



Article

Building Semantic Knowledge Graphs from (Semi-)Structured Data: A Review

Vetle Ryen ¹, Ahmet Soylu ^{2,*} and Dumitru Roman ^{3,*}¹ Department of Informatics, University of Oslo, Gaustadalléen 23B, 0373 Oslo, Norway; vetlery@ifi.uio.no² Department of Computer Science, OsloMet—Oslo Metropolitan University, Pilestredet 35, 0167 Oslo, Norway³ Sustainable Communication Technologies, SINTEF AS, Forskningsveien 1, 0373 Oslo, Norway

* Correspondence: ahmet.soylu@oslomet.no (A.S.); dumitru.roman@sintef.no (D.R.)

Abstract: Knowledge graphs have, for the past decade, been a hot topic both in public and private domains, typically used for large-scale integration and analysis of data using graph-based data models. One of the central concepts in this area is the Semantic Web, with the vision of providing a well-defined meaning to information and services on the Web through a set of standards. Particularly, linked data and ontologies have been quite essential for data sharing, discovery, integration, and reuse. In this paper, we provide a systematic literature review on knowledge graph creation from structured and semi-structured data sources using Semantic Web technologies. The review takes into account four prominent publication venues, namely, *Extended Semantic Web Conference*, *International Semantic Web Conference*, *Journal of Web Semantics*, and *Semantic Web Journal*. The review highlights the tools, methods, types of data sources, ontologies, and publication methods, together with the challenges, limitations, and lessons learned in the knowledge graph creation processes.

Keywords: Semantic Web; linked data; knowledge graphs; structured data; semi-structured data



Citation: Ryen, V.; Soylu, A.; Roman, D. Building Semantic Knowledge Graphs from (Semi-)Structured Data: A Review. *Future Internet* **2022**, *14*, 129. <https://doi.org/10.3390/fi14050129>

Academic Editors: Vijayakumar Varadarajan, Rajanikanth Aluvalu and Ketan Kotecha

Received: 31 March 2022

Accepted: 20 April 2022

Published: 24 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Knowledge graphs use graphs as the underlying model for representing data, and they are typically used for large-scale data integration and analysis, although the exact meaning of the term is debatable (see e.g., [1–3] for further discussion around definitions of knowledge graphs). Knowledge graphs have become a popular concept due the development of a new generation of Web and Enterprise applications (where data needed to be integrated in a more simplified way), advances in NoSQL graph databases (where graph data could be stored and managed at scale), and enhanced learning (where graph data structures have been proved to improve machine learning techniques). Knowledge graphs have witnessed attraction both in the scientific community and industry [4]. They are particularly used in large organisations to break down data silos and make data more accessible by lowering the bar for business analysts to perform advanced data retrieval and analysis. Companies such as Google, Amazon, and Facebook utilise machine learning and graph analytics over knowledge graphs to improve their core products, for example for providing better search results and product recommendations.

The idea of integrating data using a graph format, and eventually extracting knowledge from it, existed for many decades and led to various graph-based data models, such as directed edge-labelled graphs, heterogeneous graphs, and property graphs [2]. The Resource Description Framework (RDF), which is the data model for the Semantic Web, is based on directed edge-labelled graphs. The use of Semantic Web technologies for creating knowledge graphs both for industrial (e.g., [5,6]) and public data (e.g., [7,8]) is a natural and active line of research. This is because, firstly, the knowledge graph paradigm encompasses many of the core and long-standing ideas of the Semantic Web domain [9]. Secondly, with the use of ontologies and linked data technologies, the Semantic Web utilises an open approach based on a wide range of standards and formal logic for representing, integrating,

sharing, accessing, and reasoning over the data. Knowledge in a semantic knowledge graph could be exploited in various ways using deductive and inductive approaches. Deductive approaches concern extracting knowledge using entailment and reasoning through logical axioms and rules. RDFS and OWL enrich RDF with formal semantics, making deductive knowledge extraction possible [10]. Inductive approaches concern deriving knowledge through analysing generalised patterns in a knowledge graph. These include graph analytics, embedding, and graph neural networks. Graph analytics uses techniques such as centrality, community, connectivity, and node similarity [11,12] and also utilises graph query languages [13]. In embedding, a knowledge graph is embedded into a vector space so that it can be used for various machine learning tasks such as training for classification, regression, recommendation, etc. [14], while the graph neural network approaches model a neural network based on the topology of the knowledge graph [15].

Creating and maintaining knowledge graphs is not a trivial task, where several aspects need to be taken into consideration such as data modelling, transformation, and reconciliation [3]. With respect to the purpose of the implementation, the actors involved, the domain, the data sources, etc., pipelines used for knowledge graph creation could greatly vary. In the Semantic Web domain, the overall process usually includes mapping source data onto an ontology/schema, translating it to RDF format, and subsequently publishing the resulting data through APIs. Organisations store considerable amounts of data in (semi-)structured format, such as in relational databases, CSV files, etc., and publish data on the Web in other (semi-)structured formats, such as XML, JSON, etc. These require mapping languages and engines to transform [16], integrate, and feed data into knowledge graphs, while for unstructured data, such as free text and PDF documents, natural language processing and information extraction techniques are required for knowledge graph creation [17].

Since a considerable amount of industrial and public data is in (semi-)structured format, in this study (based on [18]), we explore knowledge graph creation within the Semantic Web domain, specifically from (semi-)structured data, through a systematic literature review. The review takes into account four prominent publication venues, namely, *Extended Semantic Web Conference*, *International Semantic Web Conference*, *Journal of Web Semantics*, and *Semantic Web Journal*. We highlight the challenges, limitations, and lessons learned. Our goal is to answer the following questions:

- What are the publication statistics on the state of the art techniques for knowledge graph creation from (semi-)structured data?
- What are the key techniques, and associated technical details, for the creation of knowledge graphs from (semi-)structured data?
- What are the main limitations, lessons learned, and issues of the identified knowledge graph construction techniques?

The rest of this paper is organised as follows. In Section 2, related work is presented, while Section 3 sets the background. Section 4 introduces the method and execution of the study and Section 5 presents the results. Finally, Section 6 discusses the results, while Section 7 concludes the paper.

2. Related Work

In this paper, we focus on research related to knowledge graph creation and publication within the Semantic Web domain, while other aspects, such as knowledge graph refinement [19], embedding [14], querying [13], and quality [20] fall outside the scope of this paper.

Pereira et al. [21] provide a review of linked data in the context of the educational domain. The study highlights the tools, vocabularies, and datasets being used for knowledge graphs. For tools, they find that the D2RQ platform [22] is the most used for mapping data to RDF, and Openlink Virtuoso [23] and Sesame [24] (now RDF4J [25]) are the most frequently used tools for storage. Regarding the vocabularies, the Dublin Core vocabulary [26] is the most used. Some of the challenges mentioned are related to data interlinking,

data integration, and schema matching. Barbosa et al. [27] provide a review of tools for linked data publication and consumption. The study highlights the supported processes and serialisation formats and provides an evaluation. Key takeaways from the review include: most of the studies focus on the use of tools for machine access; few solutions exist for the preparation phase, including licence specification and selection of datasets for reuse; and no tool was found to support all steps of the data publication process.

Avila-Garzon [28] surveys applications, methodologies, and technologies used for linked open data. The main findings from this survey include: most of the studies focus on the use of Semantic Web technologies and tools in the context of specific domains, such as biology, social sciences, research, and libraries and education; there is a gap in research for a consolidated and standardised methodology for managing linked open data; and there is a lack of user-friendly interfaces for querying datasets. Penteado et al. [29] survey methodologies used for the creation of linked open government data, including common steps, associated tools and practices, quality assessment validations, and evaluation of the methodology. Key takeaways from this study include: phases are described with different granularity levels for the creation process, but in general, they can be classified into specification, modelling, conversion, publication, exploitation, and maintenance; there are different tools, and each tool is often only used for one phase; and, the assessment of the methodologies mostly focus on specific aspects and not the methodologies as a whole.

Other relevant related works, more focused on specific aspects, include:

- Feitosa et al. [30] provide a review of the best practices used for linked data. This study finds that the use of best practices is mostly motivated by having standard practices, integrability, and uniformity, and that the most used best practice is the reuse of vocabularies.
- Pinto and Parreiras [31] provide a review of the applications of linked data for corporate environments. The study finds that enterprises experience the same challenges as linked open data initiatives. Semantic Web technologies may be complex and require highly specialised teams.
- Ali and Warraich [32] provide a review of linked data initiatives in the library domain, and they find that there are technical challenges in the selection of ontologies and link maintenance of evolving data.

Such studies on knowledge graph creation in the context of the Semantic Web focus on a specific application domain or on specific aspect, such as tools, technologies, etc.; however, no study appears to take a generic approach (irrespective of the application domain or specific aspects) to the knowledge graph creation process.

3. Background

This section focuses on what constitutes knowledge graph creation, in the context of (semi-)structured data. The knowledge graph creation process involves several phases. Depending on different factors, such as focus, intent, data sources, actors involved, etc., different phases with varying sub-tasks need to be undertaken [2,3,33,34]. The order of phases presented in this section is not to be taken as the “right” order of action, as it is not always necessary to be finished with one phase before another one starts, and different situations may motivate different orders. Methodologies used may be based bottom-up or top-down [35] or may involve pay-as-you-go approaches [36]. The phases presented below are based on existing literature [2,3,33,34] and also based on the work carried out as part of this review.

3.1. Ontology/Schema Development

In general, an ontology can be described as a formal definition of the concepts and their relations over a given domain, ranging from simple vocabularies to complex logic-based formalisms. In the context of knowledge graphs, the term is closely related to the schema of the knowledge graph, which can be described using an ontology. In the context

of this paper, we use ontology, vocabulary, and schema interchangeably. Ontologies can be created either by defining concepts and relations through domain analysis, or through analysing the available data [2]. These are often referred to as the top-down and bottom-up approaches, respectively [37]. The top-down approach is traditionally done manually, while the bottom-up approach can be done automatically and semi-automatically.

Regarding the top-down approach, it usually starts with obtaining an overview of the given domain through reading papers and books, interviewing experts, etc. The knowledge from this phase is then used to further formalise the ontology, until a specification of the ontology is obtained. The task of formalising a given domain may be substantial, depending on the domain to be modelled. An agile methodology is therefore often employed [38]. This can be obtained using competency questions throughout the development process [39], meaning that the ontology is incrementally developed by adding concepts based on questions the ontology has to answer. Other methods include using ontology design patterns [40], which enable the reuse of existing ontology designs and modelling templates. Regarding the bottom-up approach, automatic and semi-automatic approaches are often used to extract information from the given input data in order to model the ontology. These methods also relate to the approaches for automatically integrating data into a knowledge graph. The techniques may involve measuring the relevancy of entities in the data, based on count, or relations, through patterns [41].

It is also possible to combine the two approaches by using some information from the top-down approach as a basis for the further development of an ontology with the bottom-up approach, which is often described as the middle-out approach [37]. Such approaches allow validation of intermediate results, similarly to the bottom-up approach [2].

3.2. Data Preprocessing

Preprocessing of the data is of high relevance, especially since the data may be of poor quality. There are different tasks involved in data preprocessing, including (i) enrichment, i.e., adding additional information to the data, (ii) reconciliation, i.e., correctly matching entities from different sources, and (iii) cleaning, i.e., improving the quality of the data. Rahm et al. [42] classify data quality problems in data sources, differentiating between single- and multi-source and between schema- and instance-level.

Schema-level problems refer to the overall schema for the data source and may affect several data instances at once. Single-source problems are related to the poor schema design and lack of integrity constraints. This includes uniqueness violation, illegal values, violated attribute dependencies, etc. For multi-source, the problems are related to translation and integration between schemas. Schema integration problems may occur when matching entities with respect to naming and structural features. Naming problems may occur when the same object has different names or different objects have the same name. Structural problems may be the difference in data types, different integrity constraints, etc. Instance-level problems refer to the problems happening at instance level in the data. Instance-level problems cannot be prevented at the schema level. Single source problems, at the instance level, may be spelling error, duplicates, etc. Multi-source problems, at the instance level, may be contradicting values for objects, different use of units, different use of aggregation, etc.

To resolve the mentioned problems, Rahm et al. [42] mention several phases, which include data analysis, definition of transformation workflow and mapping rules, verification, transformation, and backflow of cleaned data. Several tools exist for the execution of data cleaning, including spreadsheet software, command line interface (CLI) tools, programming languages, and complex systems designed to be used for interactive data cleaning and transformation [43,44].

3.3. Data Integration

In this section, we describe the different data integration methods that exist for integrating (semi-)structured data into knowledge graphs. Data integration methods include manual integration and mapping-based integration.

Manual integration of data into a knowledge graph involves manually defining the entities of the knowledge graph directly from the source. This can be done either by writing the code directly in the given language or through a dedicated editor [45]. Some knowledge graphs on the Web are created through manual data integration, such as Wikidata [46]. The human interaction in this data integration process comes at a high cost compared to other approaches. However, this usually ensures high quality, where each statement in the knowledge graph can be manually verified as being correctly mapped to its corresponding concept.

Mapping-based approaches allow for integrating data from (semi-)structured sources, such as relational databases: CSV, JSON, XML, etc. Data are mapped through rules onto a graph or a graph view (a view of the data source, in graph format) [47]. The mapping can be done either directly or through custom statements. Direct mapping involves mapping data directly from its source. For table-structured data, a standard direct mapping involves the creation of a triple for each non-empty, non-header cell, where the subject is represented by the row, the predicate is represented by the column, and the object is represented by the value of the cell. Direct mapping from relational databases to RDF has been standardised by the W3C [48]. The flexibility of being able to create dedicated tables through SQL queries makes this a productive approach for knowledge graph creation. Direct mapping from tree-structured data, e.g., XML or JSON, is often not desirable, since this will only produce a mirroring of the data. Custom mapping involves defining statements about how data are to be mapped from its source [49]. This enables specifying how columns and rows are to be mapped. The mapping language R2RML [50] is a W3C standard that defines mapping from relational databases to RDF. Building upon this language, other languages have been proposed for mapping from other data structures. RML [51] is one of the well-known languages for mapping data also from CSV, XML, and JSON. Other languages such as FnO [52] enable defining transformations for the data to be mapped and integrating them with the aforementioned custom mapping languages.

Mapping from other knowledge graphs is also an option. This can either be done by manually recreating the target knowledge graph or by querying the target knowledge graph to obtain a sub-graph of it. This is typically done using SPARQL CONSTRUCT queries [53]. This approach usually also includes the necessity of aligning the schema/ontology of the graphs. An aspect to consider when mapping data onto a knowledge graph is whether to fully integrate the data onto the graph model or only make a graph view of the data. Fully integrating the data, often referred to as Extract–Transform–Load (ETL), does require the data to be updated from time to time. On the other hand, Ontology-Based Data Access (OBDA) techniques based on data virtualisation [16] enable a graph view of the data without materialising it by using query rewriting techniques. This means data are kept in their original place and format, and queries are specified using ontological concepts and relationships and rewritten to the query language of the underlying database system(s).

3.4. Quality and Refinement

Knowledge graph quality and refinement refers to how a knowledge graph may be assessed and subsequently improved. There are numerous frameworks and methods for both assessing and refining knowledge graphs [3,19,20]. Zaveri et al. [20] provide a survey of approaches for evaluating and assessing the quality of linked data, identifying 18 different interlinked quality dimensions as briefly described in the following.

Accessibility quality describes the quality aspects concerning how a knowledge graph may be accessed and how well it supports the user to easily retrieve data from it:

- Availability describes the degree a knowledge graph and its contents are available for interaction (e.g., through a SPARQL endpoint).

- Licensing concerns whether a knowledge graph has a license published with it (human-readable and machine-readable format).
- Interlinking describes the degree that entities, referring to the same real-world concept, are linked to each other [54].
- Security refers to the degree the knowledge graph is secured through verification of confidentiality in the communication between the consumers and the knowledge graph itself.
- Performance relates to how well the knowledge graph may handle latency, throughput, and scalability.

Intrinsic quality refers to the quality dimensions that are internal to the knowledge graph. These dimensions refer to quality dimensions that are independent of the user context:

- Syntactic validity describes the degree to which the knowledge graph content follows syntactic rules [55].
- Semantic accuracy refers to the degree to which a knowledge graph correctly represents real-world facts semantically.
- Consistency refers to how well a knowledge graph is free of contradictions in the information contained in it.
- Conciseness describes the degree to which a knowledge graph only contains relevant information [56].
- Completeness concerns how complete a knowledge graph is in comparison to all the required information.

Contextual quality refers to quality dimensions that usually depend on the context of the implemented knowledge graph [20]:

- Relevancy describes the extent it is possible to obtain relevant knowledge from a knowledge graph for the task at hand.
- Trustworthiness refers to how trustworthy the information contained in a knowledge graph is subjectively accepted to be correct.
- Understandability refers to the extent the information contained in a knowledge graph can be used and interpreted by users without ambiguity [57].
- Timeliness refers to how well a knowledge graph is up to date based on the real-world facts [58].

Representational quality refers to the dimensions describing the design aspects of the knowledge graph:

- Representational conciseness refers to the extent to which information is concisely represented in a knowledge graph.
- Interoperability refers to the extent to which a knowledge graph represents data with respect to the existing relevant vocabularies for the subject domain [57].
- Interpretability refers to how well a knowledge graph is technically capable of providing information in an appropriate serialisation and whether a machine is capable of processing the data.
- Versatility refers to which extent a knowledge graph is capable of being used in different representations and in different languages.

Tim Berners-Lee described a five-star rating scheme [59], where datasets could be awarded stars with the following criteria: data are available on the Web under an open license (1 star) in (semi-)structured format (2 stars) and a non-proprietary open format (3 stars) using open standards from the W3C (4 stars) and linked to other data (5 stars). This was extended to a seven-star scheme [60]: data are provided with an explicit schema (6 stars), and data are validated against the schema (7 stars). There is also a five-star rating scheme for the vocabulary use with the following criteria: there is dereferencable human-readable information about the used vocabulary (1 star); the information is available as

machine-readable explicit axiomatization of the vocabulary (2 stars); the vocabulary is linked to other vocabularies (3 stars); metadata about the vocabulary are available (4 stars); the vocabulary is linked to other vocabularies (5 stars) [61].

Regarding the knowledge graph refinement, several approaches exist [19], particularly in terms of completion and error detection. These methods can be categorised as either external or internal, i.e., whether external sources are being used for the process or not. Internal methods usually use machine learning or probabilistic methods, while external methods focus on using external data, such as Wikipedia or another knowledge graph [19].

3.5. Publication

Several aspects need to be taken into consideration when publishing a knowledge graph on the Web. These include aspects around how the knowledge graph will be hosted and what data will be accessible. In the context of the Semantic Web, Heath and Bizer [62] describe different methods for publishing linked data on the Web. Linked data may be published directly through static RDF/XML files, RDF embedded into HTML files, a wrapper over existing applications or APIs, relational databases, and triplestores. In addition to having published the knowledge graph, the publisher may also provide tools for accessing the data, such as SPARQL endpoints, data dumps, and search engines. Some publishers also provide documentation pages for concepts [63], or even visualisation tools for the given graph [64].

FAIR principles (findability, accessibility, interoperability, and reusability) [65], initially proposed for the publication of scientific data, are highly relevant for knowledge graph publication. The FAIR principles ensure that data can be easily found and accessed through the Web and they can easily be explored and reused by others, and these principles include requirements such as links to other datasets, provenance and licensing metadata, and the use of widely deployed vocabularies. Often with the publication of data on the Web, it is preferable to define a licence for the data. The W3C standard vocabulary Open Digital Rights Language (ODRL) [66] allows for defining permission, prohibition, and obligation statements of data, and it can be easily integrated in an RDF serialisation format. The vocabulary allows for modelling common licences such as the Apache [67] or Creative Commons [68] licenses, and it enables a standardised format for licensing linked data on the Web.

Other aspects relevant for knowledge graph publication are URI strategy and context and versioning [2]. The identity of instances in the model is also of relevance. Having a consistent URI or IRI strategy is important for easily finding the right instance on the Web and keeping it non-changing. Facts contained in the graph may only hold within a certain context, e.g., within a specific time period or domain. There are several ways of applying context to knowledge graphs, such as through reification (adding information directly to the edges) and higher arity representations [69].

4. Method and Execution

The Semantic Web domain is currently widely diverse and largely heterogeneous in terms of applications, hindering new adopters from easily navigating and implementing its technologies. It is evident that the developments throughout the years both in the research community and also in the industry have led to a wide variety of ways in building knowledge graphs based on Semantic Web technologies (e.g., see Table 1). A systematic review of the current state of the art would help in consolidating the field through highlighting and discussing the different technologies, tools, and methods currently being used along with reported experiences.

4.1. Review Protocol

In this section, we introduce the protocol used for the review reported in this paper.

Table 1. Studies included in the review.

Nr.	Title	Domain	Year	Venue	Citation
# 1	GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data	Geography	2018	ESWC	[70]
# 2	WarSampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History	History	2016	ESWC	[71]
# 3	DCAT-AP representation of Czech National Open Data Catalog and its impact	Government	2018	JWS	[72]
# 4	Publication and usage of official Czech pension statistics Linked Open Data	Government	2017	JWS	[73]
# 5	3cixty: Building comprehensive knowledge bases for city exploration	Smart city	2017	JWS	[74]
# 6	Raising interoperability among base registries: The evolution of the Linked Base Registry for addresses in Flanders	Government	2018	JWS	[75]
# 7	NanoMine: A Knowledge Graph for Nanocomposite Materials Science	Science	2020	ISWC	[76]
# 8	Enhancing Public Procurement in the European Union Through Constructing and Exploiting an Integrated Knowledge Graph	Business	2020	ISWC	[77]
# 9	DOREMUS: A Graph of Linked Musical Works	Music	2018	ISWC	[78]
# 10	The SEPSES Knowledge Graph: An Integrated Resource for Cybersecurity	Cybersecurity	2019	ISWC	[79]
# 11	Lessons Learned in Building Linked Data for the American Art Collaborative	Art	2017	ISWC	[80]
# 12	Facilitating the Analysis of COVID-19 Literature Through a Knowledge Graph	Health	2020	ISWC	[81]
# 13	ARCO: The Italian Cultural Heritage Knowledge Graph	Culture	2019	ISWC	[82]
# 14	SemanGit: A Linked Dataset from git	Computing	2019	ISWC	[83]
# 15	Ireland's Authoritative Geospatial Linked Data	Geography	2017	ISWC	[84]
# 16	One year of the OpenCitations Corpus: Releasing RDF-based scholarly citation data into the Public Domain	Bibliography	2017	ISWC	[85]
# 17	Zhishi.lemon: On Publishing Zhishi.me as Linguistic Linked Open Data	Language	2016	ISWC	[86]
# 18	Leveraging Linguistic Linked Data for Cross-lingual Model Transfer in the Pharmaceutical Domain	Language	2020	ISWC	[87]
# 19	Linked Data Publication of Live Music Archives and Analyses	Music	2017	ISWC	[88]
# 20	The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data	Bibliography	2019	ISWC	[89]
# 21	PDD Graph: Bridging Electronic Medical Records and Biomedical Knowledge Graphs via Entity Linking	Medicine	2017	ISWC	[90]
# 22	Weather Data Publication on the LOD using SOSA/SSN Ontology	Climate	2020	SWJ	[91]
# 23	The Rijksmuseum collection as Linked Data	Culture	2018	SWJ	[92]
# 24	The apertium bilingual dictionaries on the web of data*	Language	2018	SWJ	[93]
# 25	The ACORN-SAT linked climate dataset	Climate	2017	SWJ	[94]
# 26	LinkedSpending: OpenSpending becomes Linked Open Data	Business	2016	SWJ	[95]
# 27	The Open University Linked Data—data.open.ac.uk	Education	2016	SWJ	[96]
# 28	EventMedia: A LOD dataset of events illustrated with media	Media	2016	SWJ	[97]
# 29	Publishing DisGeNET as nanopublications	Medicine	2016	SWJ	[98]
# 30	Meta-data for a lot of LOD	Computing	2017	SWJ	[99]
# 31	CEDAR: The Dutch historical censuses as Linked Open Data	History	2017	SWJ	[100]
# 32	DM2E: A Linked Data source of Digitised Manuscripts for the Digital Humanities	Culture	2017	SWJ	[101]
# 33	Migration of a library catalogue into RDA linked open data	Bibliography	2018	SWJ	[102]
# 34	A Linked Data wrapper for CrunchBase	Media	2018	SWJ	[103]
# 35	Linked Web APIs dataset	Computing	2018	SWJ	[104]
# 36	The debates of the European Parliament as Linked Open Data	Government	2017	SWJ	[105]

4.1.1. Scope

The following two conferences and two journals have been selected as the main sources of the review: *Extended Semantic Web Conference (ESWC)* [106], *International Semantic Web Conference (ISWC)* [107], *Journal of Web Semantics (JWS)* [108], and *Semantic Web Journal (SWJ)* [109]. These are well-known venues, with high citation counts and rigorous review processes, specifically within the Semantic Web domain and are known to publish in-use and resource papers, which fall within the scope of this paper.

4.1.2. Selection Criteria

Inclusion and exclusion criteria were developed to make sure that only studies answering the questions raised were included. These were defined both before and during the search, since the scope of the research became clearer as the study progressed.

The inclusion criteria (IC) include:

1. The primary study should explain the creation of a specific knowledge graph as an entity.
2. The primary study should describe knowledge graph creation based on (semi-)structured sources.
3. The primary study should be published between January 2015 and January 2021.

Exclusion criteria (EC) include:

1. Primary studies not written in English are excluded.
2. The primary studies not describing the creation of a knowledge graph in the context of the Semantic Web are excluded.

4.1.3. Data Extraction Forms

Data extraction forms [110] were developed to extract data more precisely from the selected studies and to store additional data, such as title, year, study number, etc. The form was refined through the iterations of the review, to match the altering research questions. The fields used in the extraction process include the following:

- Study Number: reference number for the study.
- Title: title of the study.
- Author(s): author or authors of the study.
- Year: year of publication in the respective venue.
- Venue: venue the study is published in.
- Contributions the study proposed: described contributions in the study, including the knowledge graph.
- The application domain of the knowledge graph: application domain of the described knowledge graph.
- Phases of the construction process described: phases of the construction process described in the study.
- Tools and methods: tools and methods used and/or created in the study.
- Data types used as source: type of data used as a source.
- Ontologies: ontologies developed and/or used.
- Publication: the way knowledge graph is published.
- Semantic Web best practices: Semantic Web best practices used.
- Links to other Semantic Web graphs: other graphs linked by the constructed knowledge graph.
- Number of triples: reported number of triples in the knowledge graph.
- Use cases: reported use of the created knowledge graph.
- Assessment and evaluation: reported evaluation and assessment of the created knowledge graph.
- Limitations, errors, and lessons learned: reported limitations, errors, or lessons learned through the creation and use of the created knowledge graph.

4.2. Studies Selected

We first skimmed through the titles and abstracts of the published papers in the mentioned venues and then scanned the selected papers. The number of papers after each of these phases are:

- ISWC: Initially: 351—After Title, Abstract and Skimming: 35—After Scanning: 15
- ESWC: Initially: 234—After Title, Abstract and Skimming: 19—After Scanning: 2
- SWJ: Initially: 308—After Title, Abstract and Skimming: 27—After Scanning: 15
- JWS: Initially: 177—After Title, Abstract and Skimming: 15—After Scanning: 4

At the end, 36 papers were included in the review, and these are listed in Table 1.

5. Results

In this section, the results from the review are presented, highlighting the key take-aways. The publication statistics are shown in Figure 1, depicting the publication counts from each year. The numbers vary across the years. It can be seen that there is a clear high in 2017, with a low in 2019. This is in line with the growth in number of RDF graphs in the linked open data cloud over time as presented by Hitzler [9]. We also see that there is a minor increase from 2019 to 2020.

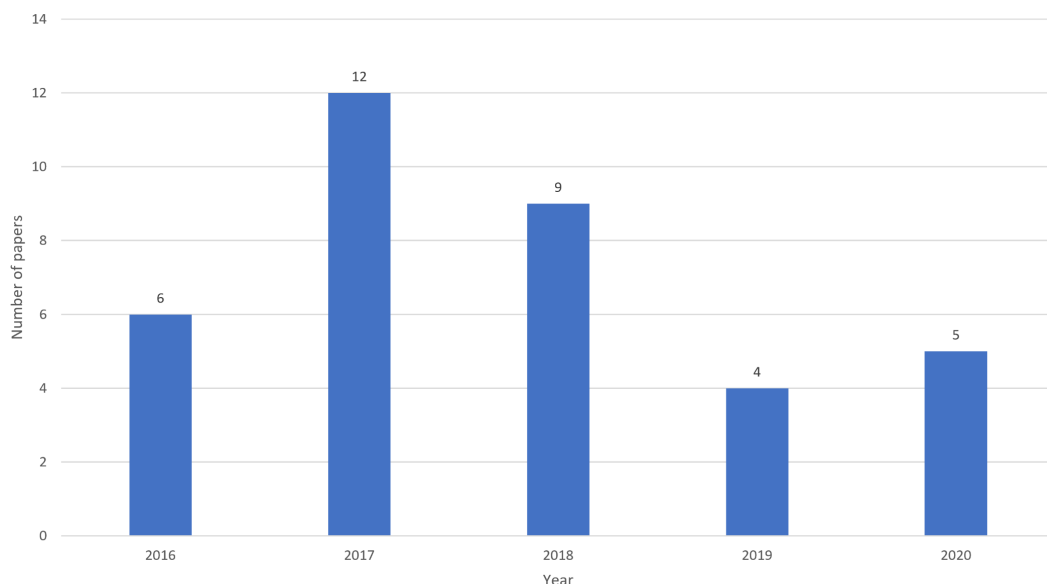


Figure 1. Number of papers per year.

The number of papers per venue is shown in Figure 2. There is a clear gap between two groups of venues, namely between ISWC and SWJ publications on one hand, and ESWC and JWS publications on the other hand. Even though the sample size and variance are not of great consideration, it may give an indication on the focus of each venue.

Regarding the domains, we see that domains such as history, culture, government, medicine, and bibliography provide a comparatively higher number of studies, while there is a spectrum of domains including art, education, climate, media (see Table 1). This indicates that research on knowledge graphs for the Semantic Web may be more popular for the public sector compared to the industry. One reason for this could be that industry could be reluctant to share data openly, while published industrial research focuses around the tools and processes (e.g., [5,6]) rather than the knowledge graphs themselves as the main contributions.

5.1. Technical Analysis

In this section, we provide an analysis of the selected works from a technical perspective in terms of their contributions, phases undertaken, and resources used.

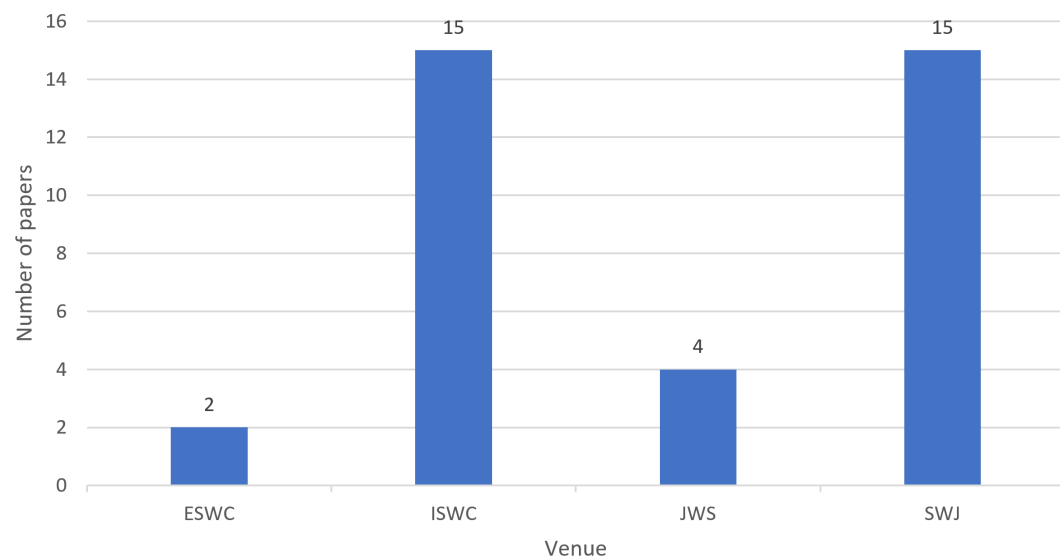


Figure 2. Number of papers per venue.

5.1.1. Contributions

The main contribution for each paper is the publication of a dedicated knowledge graph, often together with other artefacts developed (e.g., tools and ontologies). Some works include additional contributions to achieve their goal in their respective domains. These contributions include:

- The work by Achichi et al. [78] (#9) discusses contributions in terms of tools made for evaluation, data generation, construction, linking, and alignment.
- The work by Kiesling et al. [79] (#10) discusses contribution in terms of a specific ETL framework to build knowledge graphs.
- The work by Knoblock et al. [80] (#11) discusses contributions in terms of tools developed to support the processes of mapping and validation.

It should also be mentioned that several of the other papers do contribute with other specific tools. For example, Soylu et al. [77] (#8) provide a platform to explore and utilise the given knowledge graph. Pinto et al. [103] (#34) describe a Java application, developed for the data mapping in a specific project. It is also published as open source code to be modified and used. Even though the count of studies contributing additional tools is low, it does indicate that there is still room for development and that existing tools do not always seem fit.

Regarding the sizes of the knowledge graphs, 12 studies out of the 36 do not provide information on the size of the presented knowledge graphs. Table 2 presents the sizes of the knowledge graph for each study. We see that [83] (#14) is the study reporting the biggest knowledge graph, containing data from GitHub, while [105] (#36) reports the lowest, containing metadata about Web API's. The size difference highlights the applicability of knowledge graph technologies. Regarding the quality, only 19 of the studies report on knowledge graph quality based on the five to seven-star scheme, while only five of them report on the quality of the vocabulary (see Table 2).

5.1.2. Phases

We earlier categorised phases for knowledge graph construction into ontology development, data preprocessing, data integration, quality and refinement, and data publication. During the review process, we collected various tasks that fall under these categories. Table 3 presents these tasks, the number of studies using them, studies including one or more of these tasks, and the phases they belong to.

Table 2. Quality and size of the resulting knowledge graphs.

Paper	Data Quality	Vocabulary Quality	Number of Triples (ca)
#1	5 stars	-	37 millions
#2	6 stars	-	7.1 millions
#3	5 stars	-	7.4 millions
#4	5 stars	-	1.8 millions
#5	-	-	-
#6	-	-	-
#7	5 stars	-	-
#8	-	-	-
#9	5 stars	-	16 millions
#10	-	-	36 millions
#11	7 stars	-	9.7 millions
#12	-	-	-
#13	-	-	169 millions
#14	-	-	21 billions
#15	5 stars	-	-
#16	-	-	-
#17	-	-	7 millions
#18	-	-	-
#19	-	-	12 millions
#20	5 stars	-	8 billions
#21	-	-	2.3 millions
#22	5 stars	-	3.7 millions
#23	5 stars	-	22 millions
#24	5 stars	-	-
#25	-	-	61 millions
#26	-	-	113 millions
#27	5 stars	-	3.5 millions
#28	-	-	30 millions
#29	-	-	31 millions
#30	-	4 stars	110 millions
#31	5 stars	-	-
#32	5 stars	4 stars	-
#33	5 stars	-	15 millions
#34	5 stars	4 stars	83 millions
#35	5 stars	4 stars	0.5 millions
#36	5 stars	4 stars	-
Total	19	5	24

Although there are several tasks mentioned for each phase, these are often sparsely used as seen from Table 3. Ontology reuse, URI/IRI strategy, data linking, RDF transformation, and publication are the top mentioned tasks, as these could be seen among the common tasks; however, we also see other relevant phases such as evaluation and versioning, although these are not as frequent. This discrepancy could be due to the nature of the problem and solution and the extent of the work undertaken. For example, the work by Buyle et al. [75] (#6) uses a virtualisation-based approach; hence, there is no need for RDF transformation; this could apply to other tasks such as ontology modelling, data cleaning, and enrichment. A lack of tasks such as evaluation, versioning, and validation, however, is largely connected to the focus and completeness of the work.

The tasks listed are self-explanatory to a large extent. Enrichment refers to the completion of missing information or improving the precision of information available in the knowledge graph through external and internal data using knowledge graph completion techniques [111]. Data linking, in this paper, is a broader term referring to linking entities in the same knowledge graph or with entities in other graphs, and matching entity information against a set of canonical entities (i.e., reconciliation). Validation is a task used to check mostly the semantic and syntactic validity of the knowledge graph. For exam-

ple, SHACL [112] is used to validate the shape of a knowledge graph, while ontological reasoning is used to check the logical consistency of a knowledge graph.

Table 3. Phases and tasks for knowledge used for graph construction.

Phases	Total	Papers	Tasks
Ontology development	36	all	Ontology modelling (1), ontology reuse (35), ontology linking (1), URI/IRI strategy (28)
Data preprocessing	16	#3–#5, #8–#11, #13, #16, #22, #26–#28, #30, #31, #33	Data cleaning (14), conversion (2), refactoring (1), enrichment (5)
Data integration	36	all	Mapping (36), transformation (35), virtualisation (1), data linking (22)
Quality and refinement	4	#3, #4, #7–#9, #11, #13–#15, #17, #18, #24, #26, #32–#35	Evaluation (17), validation (3)
Data publication	36	all	Publication (36), update (2), graph versioning (4)

5.1.3. Resources

There are in total seven different data source types mentioned, as shown in Figure 3. There is a diverse set of data source types. The XML format is the most common type of source, indicating that much of the data may be taken from sources available on the Web. We also see JSON, CSV, and relational data are much utilised. The JSON format is typically seen in API sources and may be an easy target for knowledge graph transformation without needing much preprocessing.

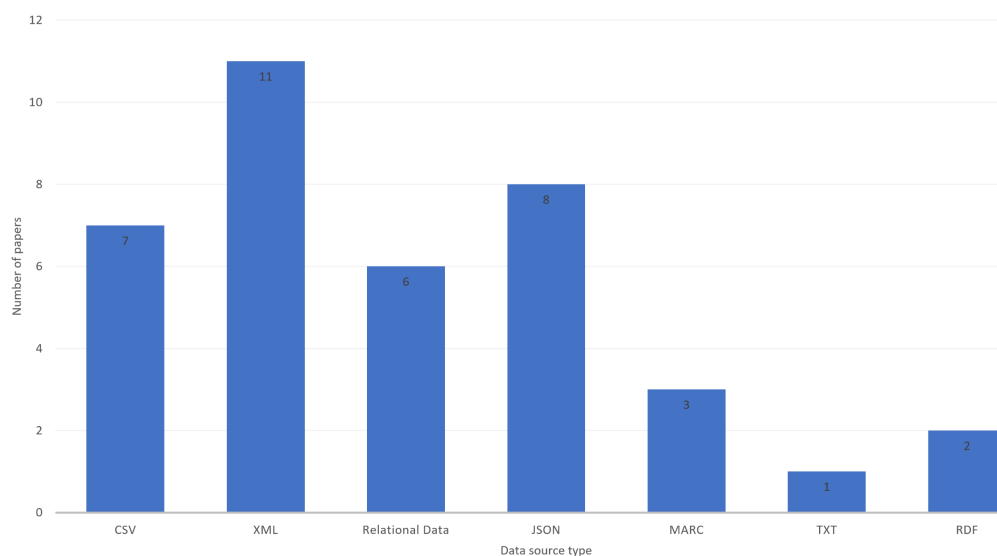


Figure 3. Data source types used.

Regarding the tools used, an account is given in Table 4. We observe that most of the listed artefacts are mentioned once and show a discrepancy in the creation of knowledge graphs; a large portfolio of tools are available (see Dominik et al. [113] for knowledge graph generation tools from XML and relational databases). Useful tools such as SWI Prolog [114] and languages such as SHACL are scarcely referenced. Heavily referenced tools include the triple stores Jena TDB (and Fuseki) [115] and OpenLink Virtuoso [23], together with other tools and languages such as SILK [116] for linking data, Pubby [117] for publishing, and XLST [118] as a mapping language. In addition to the two mentioned databases, it appears that there is a wide variety of needs and preferences for different knowledge graphs.

Considering the ontologies and vocabularies used, an account is given for the most used ontologies in Table 5. We observe there is a frequent use of metadata ontologies for describing the different datasets. This may probably be due to the fact that the creators of the knowledge graphs are actually interested in following the good practice of describing

their data, but it may also be because it becomes more relevant when the data contained come from another source. We also see that on the other side, reusing already defined classes and properties, from, e.g., Schema.org or DBpedia, is not common.

Table 4. Most used tools.

Tool	Total	Explanation	Papers
Jena TDB and Fuseki	9	Storage, publication	#2, #4, #8, #10, #21, #22, #27, #32, #33
OpenLink Virtuoso	8	Storage, publication, RDF virtualisation	#3, #4, #6, #24, #25, #26, #28, #35
XSLT	4	Data mapping, transformation	#18, #23, #25, #32
SILK	4	Reconciliation, graph linking	#5, #11, #28, #32
Pubby	4	KG publication (front-end interface)	#19, #24, #31, #32
OpenRefine	3	Reconciliation, mapping, transformation	#5, #8, #24
RML	2	Mapping, transformation	#8, #12
ELDA	2	KG publication (API)	#25, #28
D2RQ	2	Data mapping, virtualisation (relational)	#25, #29

Table 5. Most used ontologies and vocabularies.

Ontology/Vocabulary	Total	Explanation	Papers
VoID	11	Expressing metadata about RDF datasets	#1, #4, #14, #15, #16, #19, #20, #25, #30, #32, #36
SKOS	11	Representing knowledge organization systems	#2, #4, #8, #9, #19, #25, #27, #28, #29, #32, #33
Dublin Core	11	Describing metadata about resources	#2, #5, #8, #20, #23, #24, #27, #28, #32, #35, #36
PROV-O	9	Representing provenance information	#7, #15, #16, #19, #29, #30, #31, #35, #36
DCAT-AP	6	Describing public sector datasets in Europe	#3, #4, #16, #24, #30, #31
CIDOC CRM	5	Describing cultural heritage domain	#2, #11, #13, #28, #32
FOAF	5	Describing individuals and social networks	#8, #27, #28, #33, #36
RDF DATA Cube	4	Publishing multi-dimensional data, such as statistics	#3, #4, #26, #31
FRBR	4	Describing bibliographic domain	#9, #12, #13, #20
DOLCE+DnS	4	Describing natural language and human common sense	#5, #13, #25, #28
Schema.org	3	Describing common entities, such as people	#5, #8, #27
DBpedia	3	Describing common entities, such as people	#20, #26, #33
BIBO	3	Describing bibliographic things	#7, #27, #32

Regarding the specific methodologies used in creation of the knowledge graphs, there were few examples of reusing existing frameworks, both for the whole creation process and also just for the ontology development, even though there are many frameworks available. From all of the 36 chosen primary studies, only Carriero et al. [82] (#13) mention using the eXtreme Design methodology in development of the ontology. Other than this, some studies, such as Achichi et al. [78] (#9), mention the development of tools for the creation process as a framework but do not label it as a dedicated methodology.

5.2. Adoption

In this section, we present how the resulting knowledge graphs are published and exploited as well as the limitations, lessons learned, and issues reported in the studies.

5.2.1. Publishing and Exploitation

A SPARQL endpoint is the most used publication method, but we also see that the majority of the publications provide a front-end (see Figure 4). However, an interesting observation is that even though knowledge graphs do provide a unique way for data aggregation and analysis, very few of the papers provide autogenerated analytics for common metrics and data views. Even fewer provide any type of tutorial to utilise the knowledge graph. These aspects do limit the utilisation of the potential of knowledge graph features. In Figure 5, we have an overview of some of the most used external knowledge graphs and types of knowledge graphs that have been linked to (not including the ones with the smallest count). DBpedia appears the most popular knowledge graph.

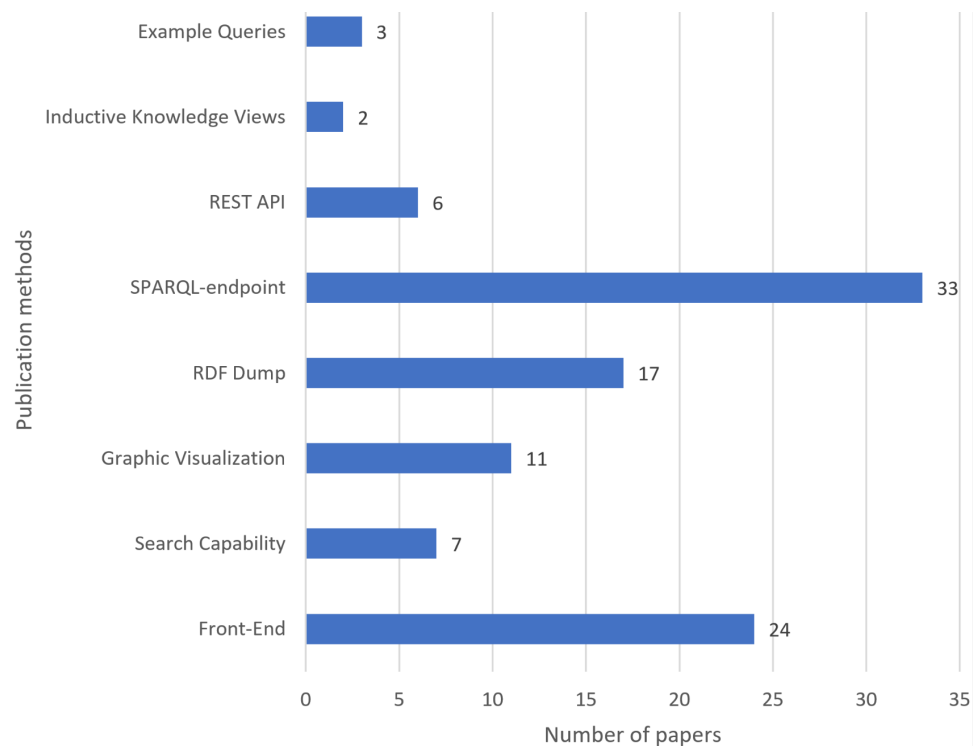


Figure 4. Publication methods.

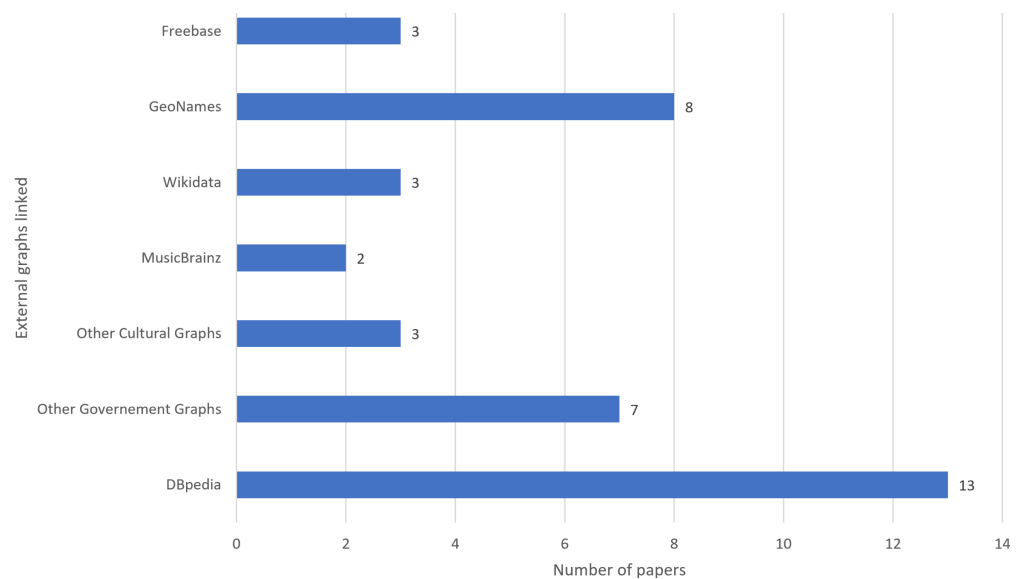


Figure 5. Links to other graphs.

Figure 6 provides an overview on how knowledge graphs are evaluated in the reviewed works. We see that linking quality to other graphs is of high interest, together with usability. Based on the technical and semantic capabilities of knowledge graphs, it is interesting to see that metrics such as correctness, graph coverage, and ontology coverage are infrequently mentioned.

Key observations regarding the employed use cases include:

- Only two studies [77,78] (#8 and #9) provide descriptions of use cases with advanced analytics, utilising the semantic and technical potential of knowledge graphs.
- Six studies provide sample queries for the respective knowledge graph.
- Twelve studies mention a use case from an external party.
- Seven studies mention use cases done by the authors themselves.

- Four studies mention potential use cases, without having an actual performed use case presented.

In total, 28 studies provide a use case, proving that usability of the graphs is of interest. However, like previously mentioned, the focus does not seem to include full utilisation of the knowledge graph capabilities.

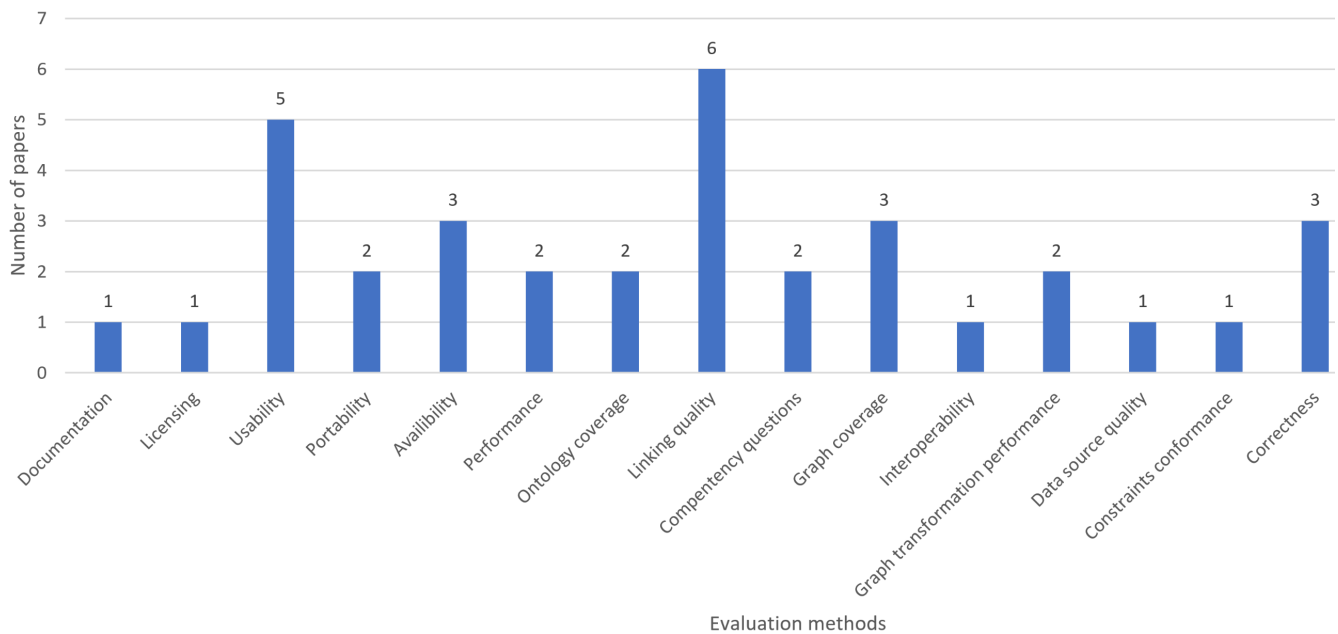


Figure 6. Evaluation methods.

5.2.2. Limitations and Lessons Learned

There is a wide variety of limitations reported. The limitations presented are related to the process used to generate the knowledge graph and the associated tools/systems. This then is often linked to future work, which also gives an insight into the primary focus of the research.

Reported limitations are listed in Table 6. We see that more data are needed to be integrated in several cases, and there is a lack of tools for exploration, showing that there is still room for improvement in this area. The limitations related to linking to the other knowledge graphs and vocabulary coverage show that there is still work to do to support these core aspects.

Table 6. Reported limitations.

Limitation	Count
Graph content	10
Tools for exploration	7
Links to other graphs	5
Coverage by vocabularies	4
Cross-language support	3
Semantic Web skills and knowledge	2
Automation	2
Evaluation	2
Use cases	2

We also extracted lessons learned from the reviewed works, and the frequent ones include:

- Publishing the knowledge graph in an alternative data format, e.g., CSV, helped those not familiar with graph technologies consume the data more easily.
- Metadata helped understandability.
- Publishing the vocabulary helped linking.
- Using a transformation pipeline made the process more maintainable.

- Users were more familiar with JSON, which helped bridge the gap between JSON-LD and RDF.
- Standard vocabulary helped understandability.
- Interlinking unveiled unknown analytics.
- Rest API made the knowledge graph more understandable.
- GitHub helped tracking data when mapping to RDF.
- Data cleaning was important.
- Validation was important to make vocabulary use understandable.
- Use of experts helped data quality.
- Linking to other graphs helped derive structural concepts.
- Keeping track of sub-graphs can be difficult.
- Having many sub-graphs resulted in several vocabularies.
- Errors related to the automation of RDF mapping were out-weighted by the potential for improvement.
- Vocabulary lookup services helped matching vocabularies.
- Vocabularies may be prone to subjectivity.

Several lessons learned related to making the knowledge graphs easier to use for non-experts were reported, which seems an important focus in many of the studies. Other than this, many of the mentions relate to specifics, such as handling of sub-graphs, or that GitHub helped tracking data in the mapping process. Even though these mentions can only be related to the respective primary study, they are valuable in pointing out issues that may be occurring in the future for other knowledge graph producers and motivate the developments of new solutions and tools.

Even though issues are expected when transforming data, few issues were reported, including:

- Five studies mention issues with poor quality of the data source, in terms of, e.g., missing data, duplicate data, and erroneous data.
- Buyle et al. [75] (#6) report an unstable SPARQL endpoint.
- Carriero et al. [82] (#13) report that the URI strategy made some of the data ambiguous.
- Daga et al. [96] (#27) report erroneous language classification in the graph based on user-provided country of origin.
- Khrouf et al. [97] (#28) report that blank nodes create issues when updating the knowledge graph.
- Rietveld et al. [99] (#30) report that the automated mapping pipeline gave erroneous results.
- Dojchinovski et al. [104] (#35) report that linking via string matching gave errors.

The reported issues are quite different, and the low count indicates that they are exclusive to their respective reports. However, some of the issues, such as trouble with URI strategy, updates, and automation are expected depending on the domain. The higher mention of data source quality issues also may be expected, but it does highlight the importance of preprocessing tools and the whole RDF transformation pipeline in general.

6. Discussions

The number of studies published on knowledge graphs creation (as a contribution) in the selected venues constitutes a minor fragment of papers published in those venues. This indicates that there is a limited interest for resource and application studies resulting in knowledge graphs, and this limits their adoption and exploitation. This lack of attention becomes even more evident between different venues. Resource and application studies documenting the knowledge graph creation process along with the resulting knowledge graphs, software and data resources used as well as lessons learned and emerging issues are essential for guiding the research in this area. A similar discrepancy also occurs at a larger scale between studies focusing on public data sources and commercial/industrial data

sources. The fact that the review largely included public knowledge graphs makes it hard for the community to assess the quality of knowledge graphs developed in the industry.

An interesting observation was the reuse of ontologies and infrequent mention of the development of new ontologies. As many as 35 of the 36 chosen studies report on reuse and the selection process of other existing vocabularies. Another interesting observation was transforming data into KG by directly transforming the original schema into an ontology. Based on the previously described ontology development methods, this may be seen as using a bottom-up approach, since the decision on ontology reuse was guided by the data and not the domain. This may be due to fact that ontology development is a demanding task; however, the ontologies used closely follow the underlying data, and hence, its model may indicate problems regarding the quality of knowledge graphs developed, since for example, a relational schema is developed with different considerations in mind compared to an ontology.

Out of the 36 studies, 16 mentioned some way of data preprocessing of the data source before mapping to RDF. The most mentioned was data cleaning, with 14 mentions. This shows the importance, also for structured sources, of data preprocessing. An interesting observation however is that this process is not always mentioned in the current published literature (e.g., [2,3,33,34]). The data sources mentioned throughout the reviewed literature are diverse, including XML, JSON, CSV, and relational databases, demanding data preprocessing and hence tools for supporting common preprocessing tasks.

Given the publication methods reported, the focus of the studies appears to be on the direct publishing of knowledge graphs and not on providing more complex views. Most of the studies provide a SPARQL endpoint for the published knowledge graph, which enables the possibility to extract deductive knowledge. However, more complex views, providing inductive knowledge, were low in numbers. Utilising the full features of knowledge graphs could motivate the adoption of the technology. There is considerable research available on this area (e.g., [2,11,13–15]). Similarly, quality of the knowledge graphs is also a sparsely developed area; some dimensions such as linking quality and usability are often used, while others such as performance and ontology coverage are less used. It is often not clear if all the necessary evaluations are conducted, raising questions on the practical use of the resulting knowledge graphs.

The limitations, issues, and lessons learned give an insight into what the authors of the studies saw as areas for potential improvement. There is a great variety in all aspects, underpinning the complexity of knowledge graph creation. One of the key elements concerns the lack of knowledge and skills on Semantic Web technologies, which is a concern not only for consumers of the knowledge graphs but also for developers, since it could be a challenge for public and private organisations to find skilled staff in this area. For example, Soylyu et al. [77] reported that the use of a RESTful approach helped non-Semantic Web developers to use the data more easily. Therefore, end user tools supporting knowledge graph consumption and various tasks related to knowledge graph construction are essential. Finally, low data source quality is also worth mentioning, since this makes the construction and use of the knowledge graphs a challenge. Particularly, public entities are required to publish more and more data openly; however, quality of the data published is a challenge that has to be addressed.

7. Conclusions

In this paper, we reported a systematic review, focusing on four selected venues, namely *Extended Semantic Web Conference*, *International Semantic Web Conference*, *Journal of Web Semantics*, and *Semantic Web Journal*. The goal was to obtain an overview over publication statistics, technical details, and issues and limitations within knowledge graph creation for the Semantic Web. Regarding the validity of results, the arguably smaller scope of venues can be considered a limitation. The venues were chosen based on general citation count and the relevancy of publications within the scope. We see that the count of studies, within the given time frame, was as much as 36 studies. Given the fact that the chosen

venues are of high relevance within the Semantic Web domain and have strict review processes, the sample of studies should give a relevant and valid view for the addressed research questions.

From the research statistics, we see that the interest in knowledge graph publication in the Semantic Web domain, throughout the four years reported, had a decrease from the first two years to the last two years, and had its peak in 2017. From the technical details, the main observations were that:

- The main phases in the construction process reported were ontology development, the RDF mapping process, and publication.
- There is a big diversity in sources, tools, and phases reported, with 11 data source types, more than 20 sub-phases, and more than 40 tools and languages used.
- There was a big focus on publication; 24 out of 36 studies reported a front end for the exploration of the knowledge graph.
- Few studies reported the use of advanced knowledge processing, i.e., inductive knowledge such as embedding, analytics, etc.
- There was little focus on the evaluation phase; most of the evaluation tasks described were qualitative and not within a dedicated phase.

Regarding the reported issues, limitations, and lessons learned, a great variety was reported, but many aspects were within the same category. The main findings were:

- There were frequent mentions of limitations related to tools.
- Many of the lessons learned were related to making knowledge graph technologies more approachable to non-experts.
- Poor data quality is often reported to be an issue.

Regarding the future work, literature reviews focusing explicitly on the tools and methodologies for creating knowledge graphs with the goal of consolidating the landscape would be one direction. This is because the findings from the review showed that knowledge graph construction in the Semantic Web domain is of high complexity with a very fragmented tool portfolio. The scope of this review was on the creation of knowledge graphs from (semi-)structured data sources; there is still room for reviews with a larger scope. Apart from extending the selected venues, another interesting direction would be to do a similar review for knowledge graph construction from unstructured data sources.

Author Contributions: Conceptualization, V.R., A.S. and D.R.; methodology, V.R., A.S. and D.R.; formal analysis, V.R.; investigation, V.R.; data curation, V.R.; writing—original draft preparation, V.R.; writing—review and editing, V.R., A.S. and D.R.; supervision, A.S. and D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the European Commission Horizon 2020 DataCloud project (grant number 101016835), the NFR BigDataMine project (grant number 309691), and the SINTEF internally funded SEP DataPipes project.

Data Availability Statement: Not applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
OWL	Web Ontology Language
CSV	Comma-Separated Values
XML	Extensible Markup Language
JSON	JavaScript Object Notation

SPARQL	SPARQL Protocol and RDF Query Language
W3C	The World Wide Web Consortium
HTML	HyperText Markup Language
URI	Uniform Resource Identifier
IRI	Internationalised Resource Identifier
TXT	Text File
API	Application Programming Interface
XSLT	Extensible Stylesheet Language Transformations
SHACL	Shapes Constraint Language
CLI	Command Line Interface
ETL	Extract–Transform–Load
ORDL	Open Digital Rights Language
ESWC	Extended Semantic Web Conference
ISWC	International Semantic Web Conference
JWS	Journal of Web Semantics
SWJ	Semantic Web Journal

References

- Gutierrez, C.; Sequeda, J.F. Knowledge Graphs. *Commun. ACM* **2021**, *64*, 96–104. [\[CrossRef\]](#)
- Hogan, A.; Blomqvist, E.; Cochez, M.; D’amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge graphs. *ACM Comput. Surv.* **2022**, *1*, 54. [\[CrossRef\]](#)
- Fensel, D.; Simsek, U.; Angele, K.; Huaman, E.; Karle, E.; Panasiuk, O.; Toma, I.; Umbrich, J.; Wahler, A. *Knowledge Graphs: Methodology, Tools and Selected Use Cases*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2020. [\[CrossRef\]](#)
- Noy, N.; Gao, Y.; Jain, A.; Narayanan, A.; Patterson, A.; Taylor, J. Industry-scale knowledge graphs: Lessons and challenges. *Commun. ACM* **2019**, *62*, 36–43. [\[CrossRef\]](#)
- Kharlamov, E.; Hovland, D.; Skjæveland, M.G.; Bilidas, D.; Jiménez-Ruiz, E.; Xiao, G.; Soylu, A.; Lanti, D.; Rezk, M.; Zheleznyakov, D.; et al. Ontology based data access in Statoil. *J. Web Semant.* **2017**, *44*, 3–36. [\[CrossRef\]](#)
- Kharlamov, E.; Mailis, T.; Mehdi, G.; Neuenstadt, C.; Özgür, Ö.; Roshchin, M.; Solomakhina, N.; Soylu, A.; Svingos, C.; Brandt, S.; et al. Semantic access to streaming and static data at Siemens. *J. Web Semant.* **2017**, *44*, 54–74. [\[CrossRef\]](#)
- Soylu, A.; Corcho, O.; Elvesæter, B.; Badenes-Olmedo, C.; Blount, T.; Yedro Martínez, F.; Kovacic, M.; Posinkovic, M.; Makgill, I.; Taggart, C.; et al. TheyBuyForYou platform and knowledge graph: Expanding horizons in public procurement with open linked data. *Semant. Web* **2021**, *13*, 265–291. [\[CrossRef\]](#)
- Roman, D.; Alexiev, V.; Paniagua, J.; Elvesæter, B.; von Zernichow, B.M.; Soylu, A.; Simeonov, B.; Taggart, C. The euBusinessGraph ontology: A lightweight ontology for harmonizing basic company information. *Semant. Web* **2022**, *13*, 41–68. [\[CrossRef\]](#)
- Hitzler, P. A review of the Semantic Web field. *Commun. ACM* **2021**, *64*, 76–83. [\[CrossRef\]](#)
- Polleres, A.; Hogan, A.; Delbru, R.; Umbrich, J. RDFS and OWL reasoning for linked data. In Proceedings of the 9th International Summer School on Reasoning Web, Mannheim, Germany, 30 July–2 August 2013; Volume 8067, pp. 91–149. [\[CrossRef\]](#)
- Iosup, A.; Hegeman, T.; Ngai, W.L.; Heldens, S.; Prat-Pérez, A.; Manhardt, T.; Chafio, H.; Capotă, M.; Sundaram, N.; Anderson, M.; et al. LDBC Graphalytics: A benchmark for large-scale graph analysis on parallel and distributed platforms. *Proc. VLDB Endow.* **2016**, *9*, 1317–1328. [\[CrossRef\]](#)
- Gulnes, M.P.; Soylu, A.; Roman, D. A graph-based approach for representing, integrating and analysing neuroscience data: The case of the murine basal ganglia. *Data Technol. Appl.* **2021**, *in press*. [\[CrossRef\]](#)
- Angles, R.; Arenas, M.; Barceló, P.; Hogan, A.; Reutter, J.; Vrgoč, D. Foundations of modern query languages for graph databases. *ACM Comput. Surv.* **2017**, *50*. [\[CrossRef\]](#)
- Wang, Q.; Mao, Z.; Wang, B.; Guo, L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2724–2743. [\[CrossRef\]](#)
- Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Networks* **2009**, *20*, 61–80. [\[CrossRef\]](#) [\[PubMed\]](#)
- Corcho, O.; Priyatna, F.; Chaves-Fraga, D. Towards a new generation of ontology based data access. *Semant. Web* **2020**, *11*, 153–160. [\[CrossRef\]](#)
- Yan, J.; Wang, C.; Cheng, W.; Gao, M.; Zhou, A. A retrospective of knowledge graphs. *Front. Comput. Sci.* **2018**, *12*, 55–74. [\[CrossRef\]](#)
- Ryen, V. Semantic Knowledge Graph Creation From Structured Data: A Systematic Literature Review. Master’s Thesis, University of Oslo, Oslo, Norway, 2021.
- Paulheim, H. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semant. Web* **2017**, *8*, 489–508. [\[CrossRef\]](#)
- Zaveri, A.; Rula, A.; Maurino, A.; Pietrobon, R.; Lehmann, J.; Auer, S. Quality assessment for linked data: A survey. *Semant. Web* **2016**, *7*, 63–93. [\[CrossRef\]](#)

21. Pereira, C.K.; Siqueira, S.W.M.; Nunes, B.P.; Dietze, S. Linked data in education: A survey and a synthesis of actual research and future challenges. *IEEE Trans. Learn. Technol.* **2018**, *11*, 400–412. [CrossRef]
22. Bizer, C.; Seaborne, A. D2RQ—Treating non-RDF databases as virtual RDF graphs. In Proceedings of the Poster presented at the 3rd International Semantic Web Conference (ISWC 2004, 2004), Hiroshima, Japan, 7–11 November 2004.
23. Available online: <https://virtuoso.openlinksw.com> (accessed on 30 March 2022).
24. Broekstra, J.; Kampman, A.; Harmelen, F.V. Sesame: A generic architecture for storing and querying RDF and RDF schema. In Proceedings of the 1st International Semantic Web Conference (ISWC 2002), Sardinia, Italia, 9–12 June 2002; Volume 2342, pp. 54–68. [CrossRef]
25. Available online: <https://rdf4j.org> (accessed on 30 March 2022).
26. Available online: <https://www.dublincore.org> (accessed on 30 March 2022).
27. Barbosa, A.; Bittencourt, I.I.; Siqueira, S.W.M.; Silva, R.d.A.; Calado, I. The use of software tools in linked data publication and consumption: A systematic literature review. *Int. J. Semant. Web Inf. Syst.* **2017**, *13*, 68–88. [CrossRef]
28. Avila-Garzon, C. Applications, methodologies, and technologies for linked open data: A systematic literature review. *Int. J. Semant. Web Inf. Syst.* **2020**, *16*, 53–69. [CrossRef]
29. Penteado, B.E.; Maldonado, J.C.; Isotani, S. Methodologies for publishing linked open government data on the Web: A systematic mapping and a unified process model. *Semant. Web* **2022**, in press. [CrossRef]
30. Feitosa, D.; Dermeval, D.; Ávila, T.; Bittencourt, I.I.; Lóscio, B.F.; Isotani, S. A systematic review on the use of best practices for publishing linked data. *Online Inf. Rev.* **2018**, *42*, 107–123. [CrossRef]
31. Pinto, V.A.; Parreiras, F.S. Enterprise linked data: A systematic mapping study. In Proceedings of the 33rd International Conference on Conceptual Modeling Workshops (ER 2014), Atlanta, GA, USA, 27–30 October 2014; Volume 8823, pp. 253–262. [CrossRef]
32. Ali, I.; Warraich, N.F. Linked data initiatives in libraries and information centres: A systematic review. *Electron. Libr.* **2018**, *36*, 925–937. [CrossRef]
33. Pan, J.Z.; Vetere, G.; Gomez-Perez, J.M.; Wu, H. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2017.
34. Kejriwal, M. *Domain-Specific Knowledge Graph Construction*; Springer: Berlin/Heidelberg, Germany, 2019.
35. Panasiuk, O.; Karle, E.; Simsek, U.; Fensel, D. Defining tourism domains for semantic annotation of Web content. In Proceedings of the ENTER2018 eTourism Conference, Jönköping, Sweden, 24–26 January 2018.
36. Sequeda, J.F.; Briggs, W.J.; Miranker, D.P.; Heideman, W.P. A pay-as-you-go methodology to design and build enterprise knowledge graphs from relational databases. In Proceedings of the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, 26–30 October 2019; Volume 11779, pp. 526–545. [CrossRef]
37. Cristani, M.; Cuel, R. A survey on ontology creation methodologies. *Int. J. Semant. Web Inf. Syst.* **2005**, *1*, 49–69. [CrossRef]
38. Pinto, H.S.; Staab, S.; Tempich, C. DILIGENT: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies. In Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004), Valencia, Spain, 22–27 August 2004; IOS Press: Amsterdam, The Netherlands, 2004; pp. 393–397.
39. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From ontological art towards ontological engineering. In Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97), Providence, RI, USA, 27–31 July 1997; The MIT Press: Cambridge, MA, USA, 1997.
40. Available online: http://ontologydesignpatterns.org/wiki/Main_Page (accessed on 19 April 2022).
41. Martínez-Rodríguez, J.; Hogan, A.; López-Arévalo, I. Information extraction meets the Semantic Web: A survey. *Semant. Web* **2020**, *11*, 255–335. [CrossRef]
42. Rahm, E.; Do, H.H. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.* **2000**, *23*, 3–13.
43. Sukhobok, D.; Nikolov, N.; Pultier, A.; Ye, X.; Berre, A.; Moynihan, R.; Roberts, B.; Elvesæter, B.; Mahasivam, N.; Roman, D. Tabular data cleaning and linked data generation with Grafterizer. In Proceedings of the European Semantic Web Conference (ESWC 2016), Crete, Greece, 29 May–2 June 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9989, pp. 134–139. [CrossRef]
44. Sukhobok, D. Tabular Data Cleaning and Linked Data Generation with Grafterizer. Master’s Thesis, University of Oslo, Oslo, Norway, 2016.
45. Kärle, E.; Simsek, U.; Fensel, D. semantify.it, a Platform for Creation, Publication and Distribution of Semantic Annotations. CoRR. 2017. Available online: <https://arxiv.org/abs/1706.10067> (accessed on 30 March 2022).
46. Vrandečić, D.; Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* **2014**, *57*, 78–85. 45/2629489. [CrossRef]
47. Lenzerini, M. Data integration: A theoretical perspective. In Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2002), Madison, WI, USA, 3–5 June 2002; ACM: New York, NY, USA, 2002; pp. 233–246. [CrossRef]
48. Available online: https://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927/#direct_mapping_rules (accessed on 19 April 2022).
49. Dimou, A. Chapter 4 creation of knowledge graphs. In *Knowledge Graphs and Big Data Processing*; Janev, V., Graux, D., Jabeen, H., Sallinger, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 59–72. [CrossRef]

50. Available online: <https://www.w3.org/TR/r2rml/> (accessed on 19 April 2022).
51. Dimou, A.; Sande, M.V.; Colpaert, P.; Verborgh, R.; Mannens, E.; de Walle, R.V. RML: A generic language for integrated RDF mappings of heterogeneous data. In Proceedings of the Workshop on Linked Data on the Web (LDOW 2014), Seoul, Korea, 8 April 2014; CEUR.org: Aachen, Germany, 2014; Volume 1184.
52. Meester, B.D.; Dimou, A.; Verborgh, R.; Mannens, E. An ontology to semantically declare and describe functions. In Proceedings of the ESWC 2016 Satellite Events, Crete, Greece, 29 May–2 June 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9989, pp. 46–49. [[CrossRef](#)]
53. Available online: <https://www.w3.org/TR/rdf-sparql-query/> (accessed on 19 April 2022).
54. Guéret, C.; Groth, P.; Stadler, C.; Lehmann, J. Assessing linked data mappings using network measures. In Proceedings of the 9th Extended Semantic Web Conference (ESWC 2019), Portorož, Slovenia, 2–6 June 2019; Springer: Berlin/Heidelberg, Germany; Volume 7295, pp. 87–102. [[CrossRef](#)]
55. Fürber, C.; Hepp, M. Swiqa—A Semantic Web information quality assessment framework. In Proceedings of the 19th European Conference on Information Systems (ECIS 2011), Helsinki, Finland, 9–11 June 2011.
56. Mendes, P.N.; Mühleisen, H.; Bizer, C. Sieve: Linked data quality assessment and fusion. In Proceedings of the 2012 Joint EDBT/ICDT Workshops (EDBT-ICDT 2012), Berlin, Germany, 30 March 2012; ACM: New York, NY, USA, 2012; pp. 116–123. [[CrossRef](#)]
57. Hogan, A.; Umbrich, J.; Harth, A.; Cyganiak, R.; Polleres, A.; Decker, S. An empirical survey of linked data conformance. *J. Web Semant.* **2012**, *14*, 14–44. [[CrossRef](#)]
58. Rula, A.; Palmonari, M.; Rubinacci, S.; Ngomo, A.N.; Lehmann, J.; Maurino, A.; Esteves, D. TISCO: Temporal scoping of facts. *J. Web Semant.* **2019**, *54*, 72–86. [[CrossRef](#)]
59. Available online: <https://www.w3.org/DesignIssues/LinkedData.html> (accessed on 19 April 2022).
60. Hyvönen, E.; Tuominen, J.; Alonen, M.; Mäkelä, E. Linked Data Finland: A 7-Star Model and Platform for Publishing and Re-Using Linked Datasets. In Proceedings of the ESWC 2014 Satellite Events, Crete, Greece, 25–29 May 2014; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8798, pp. 226–230. [[CrossRef](#)]
61. Janowicz, K.; Hitzler, P.; Adams, B.; Kolas, D.; Vardeman, C. Five stars of linked data vocabulary use. *Semant. Web* **2014**, *5*, 173–176. [[CrossRef](#)]
62. Heath, T.; Bizer, C. *Linked Data: Evolving the Web into a Global Data Space*, 1st ed.; Synthesis Lectures on the Semantic Web; Morgan & Claypool Publishers: Santa Clara, CA, USA, 2011. [[CrossRef](#)]
63. Available online: <https://www.schema.org> (accessed on 19 April 2022).
64. Available online: <https://dbpedia.org/> (accessed on 19 April 2022).
65. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
66. Available online: <https://www.w3.org/TR/odrl-model/> (accessed on 19 April 2022).
67. Available online: <https://www.apache.org/licenses/LICENSE-2.0.html> (accessed on 19 April 2022).
68. Available online: <https://creativecommons.org/licenses/> (accessed on 19 April 2022).
69. Available online: <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-star/> (accessed on 19 April 2022).
70. Regalia, B.; Janowicz, K.; Mai, G.; Varanka, D.; Usery, E.L. GNIS-LD: Serving and visualizing the geographic names information system gazetteer as linked Data. In Proceedings of the 15th International Conference on the Semantic Web (ESWC 2018), Heraklion, Crete, Greece, 3–7 June 2018; Springer: Berlin/Heidelberg, Germany, 2018; Volume 10843, pp. 528–540. [[CrossRef](#)]
71. Hyvönen, E.; Heino, E.; Leskinen, P.; Ikkala, E.; Koho, M.; Tamper, M.; Tuominen, J.; Mäkelä, E. WarSampo data service and semantic portal for publishing linked open data about the second world war history. In Proceedings of the 13th International Conference on the Semantic Web (ESWC 2016), Heraklion, Crete, Greece, 29 May–2 June 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9678, pp. 758–773. [[CrossRef](#)]
72. Klímek, J. DCAT-AP representation of Czech national open data catalog and its impact. *J. Web Semant.* **2019**, *55*, 69–85. [[CrossRef](#)]
73. Klímek, J.; Kucera, J.; Necaský, M.; Chlapek, D. Publication and usage of official Czech pension statistics Linked Open Data. *J. Web Semant.* **2018**, *48*, 1–21. [[CrossRef](#)]
74. Troncy, R.; Rizzo, G.; Jameson, A.; Corcho, Ó.; Plu, J.; Palumbo, E.; Hermida, J.C.B.; Spirescu, A.; Kuhn, K.; Barbu, C.; et al. 3city: Building comprehensive knowledge bases for city exploration. *J. Web Semant.* **2017**, *46–47*, 2–13. m.2017.07.002. [[CrossRef](#)]
75. Buyle, R.; Vanlshout, Z.; Coetzee, S.; Paepe, D.D.; Compennolle, M.V.; Thijs, G.; Nuffelen, B.V.; Vocht, L.D.; Mechant, P.; Vidts, B.D.; et al. Raising interoperability among base registries: The evolution of the linked base registry for addresses in Flanders. *J. Web Semant.* **2019**, *55*, 86–101. [[CrossRef](#)]
76. McCusker, J.P.; Keshan, N.; Rashid, S.M.; Deagen, M.; Brinson, L.C.; McGuinness, D.L. NanoMine: A knowledge graph for nanocomposite materials science. In Proceedings of the 19th International Semantic Web Conference (ISWC 2020), Athens, Greece, 2–6 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12507, pp. 144–159. [[CrossRef](#)]

77. Soyulu, A.; Corcho, Ó.; Elvesæter, B.; Badenes-Olmedo, C.; Martínez, F.Y.; Kovacic, M.; Posinkovic, M.; Makgill, I.; Taggart, C.; Simperl, E.; et al. Enhancing public procurement in the European Union through constructing and exploiting an integrated knowledge graph. In Proceedings of the 19th International Semantic Web Conference (ISWC 2020), Athens, Greece, 2–6 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12507, pp. 430–446. [\[CrossRef\]](#)
78. Achichi, M.; Lisena, P.; Todorov, K.; Troncy, R.; Delahousse, J. DOREMUS: A graph of linked musical works. In Proceedings of the 17th International Semantic Web Conference (ISWC 2018), Monterey, CA, USA, 8–12 October 2018; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11137, pp. 3–19. [\[CrossRef\]](#)
79. Kiesling, E.; Ekelhart, A.; Kurniawan, K.; Ekaputra, F.J. The SEPSES knowledge graph: An integrated resource for cybersecurity. In Proceedings of the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, 26–30 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11779, pp. 198–214. [\[CrossRef\]](#)
80. Knoblock, C.A.; Szekely, P.A.; Fink, E.E.; Degler, D.; Newbury, D.; Sanderson, R.; Blanch, K.; Snyder, S.; Chheda, N.; Jain, N.; et al. Lessons learned in building linked data for the American art collaborative. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, 21–25 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10588, pp. 263–279. [\[CrossRef\]](#)
81. Steenwinckel, B.; Vandewiele, G.; Rausch, I.; Heyvaert, P.; Taelman, R.; Colpaert, P.; Simoons, P.; Dimou, A.; Turck, F.D.; Ongenaes, F. Facilitating the analysis of COVID-19 literature through a knowledge graph. In Proceedings of the 19th International Semantic Web Conference (ISWC 2020), Athens, Greece, 2–6 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12507, pp. 344–357. [\[CrossRef\]](#)
82. Carriero, V.A.; Gangemi, A.; Mancinelli, M.L.; Marinucci, L.; Nuzzolese, A.G.; Presutti, V.; Veninata, C. ArCo: The Italian cultural heritage knowledge graph. In Proceedings of the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, 26–30 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11779, pp. 36–52. [\[CrossRef\]](#)
83. Kubitz, D.O.; Böckmann, M.; Graux, D. SemanGit: A linked dataset from git. In Proceedings of the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, 26–30 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11779, pp. 215–228. [\[CrossRef\]](#)
84. Debruyne, C.; Meehan, A.; Clinton, E.; McNerney, L.; Nautiyal, A.; Lavin, P.; O’Sullivan, D. Ireland’s authoritative geospatial linked data. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, 21–25 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10588, pp. 66–74. [\[CrossRef\]](#)
85. Peroni, S.; Shotton, D.M.; Vitali, F. One year of the OpenCitations corpus. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, 21–25 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10588, pp. 184–192. [\[CrossRef\]](#)
86. Fang, Z.; Wang, H.; Gracia, J.; Bosque-Gil, J.; Ruan, T. Zhishi.lemon: On publishing Zhishi.me as linguistic linked open data. In Proceedings of the 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, 17–21 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; Volume 9982, pp. 47–55. [\[CrossRef\]](#)
87. Gracia, J.; Fäth, C.; Hartung, M.; Ionov, M.; Bosque-Gil, J.; Veríssimo, S.; Chiarcos, C.; Orlikowski, M. Leveraging linguistic linked data for cross-lingual model transfer in the pharmaceutical domain. In Proceedings of the 19th International Semantic Web Conference (ISWC 2020), Athens, Greece, 2–6 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 12507, pp. 499–514. [\[CrossRef\]](#)
88. Bechhofer, S.; Page, K.R.; Weigl, D.M.; Fazekas, G.; Wilmering, T. Linked data publication of live music archives and analyses. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, 21–25 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10588, pp. 29–37. [\[CrossRef\]](#)
89. Färber, M. The Microsoft academic knowledge graph: A linked data source with 8 billion triples of scholarly data. In Proceedings of the 18th International Semantic Web Conference (ISWC 2019), Auckland, New Zealand, 26–30 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11779, pp. 113–129. [\[CrossRef\]](#)
90. Wang, M.; Zhang, J.; Liu, J.; Hu, W.; Wang, S.; Li, X.; Liu, W. PDD graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, 21–25 October 2017; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10588, pp. 219–227. [\[CrossRef\]](#)
91. Roussey, C.; Bernard, S.; André, G.; Boffety, D. Weather data publication on the LOD using SOSA/SSN ontology. *Semant. Web* **2020**, *11*, 581–591. [\[CrossRef\]](#)
92. Dijkshoorn, C.; Jongma, L.; Aroyo, L.; van Ossenbruggen, J.; Schreiber, G.; ter Weele, W.; Wielemaker, J. The Rijksmuseum collection as linked data. *Semant. Web* **2018**, *9*, 221–230. [\[CrossRef\]](#)
93. Gracia, J.; Villegas, M.; Gómez-Pérez, A.; Bel, N. The apertium bilingual dictionaries on the Web of data. *Semant. Web* **2018**, *9*, 231–240. [\[CrossRef\]](#)
94. Lefort, L.; Haller, A.; Taylor, K.; Squire, G.; Taylor, P.; Percival, D.; Woolf, A. The ACORN-SAT linked climate dataset. *Semant. Web* **2017**, *8*, 959–967. [\[CrossRef\]](#)
95. Höffner, K.; Martin, M.; Lehmann, J. LinkedSpending: OpenSpending becomes Linked Open Data. *Semant. Web* **2016**, *7*, 95–104. [\[CrossRef\]](#)
96. Daga, E.; d’Aquin, M.; Adamou, A.; Brown, S. The Open University linked data—data.open.ac.uk. *Semant. Web* **2016**, *7*, 183–191. [\[CrossRef\]](#)
97. Khrouf, H.; Troncy, R. EventMedia: A LOD dataset of events illustrated with media. *Semant. Web* **2016**, *7*, 193–199. [\[CrossRef\]](#)

98. Queralt-Rosinach, N.; Kuhn, T.; Chichester, C.; Dumontier, M.; Sanz, F.; Furlong, L.I. Publishing DisGeNET as nanopublications. *Semant. Web* **2016**, *7*, 519–528. [[CrossRef](#)]
99. Rietveld, L.; Beek, W.; Hoekstra, R.; Schlobach, S. Meta-data for a lot of LOD. *Semant. Web* **2017**, *8*, 1067–1080. [[CrossRef](#)]
100. Meroño-Peñuela, A.; Ashkpour, A.; Guéret, C.; Schlobach, S. CEDAR: The Dutch historical censuses as linked open data. *Semant. Web* **2017**, *8*, 297–310. [[CrossRef](#)]
101. Baierer, K.; Dröge, E.; Eckert, K.; Goldfarb, D.; Iwanowa, J.; Morbidoni, C.; Ritze, D. DM2E: A Linked Data source of Digitised Manuscripts for the Digital Humanities. *Semant. Web* **2017**, *8*, 733–745. [[CrossRef](#)]
102. Romero, G.C.; Esteban, M.P.E.; Carrasco, R.C.; Such, M.M. Migration of a library catalogue into RDA linked open data. *Semant. Web* **2018**, *9*, 481–491. [[CrossRef](#)]
103. Färber, M.; Menne, C.; Harth, A. A linked data wrapper for CrunchBase. *Semant. Web* **2018**, *9*, 505–515. [[CrossRef](#)]
104. Dojchinovski, M.; Vitvar, T. Linked Web APIs dataset. *Semant. Web* **2018**, *9*, 381–391. [[CrossRef](#)]
105. van Aggelen, A.; Hollink, L.; Kemman, M.; Kleppe, M.; Beunders, H. The debates of the European Parliament as linked open data. *Semant. Web* **2017**, *8*, 271–281. [[CrossRef](#)]
106. Available online: <https://eswc-conferences.org> (accessed on 19 April 2022).
107. Available online: <http://swsa.semanticweb.org/content/international-semantic-web-conference-iswc> (accessed on 19 April 2022).
108. Available online: <https://www.journals.elsevier.com/journal-of-web-semantics> (accessed on 19 April 2022).
109. Available online: <http://www.semantic-web-journal.net> (accessed on 19 April 2022).
110. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report; Evidence-Based Software Engineering (EBSE) Project: Staffordshire, UK, 2007.
111. Chen, Z.; Wang, Y.; Zhao, B.; Cheng, J.; Zhao, X.; Duan, Z. Knowledge graph completion: A review. *IEEE Access* **2020**, *8*, 192435–192456. [[CrossRef](#)]
112. *Shapes Constraint Language (SHACL)*; Technical Report; The World Wide Web Consortium (W3C): Cambridge, MA, USA, 2017.
113. Tomaszuk, D.; Hyland-Wood, D. RDF 1.1: Knowledge Representation and Data Integration Language for the Web. *Symmetry* **2020**, *12*, 84. [[CrossRef](#)]
114. Available online: <https://www.swi-prolog.org> (accessed on 19 April 2022).
115. Available online: <https://jena.apache.org/index.html> (accessed on 19 April 2022).
116. Available online: <http://silkframework.org> (accessed on 19 April 2022).
117. Available online: <https://www.w3.org/2001/sw/wiki/Pubby> (accessed on 19 April 2022).
118. *XSL Transformations (XSLT) Version 3.0*; Technical Report; The World Wide Web Consortium (W3C): Cambridge, MA, USA, 2017.