanings/Definitions
Security	"Preservation of confidentiality, integrity and availability of information; in addition, other properties such as authenticity, accountability, non-repudiation and reliability can also be involved." (ISO 27001)
Privacy	In the commercial, consumer context, privacy entails the protection and appropriate use of the personal information of customers, and the meeting of expectations of customers about its use. For organizations, privacy entails the application of laws, policies, standards and processes by which personal information is managed.
Trust	Trust is a very sensitive notion. In general, trust is referred to as "levels of confidence in something or someone". That is why we can view trust in something as the customers' level of confidence in using something. Trust revolves around 'assurance' and confidence that people, data, entities, information or processes will function or behave in expected ways.

Briefly, according to [55], Web of Trust can be defined as "You tell the system who you want to trust". In other words, the reliability and usefulness of Web data depends on evaluating its trustworthiness, the subjective measure of the belief which a user has that the data is "true" [43]. On the other hand, in order to be more specific about trust semantics, Grandison & Sloman's classification (2000), cited by [44], distinguish between a set of different trust classes:

- Provision trust that describes the relying party's trust in a service or resource provider.
- Access trust that describes trust in principals for the purpose of accessing resources owned by or under the responsibility of the relying party.
- Delegation trust that describes trust in an agent (the delegate) that acts and makes decision on behalf of the relying party.
- Identity trust that describes the belief that an agent identity is as claimed.
- Context trust that describes the extent to which the relying party believes that the necessary systems and institutions are in place in order to support the transaction and provide a safety net in case something should go wrong

As can be identified in the literature, Privacy, Trust, Security and Provenance have worked as key aspects of the Semantic Web in order to enable more trustworthy data exchange.

2.3.2. Security in Semantic Web

The Web 3.0 makes possible that users can access data from other users stored anywhere around the world, usually in a free and open way. Under this communication model, it is even more important to consider security aspects such as quality, confidentiality, and data control access, especially when it is worked over sensible content. In the proposals made by [31], it is highlighted that one of the biggest linked data challenges is the privacy and security of published data. In a global social network, which is distributed as well, it is required that each person can control his or her identity, that this identity is linkable between sites and it is able to be authenticated globally. With a distributed authentication is easier that each person protects their resources and defines their privacy.

The reference [71] made a research of fine-grained access control mechanisms to restrict access to specific-structured data to particular users. This research proposes the design of a lightweight vocabulary called "Privacy preference ontology PPO", which, basically collaboration and link among data creators to describe the preferences of fine-grained privacy and restrict the access to specific data that can have common preferences.

The reference [87] introduces the concept of light-vocabulary. The present a solution called "Social Semantic SPARQL security for access control – S4AC", which allows the definition of fine-grained control access policies formalized in SPARQL, language used primarily to make queries over graphs in Linked Data. In particular, it represents a model of access control with the purpose of establishing policies to restrict the access to specific data RDF [88] based on social and contextual information tags.

The reference [34] argue that the Semantic Web was designed without any security considerations. Still, today there is almost no academic work on security in terms of the Semantic Web. Rather unfortunately, there also seems to be considerable confusion about security within the Semantic Web research community, ranging from ignorance of the security problems in HTTP URIs to misuse of TLS in WebID+TLS.

The reference [46] consider that a) Confidentiality of data restricts the data access to authorized parties only; b) integrity means that the data can only be modified by authorized parties; and a) availability, availability states that the data must always be accessible when requested. He argues that topics above are often listed as the three main requirements for achieving data security, owing to these requirements are especially important in open and distributed networks. Such networks are able to store large amounts of data without having a single entity in control of ensuring the data's security. The Semantic Web applies to these characteristics as well as it aims at creating a global and decentralized network of machine-readable data. Ensuring the confidentiality, integrity, and availability of this data is therefore also important and must be achieved by corresponding security mechanisms. However, the current reference architecture of the Semantic Web does not define any particular security mechanism, yet which implements these requirements. Instead, it only contains a rather abstract representation of security.

Thus, there have been developed control access models applicable to semantic data defined around metadata at the document level, establishing security policies for each document. In the Web domain, security policies are a set of rules that define security requirements for access and modification of a document or dataset [48]. These rules specify a set of credentials to gain access with resources. Those credentials might be either authentications based on user and password or a set of characteristics that agents must meet to access resources. With this approach, there have been created several ontologies and languages in order to represent those Web Semantic policies (making possible to reason over that metadata), within the most used methods, it could be mentioned:

- Web Access Control Vocabulary (ACL): According to [19], ACL is a vocabulary to represent access control lists over Web Access Control (WAC), a decentralized system of the establishment of access resources permissions, organized around users and groups identified by URIs HTTP or WebIDs [33]. In this system, set of users hosted in any host are identified by the URI of a user's class, which can be made searches to retrieve all users that belong to a specific class. In this way, resources access can be filtered by users, either extern or intern, to the host in which the resource is hosted. The ontology specifies the class AgentClass, which defines users' groups represented by foaf:Agent [84]. Furthermore, it can represent ways of access to resources (acl:Read, acl:Write, acl:Append, acl:Control). This is how it can build triplets organized in access lists that establish permissions based on FOAF profiles.
- Privacy Preference Ontology (PPO): It is a mechanism proposed by [93] that defines a light ontology based on WAC. This mechanism allows creation of complex security preferences based on a set of attributes that one user must comply in order to access a resource. This ontology defines the main class named PrivacyPreference, and a set of properties to specify the resource to be protected, the requirements to be met and the different access privileges. The requirements that the solicitor must comply to access the resource might be: To belong to an ontological class, having certain property or value or being linked with other specific resources.
- In addition, WebID Incubator Group [15] – W3C research group, has developed FOAF+SSL [78]-[61], which is a safe authentication protocol that allows using of this kind of access control ontologies. It is a one-connection authentication system that uses the SSL layer [48, 99] virtually built in every web browser that uses HTTPS. It is based on WebID [61], a Web identification system based on URIs. FOAF+SSL uses an architecture based on public keys infrastructure (PKI) standards, using certificates X.509 that contain WebID of FOAF profiles.

Using these ontologies, languages and authentication protocols, it has been created several security models for the Semantic Web, such as Policy Enabled Linked Data Server – PeLDS [61], a triplet's RDF storage system based on SWRL rules [35] and FOAF+SSL authentication protocols.

In this context, the section below shows the used methodology to carry out a searching, analysis and classification of the literature.

3. Background

The reviewed literature was selected by a deep analysis, through a systematic review methodology according to the recommendation defined by [47], [49] and [50]. In the following sections, it will be defined the criteria and aspects to be considered for the study and how were made the searches through the databases.

(a) Study Criteria

In order to carry out this review, the following aspects related to the analysis were considered:

- Consequences of the opening data
- Security panorama of LOD
- Reliability and quality of LOD resources.

The documents include book sections, indexed magazine papers and conferences published in electronic format.

(b) Research questions

To guide this review, the next research questions are raised, they will be answered along the literature review:

- What are the security mechanisms most used in LOD?
- What security methods are used on LOD?
- What are the main Linked Data challenges based on Trust Principles?

(c) Searching on bibliography databases

The searching process consists in make research of literature that allows finding related studies with the main topic of research, in this case, is: "Security mechanisms under the Web of Trust principles on LOD resources". Additionally, the research is limited using a time interval between 1998 and 2016, to identify which has been the current view of the research topic, to define the last used methodologies in the problem solution and corroborate that the problem is of current interest. In order to identify the complete research papers, it was considered the IEEE Explorer, ACM Digital Library, Scopus, Engineering Village and Springer databases. The review process was carried out from September 2015 to September 2016. The common search keywords used in the review process are described below:

- "Linked Open Data" AND "Security Mechanisms."
- "Linked Open Data" AND "Web of Trust."
- "Linked Open Data Security" AND "Semantic Web Security Problems."
- "Linked Open Data Security Strategies"
- "Linked Open Data Problems"
- "Linked Open Data Quality"

To check the literature review, the results were filtered through the following criteria:

- The keywords
- The titles and summaries
- The evaluation methods
- The experiment results

Based on these criteria, 70 primary studies were founded and then a corpus of 59 papers were selected as they were closely related to the study topic. The other 11 were discarded because they do not offer a considerable contribution to the research topic. To carry out the classification, three groups were defined: First one, the impact on the opening of LOD, second one, the quality of data emphasizing the information trust and quality levels, and the third one, the security strategies of data under LOD principles.

4. Result Analysis

From the method described earlier, it will be presented the following classifications:

a. Documentary classification: Regarding to the selected document classification, it was identified the quantity of selected productions related to the year in which they were published, also, what is the source consulted in which more publications were found and finally which was the most selected kind of publication.

b. Topic classification: Classification regarding the relevant topics related to the security strategies (quality, reliability, and access control), followed by a ranking of most cited papers.

Finally, a summary of the literature was found which was done by topic. In this part, it shows which were the tools and methods found related to the security in LOD.

4.1 Documentary Classification

a. Publication year.

In figure 1, it is shown that most of the selected paper production for the study was made between 2013 and 2015. In [12] describes the measurement of the growing of LOD between 2009 and 2014, from the quantity of triplets and linked Datasets, where the most predominant tendency targets the increasing of Social Web publications, a decrease in independent data publications and a percentage stable tendency of open data published by the government. Not surprisingly the Social Web publications are increasing. For this study, it is clear that the research and the development of new LOD tools are increasing.

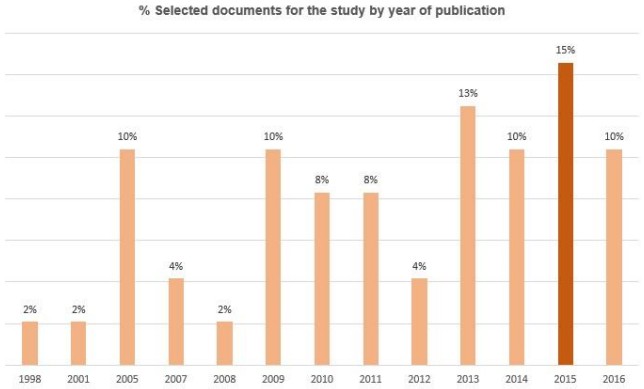

Figure 1. Selected documents for the study by year of publication. Source: Authors.

b. Data sources.

In figure 2 it can be seen that around 60% of documents that contribute to this review were found in the publication sources ACM Library, IEEE and Springer International Publishing. In the earliest searches, it was clear this tendency, due to that fact, our research was approached deeper in those sources.

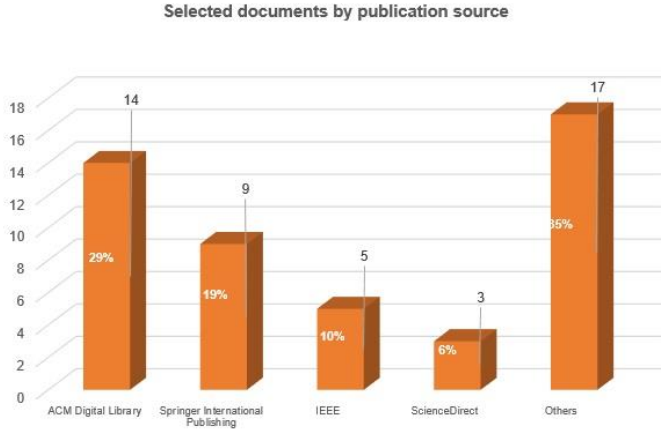

Figure 2. Selected documents by publication source. Source: Authors.

c. Type of research.

Figure 3 shows the publication resources grouped by type and participation percentage regarding to the type of found publication. The type of publication where most information about the topic was found was the Journal paper type with a 67% of the selected relevant information, which indicates that future searches about the topic should orient to this type of publication, because, among other questions, it can be easier accessed in comparison to another kind like conferences proceedings or sections books, this validates the tendency showed in figure 3.

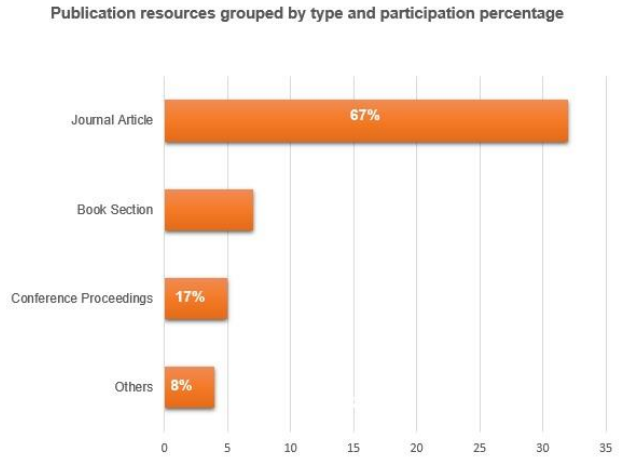

Figure 3. *Publication resources grouped by type and participation percentage. Source: Authors*

d. Document catalogue

Table 2 shows the top 10 of selected documents and their relation to the research topic. Some of them have not been cited yet, but they have a great contribution to this publication. (The complete list of papers is in Appendix No. 2).

In Table 2, the relevance of the security topic in LOD it can be identified in most of the documents. It shows the classification items for this review, related to the basic security information principles such as:

a. Trust understood as the provenance, data reliability and reputation.
b. Quality seen as data integrity.
c. Access control where studies are found, and security techniques based on policies, permissions, certificates and protocols to the access in LOD data are discussed.

Table 2. Selected documents and their relation to the research topic. Source: Authors

Classification	Author	Title	Publication resource	Source	Search criteria	Citations	Year
Quality	Thakkar, H., Endris, K., Gimenez-Garcia, J.	Are Linked Datasets fit for Open-domain Question Answering? A Quality Assessment	Journal Paper	ACM Digital Library	Keywords	2	2016
Trust	Musto, C., Narducci, F., Lops, P., Gemmis, M. De, & Semeraro, G.	ExpLOD: a framework for Explaining Recommendations based on the Linked Open Data cloud	Journal Paper	ACM Digital Library	Keywords	0	2016
Quality	Cheniki, N., Belkhir, A., Sam, Y., & Messai, N.	LODS: A Linked Open Data Based Similarity Measure	Journal Paper	IEEE	Keywords	0	2016
Quality	Piao, G., & Breslin, J. G.	Measuring Semantic Distance for Linked Open Data-enabled Recommender Systems.	Journal Paper	ACM Digital Library	Keywords	6	2016
Quality	Beek, W., Rietveld, L., Schlobach, S., and Harmelen, F.	Why the Semantic Web Needs Centralization	Journal Paper	IEEE	Keywords	2	2016
Trust	Sohn, M., Jeong, S., Kim, J., Lee, H.	Augmented context-based recommendation service framework using knowledge over the Linked Open Data cloud	Journal Paper	ScienceDirect	Keywords	0	2015
Quality	Singh, M. P.	Norms as a basis for governing sociotechnical systems	Journal Paper	ACM Digital Library	Keywords	61	2015
Quality	Behkamal, B., Kahani, M., & Bagheri, E	Quality Metrics for Linked Open Data	Book Section	Springer International Publishing	Keywords	0	2015
Quality	Yang, H.-C., & Hsu, C.-C.	Semantic Recommendation Using Linked Open Data	Journal Paper	ACM Digital Library	Keywords	0	2015
Quality	Dividino, R., Gottron, T., & Scherp, A.	Strategies for Efficiently Keeping Local Linked Open Data Caches Up-To-Date	Book Section	Springer International Publishing	Keywords	0	2015

4.2 Topic Classification

Authors like [6], identify data quality concept which is defined such as the ability to be used in a determined context, and that is characterized by its integrity, semantic representation sufficiency or the data "readability" degree. On the other hand, data that are incomplete, inaccurate, and inconsistent in the representation or have an invalid syntax, are not considered as quality data. This quality problems are tested with data validation systems and are controlled in an ongoing task that embraces the whole process. The control process should verify even the no-publication of data, the links review or the migration of technologies, etc. from this earlier definition, we might infer that access control has a dependent and directly proportional relation with data quality and it is fundamental for any safe system because without the adequate data construction structure it is difficult to establish reliable authentication processes.

According to [85], Linked Data consumption has two key quality factors: the reliability and the provenance. These factors allow users to trust in an environment where everyone can publish data, any query might find contrasting responses from different databases. Whereas the trust provides a reliability measure (general or customized) that a user might expect from the data, the provenance allows join data with its author and its creation process.

In their research, [25] describe three simple indicators to evaluate the provenance:

a. Basic provenance: This indicator checks if a Dataset has at least a triple with the cc:creator or the cc:property, editor to describes a Dataset.
b. Extended provenance: This indicator checks if each entity of a Dataset has the source requested information in a way that an agent can identify the entity origin.
c. Wealth Provenance: Provides an information measure that a Dataset has on itself, using the metadata declarations proportion of the Dataset.

The reference [9] presents a quality metrics research before publishing or loading any kind of information to the Web. Some of these metrics include semantic and syntactic accuracy that seeks to filter data that is of poor quality from their methodological perspective. Authors as [4] in their paper present the implementation of an automated tool called Roomba, which is capable of validating, correcting and generating quality metadata with results that aim to the necessity of generating better quality metrics in metadata, it is a metric to improve metrics.

In a nutshell, at the end of this review is attached all the selected studies grouped by security topics: in blue, are the studies that contribute to the methodologies to measure data quality, in orange, the studies about data source and data reliability, and finally, in gray color, the studies about security and data access. Appendix 1 presents all related studies.

Identified researches are classified according to the solutions proposed. This classification is shown in Figure 4. Regarding to the identified methods, the group that has quality metrics in terms of reports, indicators and comparisons, is the most represented group.

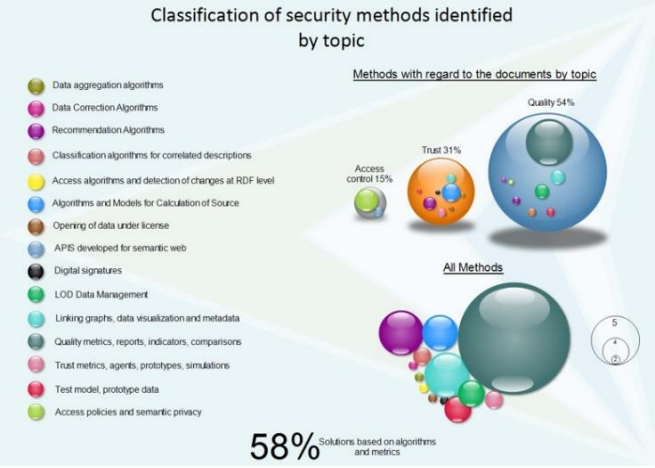

Figure 2. Tools and methods used in related studies. Source: Authors

In the same way, the most representative solutions are found in the Metrics and algorithms group, with 58% of representation. Some researches of each area, are highlighted below:

For the data quality and their linking, [68] implemented a solution called DSNotify, which addresses the problem of the broken links in Semantic Web, being at first a detector and orient of these links, and in essence, a proposed tool to manage the maintenance of the Web. The reference [96] address the problematic of semantic distance regarding similarity, using a proposed recommendation system that ends in an enhanced comparison with respect to other recommendation systems. It uses as a base the recommendation resources enabled to LOD such as Linked Data Semantic Distance (LDSD), which basically, calculate the number of direct or indirect links between two resources. As a complement, it uses normalization strategies, a statistic and probabilistic approach to calculate the semantic distance between two resources.

The reference [11] proposes an automatic evaluation metric framework of the data quality before they are published and that are quantitatively measurable for a determined Dataset. Six dimensions are defined, and each one is assessed with defined metrics.

The reference [95] present a tool called VizCurator that offers a set of tools to curate and visualize open data; it makes easier the extraction of temporary resources and the definition of temporary restrictions that the curator uses to identify contradictory or troubled facts and make them understandable. It offers a web browser based on a tree with code color, which allows the curator to navigate the RDF scheme and examine the entity, relations and extern link types. In the first level of this tree, the curator can see all entity types, and each of them can be extended to give more detailed or related information.

The reference [85] identified more relevant metrics to the quality control, from the quality linked data indicators. Using Datasets like DBpedia and Wikidata are defined crucial subsets for the domain. The experiment results suggest that the most of these network domains, the quality of Wikidata regarding most of the relevant metrics is higher than BDpedia.

The reference [78] propose a taxonomy as a classification system of the resulting data from their origin features to avoid the ambiguity. They present the reusing and establishment of metadata that include in their description, the lineage to help the data users to decide whether the data meet their searching criteria.

The reference [61] discuss options to get information of the source based on rules, focused on the metadata source model that includes two dimensions: Data creation and data access.

The reference [31] define the trust as integral part of Semantic Web from the user's point of view and not from the information itself. It is a model that integrates a subset of factors or metrics that define trust types. Moreover, it defines reasoning agents that simulate the trust user process, based on ontologies, axioms and reputation.

In [54] there is a project that allows the automatic adding of data without fissure to simplify the process of manual adding. This project adds the following of the source and justifications of why the data should be added to gain trust from the data consumer. Given that the adding algorithm is run during the query time, it is seen that the proposed algorithm is fast enough to work in real world environments. URI7 query means that a consumer sends URI of a concept as part of the query and as a response they get the description of the wished concept added to all available sources together with the global quality; the data sources are accompanied with the data of origin.

The reference [20] explore an area of Semantic Web publishing, named graphs that allow an assertive communication, due that the graphs are signed and may be assessed by information consumers using trust directives. They present a formal trust framework to be the base of trust layer of the Semantic Web using named graphs; it uses search agents and graph construction.

The reference [87] approach a solution to the Datasets that are published in LOD without the adding of any kind of metadata that specifies the access control conditions in which data are accessible. A control access model is defined, it offers to user's ways to define policies, restrict access to specific data RDF, based on social tags and contextual information.

The reference [40] propose to decentralize the user authentication and authorization applied to Semantic Web as part of the security. The collaborative creation of linked data, the possibility of editing access permissions to agents without the need to edit directly a metadata file. Access controls based on RDF and authorization based on ACL. Compatible use with traditional servers such as Apache. On the other hand, [60] propose a Linked Data Authorization (LDA) platform a top a policy language flexible enough to cover all newly emerged requirements, including context awareness. The proposed policy language leverages W3C's SPARQL query language expressiveness to protect every part of the data.

In [28], the author addresses the security environment with a project of City Data, using BigData to make decisions based on LOD. It explores the in-between data availability and uses ease. The use of APIs that link the city open information and makes it "smarter" is part of their solution to close this gap.

According to the classification defined in Annex 1, in Table 3 is presented a relation of the documents most directly cited

related with security topics according to the selected criteria analysis.

Table 3. Most cited studies. Source: Authors

Author	Title	Type of publication	Information Source	Citations	Year
Buneman, P., Khanna, S., and Wang Chiew, T.	Why and Where: A Characterization of Data Provenance	Book Section	Springer International Publishing	976	2001
Simmhan, Y. L., Plale, B., and Gannon, D.	A Survey of Data Provenance in e-Science	Journal Paper	SIGMOD Rec	843	2005
Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Den Bussche, J.	The Open Provenance Model core specification (v1.1)	Journal Paper	Elsevier B.V.	485	2011
Green, T. J., Karvounarakis, G., and Tannen, V	Provenance Semirings	Conference Proceedings	ACM Digital Library	431	2007
Bose, R., and Frew, J.	Lineage retrieval for scientific data processing: a survey	Journal Paper	ACM Digital Library	425	2005
Alexander K., Cyganiak, R., Hausenblas, M., Zhao, J.	Describing Linked Datasets on the Design and Usage of voiD, the "Vocabulary Of Interlinked Datasets.	Journal Paper	Linked Data on the Web (LDOW2009)	180	2009
Hartig, O.	Provenance Information in the Web of Data	Journal Paper	Proceedings of the Linked Data on the Web LDOW Workshop at WWW	153	2009
Gil, Y.,and Artz, D.	Towards content trust of web resources	Journal Paper	ScienceDirect	139	2007
Carroll, J., Bizer, C., Hayes, P., & Stickler, P.	Named Graphs, Provenance and Trust	Book Section	ACM Digital Library	110	2005
Kontokostas, D., & Westphal, P.	Test-driven evaluation of linked data quality	Conference Proceedings	ACM Digital Library	101	2014

5. Analysis, Considerations and Security Challenges in LOD.

According to the reviewed literature, some aspects, considerations and challenges are identified Attention should be paid to those in order to share Open Data using LOD.

5.1 Security in Semantic Data Consumption

Traditional query languages such as SQL, XPath or LDAP have suffered from security problems based on non-controlled user insertions in which is possible to chain programmed queries with information directly introduced by the user. These problems were seen in [37], where, they analyzed the same problematic applied to semantic query languages like SPARQL [90] and SPARUL [63]. In these researches are highlighted the identification of problems such as the SPARQL injection to obtain data which access can be restricted, or SPARUL injection to modify semantic data without permission. In [20] is analyzed the very same problematic, and it is proposed a solution like the entrance user control, adding a set of patches to automatically solve those problems in frameworks such as Jena or Pellet. As a result, the authors highlight the representation of RDF as several named graphs that might contain information about intellectual property, digital signatures, and ontologies with more and better self-references.

5.2 Data Provenance

One of the main aspects when assessing the authenticity of a Dataset and thereby its reliability is the data provenance that composes it. This topic has been studied in [86]. Furthermore, the W3C Provenance Incubator Group [1] defines the Data Web resource provenance as a register that describes people, entities and involved processes in production and release, or that have had any influence over that resource [31].

In relation to the kind of provenance, in work done by [64], two kinds of the provenance can be distinguished: why provenance and where provenance. The first one represents the origin involved in the creation of the information, whereas the second one represents the exact localization of where that information was taken from. In [16] it is shown another kind of provenance, how provenance, that refers to how the source participated in the information creation.

Regarding models to represent the data provenance, it has been traditionally researched in other areas, in those, there have been proposed different models, in databases, biology, and science, among others. In the area of the Web provenance, the research done by [73] are concluded a series of recommendations relative to the information representation about the data provenance, these are summarized below:

- Each resource from which the provenance wants to be identified, should be referenced with a URI, identify the data author and the process undertaken to its creation.
- It is recommendable that the resources were identified by identification provenance methods like a license or digital signature.
- The information about the provenance should be accessible for use and verification.
- It is recommendable the inclusion of a history of resource versions with temporal information about creation, modification and access of each version.

There are some representation provenance data models on the Web that implement totally or partially these recommendations:

- Open Provenance Model (OPM): It is a model of the general area, understood as an expandable and independent core of domain. OPM defines the provenance as a causal graph, where the nodes might be classes of type artifact (state of data in a specific moment), process (actions undertaken to create, modify or access the artifacts) or agents (process controllers). The graph edges define the causal relations between nodes, as well defined: used, wasControlledBy, wasTriggeredby, generated, wasDerivedFrom. On the other hand, OPM defines other concepts like accounts (partial subgraphs of the general graph used to represent different states of the same graph in different moments), or roles (allows to deeply describes some of the casual defined relations) [66].
- PROV-DM: It is a data model that it is being prepared by W3C Provenance Working Group [66], to be adopted as official, and that it is working based on OPM. It has as goal to provide a data model that can be used to represent entities, people and processes involved in the production of either data or any object. It is a model independent of the domain, but with a series of extension points defined to be extended to specific domains. The model is based on relationships between three main elements: Entity (object or data itself), Process Execution (represents an activity and influences the entities) and Agent (represents the person who launches and controls the activities). [32].

Additionally, there are several vocabularies and ontologies that serve to represent the provenance as metadata at the document level or Dataset:

- Open Provenance Model Vocabulary (OPMV) and Open Provenance Model Ontology (OPMO) that implement the bases of OPM model.
- Provenir Ontology: Used mainly to represent the source in the scientific area. It defines three classes to represent the provenance components: data (entity to represent the original product and its derivatives), process (entity to represent the processes that affect a product), agent (entity to represent who has undertaken a process) [23].
- Provenance Vocabulary: Developed to describe the data linked provenance. It is defined as an OWL modular ontology, with a simple core and general field extensible by modules. This model defines two provenance dimensions: data creation and access. It also defines concepts like actors, processes and artifacts applied to both dimensions. [59 - 66].
- Dublin Core (DC): DC offers a term vocabulary that serves to represent basic metadata of a resource. In terms of source, it has basic terms to refer to who created a resource, when and based on what, etc. [66].

It is worth mentioning that, at the expense of the ending of PROV-DM (which is foreseen to end as an official ontology of W3C), none of the data models or specific ontologies to represent the provenance has been adopted as a standard or official recommendation, even though OPM is the model normally used as reference. So that, DC is the most extended ontology to represent provenance, despite of not being a specific ontology to do so and that it does not reach the completeness of OPM or PROV-DM (it does not considerate relations beyond creation and does not deepen in the different entities that might exist in the source relations).

5.3 Provenance: a criterion for quality and trust

The provenance study is the base to establish verifiability criteria when calculating quality and trust. The verifiability means to offer the possibility of proving the accuracy and the correction degree of the information. To do so, in work done by [89] are specified several indicators to calculate the verifiability in the Semantic Web:

- The inclusion of basic information: It refers to the inclusion of at least basic data like author, editor and information collaborators, as well as references used to its creation. To do so, it is possible to use basic ontologies like DC.
- Use of dedicated vocabulary to represent the source: It refers to the use of more advanced source representation models such as OPMV, Provenance Vocabulary or PROV-DM.
- Use of digital signatures [66].

Figure 5 represents, as summarize, the classification of the literature found according to the selected criteria for the described analysis.

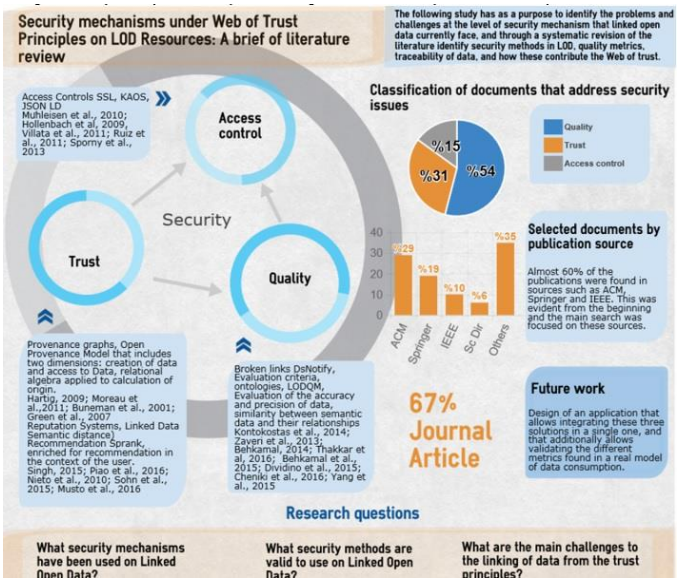

Figure 3. Classification of the literature according to the selected criteria. Source: Authors

6. Discussion

Data security concept embraces particular topics such access level and data protection. Security under LOD principles,

goes beyond of the particular security topics due to Linked Data has characteristics that allow security to be considered in a more integral way. For that reason, quality and trust characteristics are more important in information security when these characteristics are adapted to changing and fast-growing system like LOD. In order to present a discussion scenario, answers to the research questions are presented below.

6.1 What are the security mechanisms most used in LOD?

Security mechanisms found are based mainly on algorithms that implement agents to fulfill tasks, all of them important for security on LOD, such as: Searching broken links, construction or repairing links between semantic graphs, evaluation quality metrics, simulating behavior of common users, searching tasks, adding or editing data designed under intelligence artificial principles.

Among the security mechanisms that frame the solutions, it is worth to highlight the mechanism described by [30] where the data interoperability happens trough APIS that use a safe semantic layer that just can be accessed using HTTP requests associated the data URI. In [20] it had been mentioned the possibility of creating a semantic layer where the semantic agent would be used to search and construct graphs. From the above arises the problem that faces the APIS developers who work under LOD principles, which is the authentication of agents that attempt to access the exposed information by a web service [30]. Several web services suppliers have developed authentication mechanisms for their platforms; those can be used as authentication service to facilitate the user authentication in applications developed by third parties. An example of these services is Facebook Connect, developed by Facebook. An attempt to standardize and unify these authentication services is found in the OpenID standard to offer a unique authentication service for users of different platforms and applications; however, it has found a small adaptation level by users [81].

It arises the development of new authentication protocols that make a special emphasis in the achievement of a distributed authentication mechanism that reduces the role of the authentication central authority; an example of these mechanisms is BrowerID, developed by Mozilla Foundation and based on the use of the email address and the integration with the Web browser. Among the Linked Open Data Community, it also has been developed a distributed authentication mechanism known as WebID [40]. That mechanism is based on the use of basic elements of web technology, the URI that is associated with an identity using asymmetric cryptography under the form of a public certificate that web agents can use to dereference a URL. The web service that receives the request can follow the URL associated with the certificate, retrieving the document RDF with the needed information to check the certificate authenticity, as well as additional information about contacts, profile, etc., associated with that identity. By doing so, it is possible to get a really distributed authentication mechanism based on the Web principles and basic elements. At the same time, the mechanism can easily integrate with the most characteristic web agent, the web browser, due to the support of the use of certificates is now integrated into commercial web browsers [30].

Another of the mechanisms proposed in [79] where it is used the same knowledge provide by LOD, it is applied to establish trust levels by relying on search spaces that have some kind of recommendation in the context of the users based on their reputation.

6.2 What security methods are used on LOD?

All security methods that facilitate verification and authentication processes in LOD data management are used. Also, those methods that involve calculating data source. Either data generated without structure or data which does not allow a methodological provenance validation, represent a high risk for LOD security. In [59], for example, Open Data must be done under some kind of license, according to the Open Data Principles. Licenses have been one of the main strategies for provenance identification.

Security methods that are considered on LOD and have contribute to the discussion are shown in Table 4.

Table 4. Security Methods on LOD. Source: Authors

Methods	References	Tools
Access policy and semantic privacy	Mühleisen et al., 2010; Hollenbach et al, 2009, Villata et al., 2011; Ruiz et al., 2011; Sporny et al., 2013	Access Control SSL, KAOS, JSON LD
Quality Metrics	Kontokostas et al., 2014; Zaveri et al., 2013; Behkamal, 2014; Thakkar et al, 2016; Behkamal et al., 2015; Dividino et al., 2015; Cheniki et al., 2016; Yang et al., 2015	Broken links DsNotify, Evaluation criteria, ontologies, LODQM, Evaluation of the accuracy and precision of data, similarity between semantic data and their relationships
Algorithms and models for the calculation of origin.	Hartig, 2009; Moreau et al.,2011; Buneman et al., 2001; Green et al., 2007	Provenance graphs, Open Provenance Model that includes two dimensions: creation of data and access to Data, relational algebra applied to calculation of origin.
Recommendation algorithms	Singh, 2015; Piao et al., 2016; Nieto et al., 2010; Sohn et al., 2015; Musto et al., 2016	Reputation Systems, Linked Data Semantic distance) Recommendation Sprank, enriched for recommendation in the context user.

6.3 What are the main Linked Data challenges based on Trust Principles?

Public-sector information and finding a fair balance between Open Data and Privacy, is one of the biggest challenges that companies, governments and users face nowadays. The misuse of interchange information protocols when sharing information might generate new inefficiencies and introduce security risks within the organization that does not know how far to protect its information from the competence. [70].

The reference [8] present as a problem of nowadays Semantic Web that data cannot be read easily by computers. This is not a new Web problem, as data that has HTML failures nowadays still requires human intervention to be interpreted. Even though, due to the large quantity of

information generated in LOD, human intervention is not a feasible alternative in the future. In contrast with its fundamental motivation, the readability of LOD in a machine, it is an even bigger than thought obstacle. For instance, less than 10%of the Dataset is free, widely popular and much curated (now managed by WikiData), can be read by an analyzer compatible with standards such as Raptor. This percentage is even lesser for many of the introduced less common.

Recent researches have demonstrated that linked and published data in the LOD cloud are subject to frequent changes. As data changes in the cloud, the local copies need to be updated. Nevertheless, due to the limitations of the available computational resources (for example broadband to seek data, calculation time, etc.), the LOD applications might not be able to visit on the permanent basis all the LOD sources in short intervals to prove the changes. Some studies describe that the accuracy of constructed indexes on the LOD sources drops until 50% in weeks for changes made in the LOD cloud and that are not updated at the local level [27].

The reference [34] outline the three different kinds of potential attacks on Semantic Web architecture:

a. The Network attacker: On the network level, TLS is not in use currently for the majority of URIs on the Semantic Web, leading to trivial attacks on Linked Data.
b. The Web attacker: On the level of Web applications, proposed standards like WebID+TLS and the W3C Social Web standards have cryptographic security flaws.
c. The Semantic attacker: On the level of inference procedures, he shows how the preceding two levels can lead to attacks that can lead to corrupted inferences.

In general, the dependency of data retrieval and inferences based on insecure Semantic Web data can lead to attacks on trusted semantics of the Semantic Web itself. Author demonstrates that this does not have to be the case: Several standards from the IETF and W3C can be used to upgrade the Semantic Web to modern security-best practices, leading to a secure Semantic Web.

With the aim at improving the security, [65] explain that several notable technologies have been defined and integrated security standards to all phases of semantic web application development by the World Wide Web Consortium (W3C). To achieve this purpose, W3C has reported some efforts in in XML Key Management (XKMS), Security Assertion Markup Language (SAML), XML Access Control Markup Language (XACML) and Platform for Privacy Preferences (P3P). But there is a need to develop a security assessment benchmark for semantic web applications. The effective assessment of security of semantic web applications has been paid less attention so far in this regard.

7. Conclusions

Briefly, some security challenges on Linked Date are identified by [60], and are shown below:

- The need for personalized, user defined protection of vast amounts of heterogeneous data has not been considered before in such scale.
- The security protocols, procedures and tools are always a step behind in handling new security challenges.
- When it comes to sensitive data, no matter whether it is personal, social or corporate, strict rules must be applied to ensure that it is properly accessed and handled. The data owner's ability to control who and under what conditions gets accesses to their data can encourage them to expose beneficiary data for the greater goods, and at the same time, protect their privacy.
- Tools that enable security policy testing and preview of the protected data are important step towards gaining trust in the authorization platforms.
- The scale of the data to be protected, its creation velocity and its heterogeneity makes the centralized policy management unfeasible. Therefore, the policy formalism should support tools that will enable regular users to protect their own heterogeneous and distributed data.

In the last years, new tools have emerged which offer relevant access control models (MAC, DAC, RBAC, VBAC, ABAC, CBAC) and standardization efforts (XACML, WebID, WAC, P3P, APPEL, ODRL) [52], for instance. These tools offer different kind of solutions for LOD security problems, applying preventive and corrective approaches. Also, these tools work topics like design standards for semantic data management before it has been published, quality and trust criteria. In addition, these tools have components to calculate the provenance, licensing, digital signatures and authentication among other metrics.

However, the large volume of data generated in the LOD cloud is still limited by technological barriers with problems such as updating information in real time, and in a safely and reliably way [27]. At the security level, implementation of intelligent and automated agents with different tasks seems to be the most viable solution in the medium term. To delegate tasks that agents can perform more efficiently, considering the volume of information, seems to be a very safe method of control if authentication would be governed by the same rules, as well as the calculations of the trust measurement.

In this context, automatic data aggregation in order to simplify the manual aggregation process, described by [54], provides a strategy to follow-up of the provenance and justifications, which allows data consumer to generate trust, and they can specify the conflicts resolution optionally. In addition, [95] propose a tool called Vizcurator, which provides a set of tools for Open Data healing and visualization. This tool facilitates the resource extraction and the definition of temporal constraints. Using those constraints, curators can identify contradictory or conflicting facts and make them understandable. These proposals, like the distributed authentication mechanism such as WebID, consolidate what would be considered as the important aspects to have security account in the LOD data.

However, according to [34], so far there is no research oriented to that the security preserve the semantic level, which go beyond the mere "idea of semantic web security standardization". Instead, enterprise and government users simply believe that it is best to "access control or security occurs at the layer of the HTTP access and protocols, and not at the linked data layer."

Taking into account that In the original Semantic Web architecture design, the trust layer was envisioned to address authentication, identification, and proof checking, but did not mention trust in the content itself (entity trust); The reference [31] argue that Entity trust is a trust judgment regarding an entity based on its identity and its behavior, and is a blanket statement about the entity. This is insufficient in many situations that require selecting among sources of information. On the other hand, Content trust is a trust judgment on a particular piece of information in a given context. When considering content trust, one must determine what defines a unit of content and how it can be described.

8. Acknowledgment

This research has been development within the framework of the doctoral research project on Linked Data, at Universidad Distrital Francisco José de Caldas. In the same way, the subject matter is working as a line of Research Group GIIRA.

References

[1] K. Alexander, R. Cyganiak, M. Hausenblas, J. Zhao. "Describing Linked Datasets on the Design and Usage of VoiD - The Vocabulary of Interlinked Datasets". Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09). 2009. Retrieved from: http://richard.cyganiak.de/2008/papers/void-ldow2009.pdf

[2] M. Alouane H. El Bakkali. "Security, privacy and trust in cloud computing: A comparative study". International Conference on Cloud Technologies and Applications (CloudTech), Marrakech, pp. 1-8. 2015. DOI: 10.1109/CloudTech.2015.7336995. Retrieved from https://ieeexplore.ieee.org/document/7336995/.

[3] D. Artz, Y. Gil. "A survey of trust in computer science and the Semantic Web". Web Semantics: Science, Services and Agents on the World Wide Web, Volume 5, Issue 2, pp. 58-71. 2007. ISSN 1570-8268. Retrieved from https://www.isi.edu/~gil/papers/jws-trust-07.pdf

[4] A. Assaf, R. Troncy, A. Senart. "What's up LOD cloud? Observing the state of linked open data cloud metadata". Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 247–254. 2015. Retrieved from: http://doi.org/10.1007/978-3-319-25639-9_40

[5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives. "DBpedia: A nucleus for a Web of Open Data". Lecture Notes in Computer Science LNCS, pp. 722–735. 2007. Retrieved from: http://doi.org/10.1007/978-3-540-76298-0_52

[6] R. Ávila Alonso, V. Ortiz. "Principles application of Linked Open Data to the list of subject headings of Library of the Polytechnic University of Madrid". Universidad Carlos III, Madrid. 2014. Retrieved from http://eprints.rclis.org/34112/

[7] F. Bauer, M. Kaltenbock. "Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers". Edition mono/monochrom, Vienna, Austria. ISBN: 978-3-902796-05-9. 2012. Retrieved from https://www.reeep.org/LOD-the-Essentials.pdf

[8] W. Beek, L. Rietveld, S. Schlobach, F. Van Harmelen. "Why the Semantic Web Needs Centralization?". IEEE Internet Computing (Volume: 20, Issue: 2. pp. 78–81. 2016. Retrieved from http://ieeexplore.ieee.org/document/7420493/?reload=true

[9] B. Behkamal "Metrics-driven framework for LOD quality assessment". Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, pp. 806–816. 2014. Retrieved from http://doi.org/10.1007/978-3-319-07443-6_54

[10] B. Behkamal, E. Bagheri, M. Kahani, M. Sazvar. "Data accuracy: What does it mean to LOD?". 4th International Conference on Computer and Knowledge Engineering (ICCKE). pp. 80–85. 2014. Retrieved from http://doi.org/10.1109/ICCKE.2014.6993457

[11] B. Behkamal M. Kahani E., Bagheri. "Quality Metrics for Linked Open Data". Database and Expert Systems Applications. DEXA 2015. International Conference on Data Management in Cloud, Grid and P2P Systems. Lecture Notes in Computer Science, vol 9261. Springer, Cham. pp 144-152. 2015. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-22849-5_11

[12] F. Benedetti, S. Bergamaschi, L. Po. "Online Index Extraction from Linked Open Data Sources". CEUR Workshop Proceedings. Vol-1267 Linked Data for Information Extraction. 2014. Retrieved from http://ceur-ws.org/Vol-1267/LD4IE2014_Benedetti.pdf

[13] T. Berners-Lee, R. Fielding, L. Masinter. "STD 66 - RFC 3986. Uniform Resource Identifier (URI): Generic Syntax". RDF Editor. 2005. Retrieved from https://www.rfc-editor.org/info/rfc3986

[14] C. Bizer, T. Heath, T. Berners-Lee. "Linked data - the story so far". International Journal on Semantic Web and Information Systems, 5(3), pp 1-22. DOI:10.4018/jswis.2009081901. 2009. Retrieved from https://eprints.soton.ac.uk/271285/

[15] R. Bose, J. Frew. "Lineage retrieval for scientific data processing: a survey". ACM Computing Surveys, 37(1), pp 1–28. 2005. Retrieved from http://doi.org/10.1145/1057977.1057978

[16] J. Boyle. "James Boyle: A natural experiment". Financial Times. 2004. Retrieved from http://www.ft.com/intl/cms/s/2/4cd4941e-3cab-11d9-bb7b-00000e2511c8.html#axzz49ibVnDri.

[17] C. Böhm, F. Naumann, Z. Abedjan, D. Fenz, T. Grütze, D. Hefenbrock, D. Sonnabend. "Profiling linked open data with ProLOD". IEEE 26th International Conference on Data Engineering Workshops (ICDEW 2010), pp 175–178. 2010. https://doi.org/10.1109/ICDEW.2010.5452762

[18] S. Buchholtz, M. Bukowski, A. Śniegocki. "Big and Open data in Europe: A growth engine or a missed opportunity?". Warsaw Institute for Economic Studies (WISE Institute). 2014. Retrieved from http://wise-europa.eu/en/2014/02/26/big-open-data-in-europe-a-growth-engine-or-a-missed-opportunity/

[19] P. Buneman, S. Khanna, T. Wang-Chiew. "Why and Where: A Characterization of Data Provenance". ICDT - International Conference on Database Theory. Database Theory — ICDT 2001, pp 316–330. 2001. Retrieved from http://link.springer.de/link/service/series/0558/bibs/1973/19730316.htm

[20] J. Carroll, C. Bizer, P. Hayes, P. Stickler. "Named Graphs, Provenance and Trust". WWW '05 Proceedings of the 14th international conference on World Wide Web, pp. 613–622. Japan. 2005. Retrieved from http://doi.org/10.1145/1060745.1060835

[21] N. Cheniki, A. Belkhir, Y. Sam, N. Messai. "LODS: A Linked Open Data Based Similarity Measure". IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), pp 229–234. 2016. Retrieved from https://doi.org/10.1109/WETICE.2016.58

[22] D. Corsar, P. Edwards, J. Nelson. "Personal privacy and the web of linked data". 13th International Conference on Society, Privacy and the Semantic Web - Policy and Technology - Volume 1121 (PrivOn'13), Vol. 1121. pp 11-21.

2013. Retrieved from http://ceur-ws.org/Vol-1121/privon2013_paper2.pdf.

[23] Creativecommons. Creative Commons. 2019. Retrieved from http://creativecommons.org/

[24] E. Damiani E., S. De Capitani di Vimercati C. Fugazza, P. Samarati. "Extending Policy Languages to the Semantic Web". Web Engineering. International Conference on Web Engineering. Lecture Notes in Computer Science, vol 3140, pp 330 - 343. Springer, Berlin, Heidelberg. 2004. Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-27834-4_41.

[25] J. Debattista, C. Lange, S. Auer. "Luzzu - A framework for linked data quality assessment". ISWC 2015 Posters and Demonstrations Track, 14th International Semantic Web Conference, ISWC 2015. 2015. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84955600970&partnerID=tZOtx3y1

[26] G. Demartini, D. Difallah, P. Cudré-Mauroux. "Large-scale Linked Data Integration Using Probabilistic Reasoning and Crowdsourcing". The VLDB Journal, 22(5), pp 665–687. 2013. Retrieved from http://doi.org/10.1007/s00778-013-0324-z

[27] R. Dividino, T. Gottron, A. Scherp. "Strategies for Efficiently Keeping Local Linked Open Data Caches Up-To-Date". The 14th International Semantic Web Conference. 2015. Retrieved from http://iswc2015.semanticweb.org/sites/iswc2015.semanticweb.org/files/93670303.pdf

[28] M. Fox. "City Data: Big, Open and Linked". Municipal Interfaces, November, pp. 19-25. 2013. Retrieved from https://www.researchgate.net/publication/262674890_City_Data_Big_Open_and_Linked

[29] Q. Gao, G. Houben. "A Framework for Trust Establishment and Assessment on the Web of Data". 19th International World Wide Web Conference. 2010. Retrieved from http://ramb.ethz.ch/CDstore/www2010/www/p1097.pdf

[30] A. Garrote Hernández. "APIs semánticas para la web orientada a datos enlazados". PhD Thesis. Universidad de Salamanca (España). 2014. Retrieved from http://gredos.usal.es/jspui/handle/10366/124158

[31] Y. Gil, D. Artz. "Towards content trust of web resources". Journal of Web Semantics, Volume 5, Issue 4, December 2007, pps 227-239. 2007. Retrieved from http://doi.org/10.1016/j.websem.2007.09.005

[32] GNU. "GNU Lesser General Public License". GNU Project - Free Software Foundation. 2016. Retrieved from http://www.gnu.org/licenses/lgpl.html

[33] T. Green, G. Karvounarakis, V. Tannen. "Provenance Semirings". PODS '07 Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. pp 31-40. 2007. Retrieved from: http://doi.org/10.1145/1265530.1265535

[34] H. Halpin. "Semantic Insecurity: Security and the Semantic Web". PrivOn 2017 - Workshop Society, Privacy and the Semantic Web - Policy and Technology, pp.1-10. 2017. Retrieved from https://hal.inria.fr/hal-01673291/document.

[35] O. Hartig. "Provenance Information in the Web of Data". Proceedings of the Linked Data on the Web LDOW Workshop at WWW, 39(27), pp 1–9. http://doi.org/10.1016/S0040-4039. 2009. Retrieved from http://events.linkeddata.org/ldow2009/papers/ldow2009_paper18.pdf

[36] O. Hartig, J. Zhao. "Using web data provenance for quality assessment". SWPM'09 Proceedings of the First International Conference on Semantic Web in Provenance Management - Volume 526, pp 29-34. 2009. Retrieved from https://dl.acm.org/citation.cfm?id=2889881

[37] O. Hartig, J. Zhao. "Publishing and Consuming Provenance Metadata on the Web of Linked Data". International Provenance and Annotation Workshop, pp. 78–90. 2010. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-17819-1_10

[38] T. Hernández-Pérez, M. García-Moreno. "Open data and data repositories: a new challenge for librarians". The professional information, 22(3), pp 259–263. 2013. Retrieved from: http://doi.org/10.3145/epi.2013.may.10

[39] P. Hitzler, K. Janowicz. "Linked Data, Big Data, and the 4th Paradigm". Semantic Web Journal, pp 233–235. 2013. Retrieved from http://iospress.metapress.com/index/552L1266G1655N4V.pdf

[40] J. Hollenbach, J. Presbrey, T. Berners-Lee. "Using RDF metadata to enable access control on the social semantic web". CEUR Workshop Proceedings. 2009. Retrieved from http://dig.csail.mit.edu/2009/Papers/ISWC/rdf-access-control/paper.pdf

[41] K. Holtman, A. Mutz. "RFC 2295 - Transparent Content Negotiation in HTTP". RFC Editor. 1998. Retrieved from https://www.rfc-editor.org/info/rfc2295

[42] S. Hosseinzadeh, S. Virtanen, N. Diaz Rodriguez, J. Lilius. "A semantic security framework and context-aware role-based access control ontology for smart spaces". SBD '16 Proceedings of the International Workshop on Semantic Big Data, pp 1-6. 2016. Retrieved from https://dl.acm.org/citation.cfm?id=2928300

[43] I. Jacobi, L. Kagal, A. Khandelwal. "Rule-Based Trust Assessment on the Semantic Web". International Workshop on Rules and Rule Markup Languages for the Semantic Web. Lecture Notes in Computer Science, vol 6826, pp 227-241. 2011. Retrieved from https://link.springer.com/chapter/10.1007/978-3-642-22546-8_18

[44] A. Josang, R. Ismail, C. Boyd. "A survey of trust and reputation systems for online service provision". Decision Support Systems, Volume 43, Issue 2, pp. 618-644. 2007. Retrieved from https://www.oasis-open.org/committees/download.php/28303/JIB2007-DSS-Survey.pdf

[45] M. Joshi, S. Mittal, K. Joshi, T. Finin. "Semantically Rich, Oblivious Access Control Using ABAC for Secure Cloud Storage". IEEE International Conference on Edge Computing (EDGE). 2017. Retrieved from https://ieeexplore.ieee.org/document/8029268/.

[46] A. Kasten. "Secure semantic web data management: Confidentiality, Integrity, and Compliant Availability in Open and Distributed Networks". Doctoral Thesis, Universitat Koblenz-Landau. 2016. Retrieved from https://kola.opus.hbz-nrw.de/frontdoor/index/index/docId/1393

[47] S. Keele. "Guidelines for performing systematic literature reviews in software engineering: Technical report". EBSE Technical Report EBSE. 2007. Retrieved from https://userpages.uni-koblenz.de/~laemmel/esecourse/slides/slr.pdf

[48] W3C. Provenance Working Group Wiki. 2011. Retrieved from https://www.w3.org/2011/prov/wiki/Main_Page

[49] B. Kitchenham. "Procedures for performing systematic reviews". Keele University, 33, 2004. Retrieved from http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

[50] B. Kitchenham, O. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman. "Systematic literature reviews in software engineering–a systematic literature review". Information and software technology, 51(1), pp 7-15. 2009. Retrieved from http://www.sciencedirect.com/science/article/pii/S0950584910000467

[51] Kirrane et al., 2016. Kirrane, S., Mileo, A., Decker, S. (2016). Access control and the Resource Description Framework: A survey. Semantic Web. 8. 1-42. Retrieved from http://www.semantic-web-journal.net/system/files/swj1084.pdf

[52] S. Kirrane, A. Mileo, S. Decker. "Access control and the Resource Description Framework: A survey". Semantic Web,

8, pp 311-352. 2017. Retrieved from http://www.semantic-web-journal.net/system/files/swj1280.pdf.

[53] S. Kirrane, S. Villata, M. d'Aquin. "Privacy, Security and Policies: A review of Problems and Solutions with Semantic Web Technologies". Semantic Web, Volume 9, Number 2, pp. 153-161. 2018. Retrieved from http://www.semantic-web-journal.net/system/files/swj1801.pdf

[54] T. Knap, J. Michelfeit, M. Necasky. "Linked Open Data Aggregation: Conflict Resolution and Aggregate Quality". 2012 IEEE 36th Annual Computer Software and Applications Conference Workshops, pp 106–111. 2012. Retrieved from https://doi.org/10.1109/COMPSACW.2012.29

[55] C. Knoblock. "The Semantic Web". University of Southern California. 2009. Retrieved from https://www.isi.edu/integration/courses/csci548_2009/slides09/SemanticWeb.pdf

[56] D. Kontokostas, P. Westphal, S. Auer, S. Hellmann, J. Lehmann, R. Cornelissen, A. Zaveri. "Test-driven evaluation of linked data quality". WWW '14 Proceedings of the 23rd international conference on World Wide Web, pp 747-758. DOI: http://doi.org/10.1145/2566486.2568002. 2014. Retrieved from https://dl.acm.org/citation.cfm?doid=2566486.2568002.

[57] J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, E. Almasi. "Open data: Unlocking innovation and performance with liquid information". McKinsey Digital. 2013. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

[58] S. McCarron. "XHTML+RDFa 1.1 - Third Edition". W3C. 2015. Retrieved from https://www.w3.org/TR/xhtml-rdfa/

[59] P. Miller, R. Styles, T. Heath. "Open data commons, a license for open data". CEUR Workshop Proceedings. 2008. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.6655&rep=rep1&type=pdf

[60] I. Mishkovski, R. Stojanov, S. Gramatikov, D. Trajanov. "Linked Data Authorization Platform". IEEE Access, Vol 6, pp 1189 - 1213. 2017. Retrieved from https://www.researchgate.net/publication/321328847_Linked_Data_Authorization_platform

[61] L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J.M yers, B. Plale, Y. Simmhan, E. Stephan, J. Vanden Busschef. "The Open Provenance Model core specification (v1.1)". Future Generation Computer Systems, 27(6), pp 743–756. 2011. http://doi.org/10.1016/j.future.2010.07.005

[62] C. Musto, F. Narducci, P. Lops, M. Gemmis, G. Semeraro. "ExpLOD : a framework for explaining Recommendations based on the Linked Open Data cloud". Proceedings of the 10th ACM Conference on Recommender Systems. pp 151-154. 2016. Retrieved from https://dl.acm.org/citation.cfm?id=2959173&dl=ACM&coll=DL&CFID=998636676&CFTOKEN=54005840

[63] H. Mühleisen, M. Kost, J. Freytag. "SWRL-based Access Policies for Linked Data". Proceedings of the Second Workshop on Trust and Privacy on the Social and Semantic Web. 2010. Retrieved from http://hannes.muehleisen.org/SPOT2010-muehleisen.pdf

[64] J. Cheney, L. Chiticariu, W. Tan. "Provenance in Databases: Why, How, and Where". Foundations and Trends in Databases, Vol. 1, No. 4, pp 379–474. 2007. Retrieved from http://homepages.inf.ed.ac.uk/jcheney/publications/provdbsurvey.pdf

[65] U. Noor, Z. Rashid. "Secure Semantic Web Application Development: Present and Future". IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), pp. 694-699. DOI: 10.1109/CSE-EUC-DCABES.2016.263. 2016. Retrieved from https://ieeexplore-ieee-org.bdigital.udistrital.edu.co/document/7982325/

[66] Open Data Commons. "Legal tools for Open Data". 2011. Retrieved from http://opendatacommons.org/

[67] V. Papavasileiou, G. Flouris, I. Fundulaki, D. Kotzinos, V. Christophides. "High-level Change Detection in RDF(S) KBs". ACM Trans. Database Syst. 2013. Retrieved from http://doi.org/10.1145/2445583.2445584

[68] N. Popitsch, B. Haslhofer. "DSNotify - A solution for event detection and link maintenance in dynamic datasets". Journal of Web Semantics, 9(3), pp 266–283. 2011. Retrieved from http://doi.org/10.1016/j.websem.2011.05.002

[69] A Rafea. "Trust and Proof in the Semantic Web". Knowledge Engineering, the American University in Cairo. 2007. Retrieved from http://www.cse.aucegypt.edu/~csci585/StudentsProjectsSpring07/Jarhi&GaalyReport.pdf

[70] C. Ruiz, G. Álvaro, J. Gómez-Pérez. "A framework and implementation for secure knowledge management in large communities". Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies. 2011. Retrieved from http://doi.org/10.1145/2024288.2024312

[71] O. Sacco, A.Passant. "A privacy preference ontology (PPO) for linked data". CEUR Workshop Proceedings. 2011. Retrieved from http://events.linkeddata.org/ldow2011/papers/ldow2011-paper01-sacco.pdf

[72] S. Sahoo, A. Sheth. "Provenir ontology: Towards a Framework for eScience Provenance Management". Microsoft eScience Workshop, Pittsburgh. 2009. Retrieved from http://knoesis.wright.edu/library/resource.php?id=741

[73] S. Sampaio, C. Dong, P. Falcone. "Incorporating the Timeliness Quality Dimension in Internet Query Systems". International Conference on Web Information Systems Engineering - WISE 2005, pp 53-62. 2005. Retrieved from https://link.springer.com/chapter/10.1007/11581116_6

[74] R. Sandhu, D. Ferraiolo, D. Kuhn. "NIST model for role-based access control: Towards a unified standard". Proceedings of the ACM Workshop on Role-Based Access Control, pp 47-63. 2000. Retrieved from https://www.nist.gov/publications/nist-model-role-based-access-control-towards-unified-standard

[75] M. Singh. "Norms as a basis for governing sociotechnical systems". IJCAI International Joint Conference on Artificial Intelligence, pp 4207–4211. 2015. Retrieved from http://doi.org/10.1145/0000000.0000000

[76] A. Shiri. "Linked Data Meets Big Data: A Knowledge Organization Systems Perspective". Advances in Classification Research Online, 24(1), pp 16–20. 2014. Retrieved from http://doi.org/10.7152/acro.v24i1.14672

[77] C. Sifaqui, F. Cifuentes-Silva, J. Labra-Gayo. "Towards an Architecture and Adoption Process for Linked Data Technologies in Open Government Contexts: A Case Study for the Library of Congress of Chile", Proceedings of the 7th International Conference on Semantic Systems, pages 79-86. 2011. Retrieved from http://doi.org/10.1145/2063518.2063529

[78] Y. Simmhan, B. Plale, D.Gannon, D. "A Survey of Data Provenance in e-Science". SIGMOD Rec., Volume 34 Issue 3, pp 31-36. 2005. Retrieved from https://dl.acm.org/citation.cfm?id=1084812

[79] M. Sohn, S. Jeong, J. Kim, H. Lee. "Augmented context-based recommendation service framework using knowledge over the Linked Open Data cloud". Pervasive and Mobile Computing, Volume 24, pp 166-178. 2015. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S1574119215001479?via%3Dihub

[80] M. Sporny, G. Kellogg, M. Lanthaler. "Json-Ld 1.0. A JSON Based Serialization for Linked Data", W3C Recommendation, pp 1–33. 2013. Retrieved from https://www.researchgate.net/publication/259671337_JSON-LD_10_-_A_JSON-based_Serialization_for_Linked_Data

[81] S. Sun, E. Pospisil, K. Beznosov. "What Makes Users Refuse Web Single Sign-On? An Empirical Investigation of OpenID Categories and Subject Descriptors". Proceedings of the Seventh Symposium on Usable Privacy and Security. 2011. Retrieved from https://dl.acm.org/citation.cfm?id=2078827.2078833

[82] M. Sule, M. Li, G., Taylor. "Trust Modeling in Cloud Computin"g. IEEE Symposium on Service-Oriented System Engineering (SOSE), pp. 60-65. DOI: 10.1109/SOSE.2016.32. 2016. Retrieved from https://ieeexplore.ieee.org/document/7473010/

[83] Z. Syed, A. Padia, L, Mathews, T. Finin, A. Joshi. "UCO: A Unified Cybersecurity Ontology". Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security. 2016. Retrieved from https://ebiquity.umbc.edu/paper/html/id/722/UCO-A-Unified-Cybersecurity-Ontology

[84] W. Tan. "Provenance in Databases: Past, Current, and Future". Sigmod 2007. Retrieved from http://sites.computer.org/debull/A07dec/wang-chiew.pdf

[85] H. Thakkar, K. Endris, J. Gimenez-Garcia, J. Debattista, C. Lange, S. Auer. "Are Linked Datasets fit for Open-domain Question Answering? A Quality Assessment". Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics, WIMS 2016. 2016. Retrieved from https://dl.acm.org/citation.cfm?id=2912857&preflayout=tabs

[86] G. Tummarello, C. Morbidoni, P. Puliti, F. Piazza. "Signing Individual Fragments of an RDF Graph". Special interest tracks and posters of the 14th international conference on World Wide Web, pp. 1020–1021. 2005. http://doi.org/10.1145/1062745.1062848

[87] S. Villata, N. Delaforge, F. Gandon, A. Gyrard. "An Access Control Model for Linked Data". Move to Meaningful Internet Systems: OTM 2010 Workshops. 2010. Retrieved from http://doi.org/10.1007/978-3-642-16961-8

[88] W3C. "Resource Description Framework (RDF)". 2014. Retrieved from https://www.w3.org/RDF/

[89] W3C. "PingTheSemanticWeb". 2016. Retrieved from https://www.w3.org/wiki/PingTheSemanticWeb

[90] S. Weibel, J. Kunze, C. Lagoze, M. Wolf. "Dublin core metadata for resource discovery". Internet Engineering Task Force RFC. RFC Editor. 1998. Retrieved from http://www.hjp.at/doc/rfc/rfc2413.html

[91] H. Yang. C Hsu. "Semantic Recommendation Using Linked Open Data". ASE BD&SI '15 Proceedings of the ASE Big Data & Social Informatics. 2015. https://dl.acm.org/citation.cfm?id=2818933

[92] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, S. Auer, P. Hitzler. "Quality assessment methodologies for linked open data". Semantic Web Journal. 2013. Retrieved from http://semantic-web-journal.net/system/files/swj414.pdf

[93] J. Zhao, A. Miles, G. Klyne, D. Shotton. "Linked data and provenance in biological data webs". Briefings in Bioinformatics, 10(2), pp 139–152. 2008. Retrieved from http://doi.org/10.1093/bib/bbn044

[94] L. Zhao, R. Ichise. "Graph-based Ontology Analysis in the Linked Open Data". I-SEMANTICS '12 Proceedings of the 8th International Conference on Semantic Systems, pp 56–632012. DOI: http://doi.org/10.1145/2362499.2362508

[95] B. Ghadiri, C. Christodoulakis, S. Hassas Yeganeh, R. Miller, K. Lyons, O. Hassanzadeh. "VizCurator: A Visual Tool for Curating Open Data". In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, pp 195-198. 2015. DOI: http://dx.doi.org/10.1145/2740908.2742845

[96] G. Piao, J. Breslin. "Measuring semantic distance for linked open data-enabled recommender systems". In Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16). ACM, pp 315-320. 2016. DOI: https://doi.org/10.1145/2851613.2851839

[97] Ontotext, "What are Linked Data and Linked Open Data?". 2018. Retrieved from https://ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/

[98] W3C. "Linked Open Data". 2018. Retrieved from https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

[99] M. Aymen Chalouf, F. Krief. "A Secured Service Level Negotiation in Ubiquitous Environments". International Journal of Communication Networks and Information Security, Vol 1 num, 2, pp 9-18. 2009. Retrived from http://www.ijcnis.org/index.php/ijcnis/article/view/9/9

[100] A. Rezakhani, H. Shirazi, N. Modiri. A novel access control model based on the structure of applications. International Journal of Communication Networks and Information Security (IJCNIS), Vol 8 Num 2. ISSN: 2076-0930. 2016. Retrieved from http://www.ijcnis.org/index.php/ijcnis/article/view/1994