Towards Blockchain and Semantic Web

Juan Cano-Benito$^{(\boxtimes)}$, Andrea Cimmino$^{(\boxtimes)}$, and Raúl García-Castro$^{(\boxtimes)}$

Universidad Politécnica de Madrid, Madrid, Spain
{jcano,cimmino,rgarcia}@fi.upm.es

Abstract. Blockchain has become a pervasive technology in a wide number of sectors like industry, research, and academy. In the last decade a large number of tailored-domain problems have been solved thanks to the blockchain. Due to this reason, researchers expressed their interest in combining the blockchain with other well-known technologies, like Semantic Web. Unfortunately, as far as we known, in the literature no one has presented the different scenarios in which Semantic Web and blockchain can be combined, and the further benefits for both. In this paper, we aim a providing an in-depth view of the beneficial symbiotic relation that these technologies may reach together and report the different scenarios that we have identified in the literature to combine Semantic Web and blockchain.

Keywords: Blockchain · Semantic Web · Semantic blockchain

1 Introduction

In the last decade the blockchain technologies have become a pervasive in our world [1]. Sectors like finance, security, IoT, or public services have benefited from the quantum leap that block chain has brought [2]. The wide range of domains in which this technology has been used has led researchers to elicit and analyse the problems and challenges related to the use of blockchain technologies [3].

One of the interests that researchers have shown lately is to combine the Semantic Web and blockchain technologies [4,4,5]. The reason of this interest relies on the symbiotic relationship that enhances both technologies, and the potential that can be reached combining them [6]. As far as we known current literature focuses mainly in applications that rely on blockchain and Semantic Web, with the exception of English et al. who presented the only article that analyses the benefits of combining these technologies [7]. The work of English et al. provides an overview of what the Semantic Web can do for the blockchain and vice-versa, nevertheless their work focuses on covering a large number of topics, and not only the benefits, and thus, they lack of an in-depth analysis of the benefits and the scenarios in which both technologies are combined.

In this paper we aim to extend the work of English et al. [7], by providing an in-depth analysis of the benefits that blockchain may find by relying on Semantic Web and vice-versa. In addition, our goal is to provide an overview

of the different scenarios and approaches to combine blockchain with Semantic Data by analysing the advantages and disadvantages of the different scenarios.

The rest of this article is organised as follows: Sect. 2 introduces the key concepts of the blockchain and the Semantic Web. Thirdly, Sect. 3 presents the different benefits that both technologies may offer to the other. After that, Sect. 4 introduces the different scenarios that we have identified to combine blockchain and Semantic Web. Finally, Sect. 5 recaps our conclusions.

2 Preliminaries

In this section we aim at introducing the key-concepts of the blockchain and the Semantic Web, as well as, the main characteristics of both. Our goal is not to provide an in-depth description, instead we aim at describing only the key-concepts to later explain on top of these the benefits and the scenarios.

2.1 Blockchain

The blockchain counts with several implementations each of which has some differences, e.g. Bitcoin or Ethereum. Nevertheless, all these implementations follow a common bottom line. The blockchain can be seen as a shared database among several peers who validate its content without the intervention of a third part. Some of the key concepts of blockchain are the following ones:

Definition 1 (Chain). *The chain is a list of blocks. When a peer aims at adding a new block a specific procedure has to be followed: the hash of the new block is computed relying on the hash of the last block in the chain, then the block is forwarded to all the peers who share the chain, and is included only when the rest of the peers validate the new block. In addition, when something is included in the chain it becomes immutable, and thus, that information is no longer modifiable by anyone.*

Definition 2 (Block). *The blocks are the minimal storage unit of the blockchain. They have some meta-data, like the hash of the previous block and the current hash, and the actual data, i.e., their content. The content can be expressed in multiple formats regardless if another block in the same chain used a different one. Depending on the current implementation the meta-data may change. In addition, the content of a block is limited to a maximum amount of data that may change depending on the implementation, e.g., Bitcoin supports up to 80 bytes [8] whereas Ethereum supports only 32 bytes [9].*

Depending on the peers and the grants to read/write in the chain that they have, three types of blockchains are distinguished:

- Public blockchain: any peer is able to read in the chain, even anonymously, and write blocks in such chain.

– Consortium blockchain: peers may only read the chain if an administrator invites them, however, once invited, all the peers are able to write new blocks in the chain.
– Private blockchain: peers may only read the chain if they received an invitation from an administrator, in addition, further permissions must be granted in order to write in the chain.

On the one hand, a well-known drawback in blockchain is its scalability, blocks with large amount of data entails that a larger storage space is required, and thus, the propagation of the block to the rest of the peers in the network will be slower [2]. On the other hand, since the chain is shared by several peers and the blocks are build relying on the previous ones the blockchain is more secure than other data stores [10].

2.2 Semantic Web

The Semantic Web is an alternative to the classic Web of Documents that counts with a stack of bespoke technologies [11]. The collection of Semantic Web technologies is large, following we present the more relevant ones and the key capabilities that Semantic Web technologies support.

Definition 3 (RDF). *The W3C has promoted a standardised formal language called Resource Description Framework (RDF) [12], which allows to describe data regardless the format used to express it. Data is expressed in form of triplets in which the first and the second elements are known as subject and predicate, respectively, and are URIs, and the third element is known as object, which can be either a URI referencing a subject or a literal. The data expressed in RDF is modelled as a graph.*

Definition 4 (Virtual RDF). *Data expressed in RDF is usually stored in a file or a triple store. In the literature there are several proposals that relying on that stored RDF produce on the fly new RDF data, which is called virtual RDF. Another approach to generate virtual RDF is publishing RDF from heterogeneeus datasources relying on specifications provided by users [13,14].*

Definition 5 (Linked Data). *The RDF data published following the principles proposed by Tim Berners-Lee [15] is known as Linked Data. These principles refer to a set of data quality requirements that the data must meet:*

1. *Use URIs as names for resources*
2. *Use HTTP URIs so that people can look up those names*
3. *When someone looks up a URI, provide useful information using the standards (RDF, SPARQL)*
4. *Include links to other URIs so that they can discover more things*

According to this principles two features must be remarked, on the one hand, the resources are identified relying on URIs which are dependant on the domain

name space (DNS) used to publish the data. This means that if the DNS changes the name space provided, then the resources will no longer be identified by these URIs nor retrievable. On the other hand, the data is stored following a decentralised approach, however when consuming such data it will appear as an unique dataset thanks to the links between datasets and the online availability of the resources.

Definition 6 (RDFs and OWL). *The W3C has promoted two formal languages to model data [16,17], i.e., RDFs and OWL. The use of this languages aim at defining ontologies that are formal models. One of the characteristics of the ontologies is the fact that they support reasoning over the data they model. They allow to validate the consistency of data, or generate new data that is not explicitly defined or stored by using reasoning mechanisms. There are multiple standard ready-to-use ontologies in the literature for a wide range of domains [18].*

Definition 7 (Ontology mappings). *Linked data allows to store data in different datasets, and, at the same time, thanks to the links consuming such data as a whole. Ontologies have a similar mechanism known as mappings that allow to relate one ontology to another [19], meaning that even if a local ontology is used to define some data if the mappings exists the same data can be automatically modelled according to another ontology referenced by the mappings.*

Definition 8 (SPARQL). *The W3C has promoted a formal language to query data express in RDF following an ontology [20], i.e., SPARQL. Assuming there is an engine that reads all the data, then SPARQL allows to query and consume the data from one dataset, or several datasets at once, i.e., SPARQL federation [21].*

Definition 9 (Data shapes). *In order to validate RDF data against a set of constraints the W3C promoted a new language known as Shapes Constraint Language [22], also known as data shapes. This language allows to specify a wide range of constraints, from the structure that data should follow to how literals should look like according to a regular expression. In addition, the data shapes allow to define some so-called SPARQL definitions that create virtual RDF.*

3   Benefits

The work presented by English et al. [7] introduced namely one benefit that Semantic Web offers to the blockchain, i.e., by using ontologies to model the meta-data of the blocks practitioners may perform queries using SPARQL. Unfortunately, such approach does not query the content of the blocks since their data was not expressed in RDF, as pointed out by the authors. On the other hand, the benefit that Semantic Web may find by using blockchain according to English et al. is the following one: IRIs to identify resources in Semantic Web depend on a DNS, by using blockchain technologies the identification of resources can be relayed on the hashes of a blockchain that will point to their related properties, achieving decentralise the domain name system (DNS).

However, after analysing both technologies we have reached a richer and detailed list of benefits. Considering the Semantic Web technologies, the blockchain may find the following benefits:

1. **One language multiple formats:** RDF is not bounded to a specific format therefore the data described following this standard can be expressed in multiple formats, e.g., RDF/XML, JSON-LD, TURTLE. Blockchain may rely on RDF to write the content of the blocks, using the format that suits better a specific domain problem.
2. **Model data following well-known standards:** Semantic Web technologies like RDF, SPARQL, RDFs, or OWL are well-known W3C standards. This entails that there is a global consensus which makes them reliable and trustworthy. In addition, there is a large amount of standard ontologies ready-to-use to describe the data of multiple domains easing the modelling task for a given problem.
3. **Linking of data:** one of the properties that promotes RDF is to reference data from other datasets by relying on links, obtaining as a result a global view of the data although the storage of the different fragments is distributed through the web. Linking data is a well-known, and largely address, challenge in the Semantic Web community [23]. Blockchain may benefit from this feature by linking the content of the blocks with external datasets, or even with the content of other blocks in the same or in a different chain.
4. **Multiple data models:** ontologies count with mappings to relate their properties and classes to other ontologies. This feature means that when a blockchain relies on an ontology that has mappings to another ontology, the data described with the former ontology can be translated to the model of the latter automatically. This is specially suitable when data must be modelled at the same time differently depending on who is consuming such data.
5. **Search over blockchain:** assuming the blockchain relies on ontologies and RDF to describe their meta-data and their content, then, practitioners may use SPARQL to query the chain. However, this capability requires a third part service that reads the blockchain and executes the queries.
6. **Blockchain data and meta-data validation:** once a chain relies on ontologies and RDF it may benefit from the shapes to validate its data and meta-data. The shapes report the errors in the model and the content of a specific RDF document.
7. **Blockchain consistency validation:** the Semantic Web counts with reasoning engines that allow to check the consistency of the data. In this case the blockchain should count with a third-party service to perform the reasoning and have its meta-data expressed relying on an ontology.
8. **Virtual RDF:** data expressed in RDF is not always stored, instead sometimes such data is generated on the fly. This generated RDF is known as virtual RDF. One approach consist in using reasoning engines, which infer new data. An alternative approach is to infer virtual RDF relying on data shapes. Finally, another approach relies on graphs embedding that anal-

yse the current data and create new knowledge relying on machine learning proposals [24].

9. **Virtual RDF services:** Semantic Web counts with some engines that relying on specifications are able to translate on the fly data from heterogeneous datasources into RDF [13,14]. A blockchain may benefit from this kind of engines, known as virtual RDF services, by storing data in non-RDF formats and relying on these third-party services to generate at the same time their content in RDF.

10. **Interoperability:** relaying on semantic web technologies, and using standard ontologies to model data, both meta-data and content, the blockchain becomes interoperable. Been interoperable means that an information system is able to transparently interact with other interoperable systems, e.g., other blockchains, databases, or services. In addition, third-party systems can discover interoperable systems and know how to access their data automatically.

The benefits that Semantic Web may find by using blockchain are:

1. **Data decentralisation:** on the one hand, the Semantic Web aims at storing data following a decentralised approach, on the other hand, blockchain is a decentralised data store. Therefore, Semantic Web may benefit from the decentralised nature of the blockchain in order to store the data. In addition, the fact that the chain is shared by several peers increases the data availability, which normally depends only on one service.

2. **Identifiers for RDF resources:** one of the well-known benefits that blockchain brings to the Semantic Web is the generation of identifiers that do not depend on a DNS [7], e.g., Ethereum Name Service (ENS).

3. **Data immutability:** Semantic Web has been adopted by several public entities to provide open, accessible, and transparent data, e.g., open government initiatives [25]. These scenarios require that data published and its provenance can be trusted. One of the well-known properties of blockchain is that once data is published it cannot be modified. This makes blockchain the perfect technology to be used in these scenarios. For instance, a Danish political party in 2014 relayed on the immutability that blockchain offers to perform its internal elections [1].

4. **Data transparency and privacy:** since blockchain counts with different reading and writing permissions, the data stored is immutable, and information is shared and decentralised. Thus, applications of Semantic Web that aim at promoting transparency with privacy policies find in blockchain the perfect technology to rely on. A clear scenario that may benefit from this is the clinical data in which data is sensitive and must follow strict privacy policies [26].

5. **Crowdsourcing data:** a large number of Semantic Web applications rely on the participation of external entities, humans or machines. Public blockchains bring the perfect technology for this kind of applications, since external entities will be able to publish data and at the same time, provenance, trust, and transparency will be guaranteed.

4   Semantic Web and Blockchain Scenarios

The Semantic Web consists of a set of technologies that focus on data, i.e., how is described, modelled, and linked. On the other hand, the blockchain is a technology that aims at storing data and share such data among a set of peers, who validate the content without the intervention of a third-party services. It is clear that blockchain will benefit by storing or expressing its data using the Semantic Web technologies, and on the other way around, the Semantic Web will benefit from the blockchain due to the decentralisation and the data immutability that this technology offers.

As far as we know, there is only one article that address some preliminary ideas about how to combine these technologies, i.e., Ugarte [27], who presented three scenarios in which Semantic Web and blockchain could be combined. Starting from this article and analysing the state of the art, we have identified a total of six different scenarios, which we explain in the following sub-sections.

4.1   Blockchain with Semantic Meta-data

This scenario is the first step to integrate Semantic Web and blockchain technologies. This show-case depicted in Fig. 1 consists of a chain of blocks in which the meta-data is expressed following an ontology, and the content of the blocks is expressed in a non-RDF format. Some ontologies have been proposed for this purpose [28], however none is a standard yet.
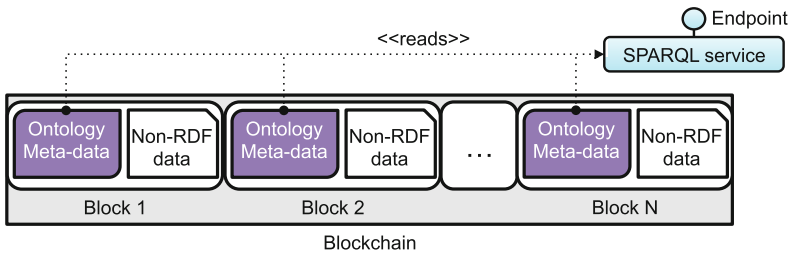

Fig. 1. Blockchain with meta-data referencing an ontology

The main benefit of this scenario is that practitioners may perform search queries considering the meta-data of the blocks. However this benefit requires to have an external service that reads the blockchain and is able to process SPARQL queries. For instance, with this benefit an user could search all the blocks which hash follows a provided regular expression.

The main drawback of this approach is that only the meta-data of the blocks is expressed using Semantic Web technologies. Therefore, in this scenario executing SPARQL queries over the meta-data is the only feasible benefit from our list.

4.2 Blockchain with RDF Content

In this show-case the approach to combine Semantic Web and blockchain relies on storing data in the blocks using RDF, as shown by Fig. 2. This scenario is complementary to the previous one. The content of the blocks may be expressed in any format that RDF supports, e.g., JSON-LD or XML/RDF.
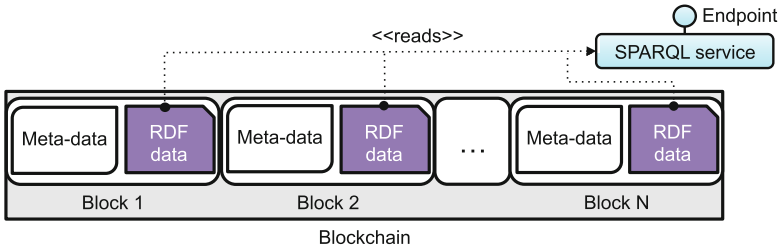

Fig. 2. Blockchain with RDF data

Assuming there is an external service that reads the blockchain and is able to process SPARQL queries, then, in this scenario the blockchain obtains all the benefits of our list. On the one hand, practitioners can execute SPARQL queries over the content of the blocks and their meta-data, which were described using an ontology. On the other hand, in case of having links to other data sources or ontologies these links should be stored the chain.

The main drawback of this scenario has not been studied yet, as far as we know. RDF formats like RDF/XML, JSON-LD, or TURTLE, are very verbose and require large number of characters. On the contrary, the amount of data that can be stored in each block is limited to a small amount of characters. As a result, using RDF entails that the chain will contain a larger amount of blocks to express the same information that could be expressed with a non-RDF format. Having a large number of blocks that contain large amount of data may drop the blockchain efficiency. As far as we know, authors have not presented a research work that relies this scenario or analyses its feasibility in terms of efficiency.

4.3 Blockchain and Virtual RDF

This scenario consists in a blockchain and a virtual RDF service. Virtualisation services take as input a data source, i.e., the blockchain, and generate RDF as depicted by Fig. 3. Some services publish the data as a dataset and count with a SPARQL query endpoint, others only generate an RDF dump that must be stored in a triple store in order to query the data.

This approach counts with all the benefits that we reported without the problem that entails storing directly RDF in the blockchain. Most of the virtual RDF services offer the capability of linking data and combining several data sources. Therefore, links between data can be generate on the fly, or stored in another
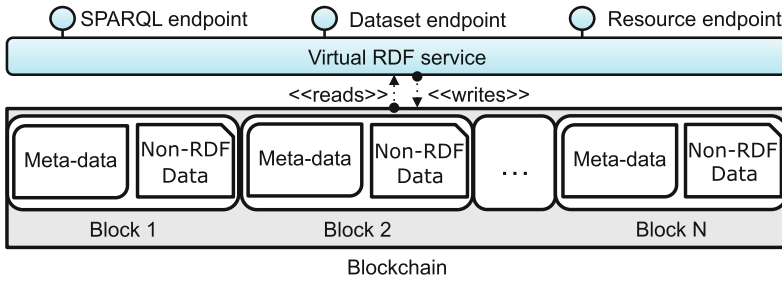
Fig. 3. Blockchain with a Virtual RDF Service to publish its content

data source and combined with the RDF virtual data from the blockchain. Something similar can be done with the ontology mappings.

The main drawback of this approach is that requires to rely on a third-party service to generate virtual RDF. As far as we know, authors have not presented a research work that presents the results of any third-party service that generates virtual RDF from a blockchain.

4.4   Blockchain with External Pointers

In this scenario there is a blockchain and an RDF dataset as depicted by Fig. 4. The bottom line is to rely on the blockchain to uniquely identify fragments of data from the RDF dataset, avoiding in this case the DNS problem. Set of triplets from the RDF dataset that share the same subject will be related to a hash from the blockchain [29], which will be an alternative identifier independent from the DNS used to identify the URI of such subjects.
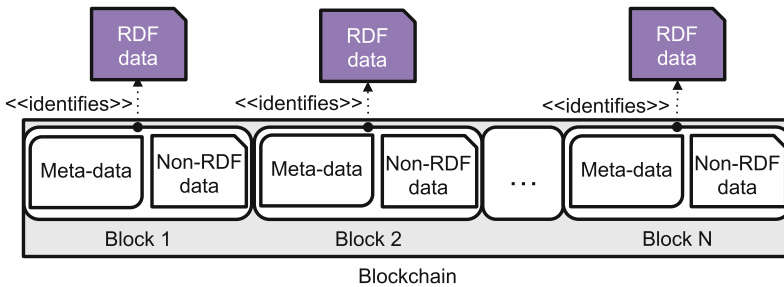

Fig. 4. Blockchain providing DNS-free identifiers for RDF fragments

In this scenario the blockchain does not obtain any benefit from the Semantic Web technologies. On the other hand, the Semantic Web technology is the one that benefits from the blockchain. The RDF data in this scenario has an alternative identifier that is independent from the DNS, entailing that resources are uniquely identified even if the DNS changes over time.

4.5   Blockchain Referencing Another Blockchain

In this scenario there are two blockchains as depicted by Fig. 5. One chain is used to identify RDF resources that are stored in the other chain following any of the approaches reported in this section. As a result, in this scenario RDF data will be immutable, transparent, and double identified (by the URIs and the hashes of the first blockchain).

In this scenario the Semantic Web counts with all the benefits that we reported from the blockchain.
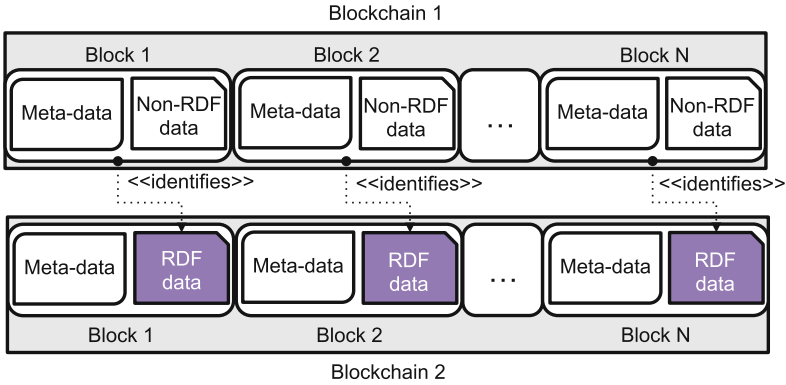

Fig. 5. Blockchain providing identifying the RDF stored in another chain

4.6   Semantic Blockchain

This scenario consists of a forked blockchain implementation that is meant to use Semantic Web technologies from the beginning, as depicted in Fig. 6. As far as we known, there is no such implementation yet but considering the relevance of the benefits that Semantic Web offers to the blockchain and vice-versa is likely that one implementation will be proposed.

This approach will count with all the benefits that we reported, the ones that the Semantic Web offers to blockchain, and the other way around.
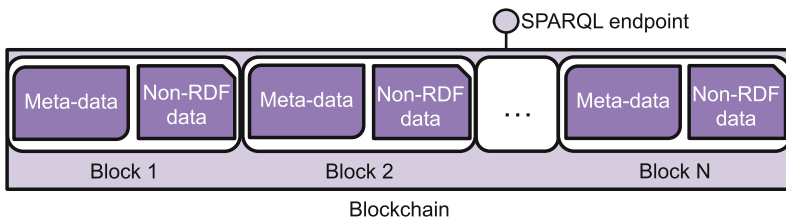

Fig. 6. Blockchain implementation relaying on Semantic Web

5   Conclusions

Recently blockchain as become a relevant technology to solve a wide-range of problems in different domains. Some researchers have expressed their interest in combining blockchain with Semantic Web, since the former offers special features to store data, and the latter is used to model and publish data. In this paper we analysed the benefits that blockchain may offer to the Semantic Web technologies, and vice-versa. In addition, we reported six different scenarios that show how these two technologies can be combined, considering which benefits they will gain and which drawbacks will have.[1]

References

1. Pilkington, M.: 11 blockchain technology: principles and applications. Res. Handb. Digit. Transform. **225**, 225–253 (2016)
2. Zheng, Z., Xie, S., Dai, H.N., Wang, H.: Blockchain challenges and opportunities: A survey. Work Pap.-2016 (2016)
3. Lin, I.C., Liao, T.C.: A survey of blockchain security issues and challenges. IJ Netw. Secur. **19**(5), 653–659 (2017)
4. Ruta, M., Scioscia, F., Ieva, S., Capurso, G., Di Sciascio, E.: Semantic blockchain to improve scalability in the internet of things. Open J. Internet Things (OJIOT) **3**(1), 46–61 (2017)
5. Panarello, A., Tapas, N., Merlino, G., Longo, F., Puliafito, A.: Blockchain and iot integration: a systematic survey. Sensors **18**(8), 2575 (2018)
6. Sikorski, J.J., Haughton, J., Kraft, M.: Blockchain technology in the chemical industry: machine-to-machine electricity market. Appl. Energ. **195**, 234–246 (2017)
7. English, M., Auer, S., Domingue, J.: Block chain technologies & the semantic web: a framework for symbiotic development. In: Lehmann, J., Thakkar, H., Halilaj, L., Asmat, R. (eds.) Computer Science Conference for University of Bonn Students, pp. 47–61 (2016)
8. Bartoletti, M., Pompianu, L.: An analysis of bitcoin op_return metadata. FC 2017. LNCS, vol. 10323, pp. 218–230. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70278-0_14
9. Wood, G., et al.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj. Yellow Pap. **151**, 1–32 (2014)
10. Maly, R.J., Mischke, J., Kurtansky, P., Stiller, B.: Comparison of centralized (client-server) and decentralized (peer-to-peer) networking. Semester thesis, ETH Zurich, Zurich, Switzerland, pp. 1–12 (2003)
11. Berners-Lee, T., Hendler, J., Lassila, O., et al.: Semant. Web. Scientific american **284**(5), 28–37 (2001)
12. Brickley, D., Guha, R.V., McBride, B.: RDF schema 1.1. W3C recomm. **25**, 2004–2014 (2014)

---

[1] https://www.delta-h2020.eu/.

13. Lefrançois, M., Zimmermann, A., Bakerally, N.: A SPARQL extension for generating RDF from heterogeneous formats. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., Hartig, O. (eds.) ESWC 2017, Part I. LNCS, vol. 10249, pp. 35–50. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58068-5_3
14. Dimou, A., Vander Sande, M., Colpaert, P., Verborgh, R., Mannens, E., Van de Walle, R.: RML: A generic language for integrated RDF mappings of heterogeneous data. In: LDOW (2014)
15. Bizer, C., Heath, T., Berners-Lee, T.: Linked data: the story so far. In: Semantic Services, Interoperability and Web Applications: Emerging Concepts, pp. 205–227. IGI Global (2011)
16. McGuinness, D.L., Van Harmelen, F., et al.: Owl web ontology language overview. W3C Recomm. **10**(10), 2004 (2004)
17. Nejdl, W., Wolpers, M., Capelle, C., Wissensverarbeitung, R., et al.: The RDF schema specification revisited. In Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik, Modellierung 2000 (2000)
18. Vandenbussche, P.Y., Atemezing, G.A., Poveda-Villalón, M., Vatant, B.: Linked open vocabularies (LOV): a gateway to reusable semantic vocabularies on the web. Semant. Web **8**(3), 437–452 (2017)
19. Uschold, M.: Achieving semantic interoperability using RDF and OWL-v4, 2005 (2013)
20. Harris, S., Seaborne, A., Prud'hommeaux, E.: Sparql 1.1 query language. W3C Recomm. **21**(10), 778 (2013)
21. Buil-Aranda, C., Arenas, M., Corcho, O.: Semantics and optimization of the SPARQL 1.1 federation extension. In: Antoniou, S., et al. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 1–15. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21064-8_1
22. Knublauch, H., Kontokostas, D.: Shapes constraint language (SHACL). W3C Candidate Recomm. **11**(8) (2017)
23. Nentwig, M., Hartung, M., Ngonga Ngomo, A.C., Rahm, E.: A survey of current link discovery frameworks. Semant. Web **8**(3), 419–436 (2017)
24. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
25. Sidoroff, T., Hyvönen, E.: Semantic e-goverment portals-a case study. In: Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05, vol. 7 (2005)
26. Nugent, T., Upton, D., Cimpoesu, M.: Improving data transparency in clinical trials using blockchain smart contracts. F1000Research **5** (2016)
27. Ugarte, H.: A More Pragmatic Web 3.0: Linked Blockchain Data. Bonn, Germany (2017)
28. Pfeffer, J., et al.: Ethon-an Ethereum Ontology (2016)
29. García-Barriocanal, E., Sánchez-Alonso, S., Sicilia, M.-A.: Deploying metadata on blockchain technologies. In: Garoufallou, E., Virkus, S., Siatri, R., Koutsomiha, D. (eds.) MTSR 2017. CCIS, vol. 755, pp. 38–49. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70863-8_4