An End-to-End Framework for Integrating and Publishing Linked Open Government Data

Abida, Rabeb; Hachicha Belghith, Emna; Cleve, Anthony

Published in:
Proceedings of the 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2020)

Publication date:
2020

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

An End-to-End Framework for Integrating and Publishing Linked Open Government Data

Abida, Rabeb; Hachicha Belghith, Emna; Cleve, Anthony

Published in:
Proceedings of the 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2020)

Publication date:
2020

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

An End-to-End Framework for Integrating and Publishing Linked Open Government Data

Rabeb Abida
Namur Digital Institute
PReCISE research center
University of Namur, Belgium
rabeb.abida@unamur.be

Emna Hachicha Belghith
Namur Digital Institute
PReCISE research center
University of Namur, Belgium
emna.hachicha@unamur.be

Anthony Cleve
Namur Digital Institute
PReCISE research center
University of Namur, Belgium
anthony.cleve@unamur.be

***Abstract*—Linked Government Statistical Data on the Web are significantly increasing in terms of variety, which makes challenging to determine the quality of this data that explicitly become accessible to data consumers. Converting and publishing such data involves several challenges, e.g., data unifying, design decisions, knowledge extraction. In this paper, we aim to address this issue by proposing an end-to-end framework, based on *linked open data technologies*. This framework assists the user to produce and publish structured Linked Data in an e-Government context. Our aim is to enable data consumers to access to an end-to-end solution, allowing them to produce and publish high-quality data on the web, that meet their needs. To assess our framework, we use a real case study, relying on a local government catalogue.**

***Index Terms*—Linked Open Data, Open Government data, Semantic Web, Data publishing**

I. Introduction

In the last years, open data has become an important trend, especially in the context of government data (i.e., Open Government Data), since it offers citizens and public institutions a transparent, free and customized access to such information. It also enable the development of innovative services, both in the public and private sectors [1], [2]. In such a context, several government projects (e.g., in France, United States, Belgium) have been launched in order to provide free access to different data catalogues such as *data.gov*[1], *data.gouv.fr*[2], and *data.gov.be*[3] in diverse domains (i.e., geospatial, statistics, transport, education, population, etc).

Despite this increasing attention, the open data available on the web are still of a high level of heterogeneity: They are available in various formats (i.e., spreadsheets, CSV, relational database dumps, RDF), which make challenging to determine the quality of data to be published. The advent of Linked Open Data (LOD)[4] technologies allows for more efficient access and facilitates knowledge extraction and enrichment from heterogeneous data sources. Yet, limited work exists, in the R&D community, for supporting the end-to-end process of LOD production and publishing in a e-Government context. There is a clear lack of automatically-supported, integrated solutions that enable (i) combining appropriate tools to assist users, especially non-expert users, to efficiently manipulate data sets from extraction to publication, and (ii) having a follow-up on the progress during this process as well as a semantic visualization of its data.

Some approaches have been proposed, mainly with a particular, non-integrated focus on data conversion [3], on data formalization [4], or on knowledge extraction [5], [6]. Other proposals have been made for publishing linked open government data (LOGD) [3], [7]–[9] but without considering data refinement nor semantic visualizations.

In this paper, we aim to fill the above research gap by proposing an integrated framework for semi-automatically assisting citizens in the **Pub**lishing of **L**inked **O**pen **G**overnment **D**ata (*Pub-LOGD)*. Our framework consists in an end-to-end solution that covers five main data analysis steps (i) *data transformation*, (ii) *data interlinking*, (iii) *data storage*, (iv) *data visualization* and (v) *data publishing*. It relies on a *Docker-based* implementation using a combination of software tools in order to provide users with a facilitated and customized access to the data sets available on the web.

Our purpose is to transform raw data to a high-quality government dataset published on the web, by also providing semantic visualizations which make citizens aware of the extracted knowledge to enhance decision-making. We have made a preliminary evaluation of our proposal through a case study, based on a local population dataset.

The remainder of the paper is organized as follows. Section 2 gives an overview of our framework and details the followed steps it supports. Section 3 describes the application of our framework in a particular Belgium government catalogue in order to evaluate our solution. Section 4 presents the experimental results. Section 5 positions our work with respect to the related literature. Section 6 concludes the paper and anticipates our future work.

II. Approach Overview

In this section, we present a global overview of our framework *Pub-LOGD*, that consists of a pipeline of five main data processing steps for publishing government data as linked data. Our purpose is to allow public institutions as well as citizens to access to an end-to-end framework that guides them with the appropriate tools during the entire publishing process, thereby offering them relevant data visualizations.

[1]https://www.data.gov/open-gov/
[2]https://www.gouvernement.fr/en/openness-of-public-data
[3]https://data.gov.be/fr
[4]https://www.w3.org/DesignIssues/GovData.html

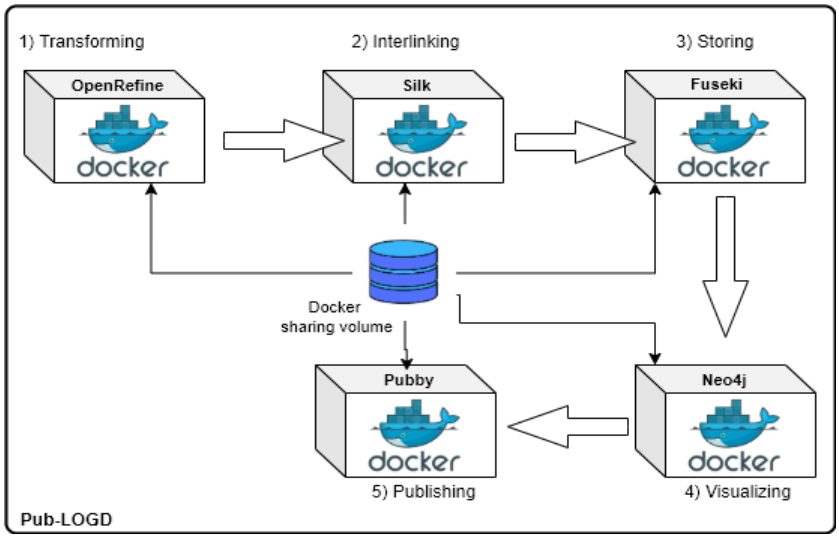

Fig. 1: Pub-LOGD Overview for publishing LOGD as Linked Data.

As shown in Fig. 1, our framework follows an iterative and incremental process which covers five successive activities: (1) Transforming data (i.e., extracting, unifying, cleaning, and converting to RDF), (2) Interlinking data with other data sources, (3) Storage, (4) Visualization, and (5) Publishing on the Web. The integration of the tools supporting each step has been achieved using the Docker container technology. At first, an extracting phase is applied on multiple data sources given as input, which involves a set of discovery actions depending on the user requirements. Two additional phases are then realized that are data unification and data cleaning. All these phases are achieved via *Open Refine*[5] which is an open-source tool that offers powerful capabilities for data formatting and tidying up. Afterwards, an interlinking phase is applied to enrich our LOGD with useful related data extracted from other data sources. The latter step is realized by *SILK*[6] tool. A data storage phase is then performed on the generated LOGD by means of the *Fuseki*[7] tool. A semantic visualization can then be triggered by the user via the use of *neo4j*[8]-based tool which reveals implicit contexts and hidden relationships among the data. Finally, a publishing phase is carried out via the use of the Pubby tool[9] in order to publish the according RDF triplets on the web.

During each step of our process, the user is assisted with guidelines to properly execute the according sub-tasks and he cannot go to the next step without completing the previous one in a correct way. Moreover, the user is provided with a workspace file that consists in a docker sharing volume in order to facilitate the access to his data. More details about each of the steps are presented in the next section.

[5]https://code.google.com/archive/p/google-refine/
[6]http://silkframework.org/
[7]https://jena.apache.org/documentation/fuseki2/
[8]https://neo4j.com/
[9]http://wifo5-03.informatik.uni-mannheim.de/pubby/

We have established a docker-based framework which formally specifies unstructured LOGD, order and publish them in an easy and customized way for the user. We believe that our approach contributes to making LOGD available to citizens in an easier and higher quality manner, by taking advantage of the cumulative capabilities of the appropriate tools in one single framework, and by providing informative data visualization.

III. Assistance for Publishing Linked Government Data

In this section, we briefly detail the different steps supported by our *Pub-LOGD* Framework. More details on the implementation of our framework are available online[10].

A. Data Transformation

The data transformation step consists in extracting statistical data from heterogeneous data sources. To this end, we use *Dcat Browser* as an extension to *Open Refine*. This tool uses the DCAT standard[11] to describe datasets of the public sector in Europe (e.g., http://data.gov.uk, http://open-data.europa.eu, etc).

Then a data unification and cleaning step is applied. The goal of this step is to unify and clean up the data by fixing errors, removing duplicates and preparing them for transformation. *Open Refine* is the main tool that is used in this step. We found it suitable in our context since it provides a user-friendly user interface, especially for non-technical, non-expert users, which is the case for most citizens [10]. Additionally, it enables data clustering which further helps understanding the data.

Once we have cleaned and unified data, it is necessary to define an ontological model to share and annotate our statistical governmental data. To do so, we rely on a domain

[10]https://github.com/123rabida123/WebAppOpenData
[11]http://rd-alliance.github.io/metadata-directory/standards/dcat-data-catalog-vocabulary.html

expert for modeling statistical metadata. Then, we develop the corresponding RDF Extension in *Open Refine* to enable modeling and exporting data in RDF format. This step contains two sub-tasks: data reconciliation and data modeling.

Data reconciliation is a feature that allows to semi-automatically link data to external databases (e.g., Freebase, NCBI taxonomy, DBPedia) in order to align the data with these databases. Its process is semi-automatic because the end-user has the possibility to approve the results interactively and/or to select those to modify from a list of results.

Afterwards, it is necessary to define a data structure to model Statistical data of LOGD in an ontological manner. The data modeling step aims at building a skeleton to define the statistical data structure by reusing available vocabularies/ontologies as much as possible to define namespace prefixes. At this stage, URIs play an important role in the discovery and interoperability of statistical data on the web [11]. RDF extensions allow to import vocabularies available on the web regardless of the format used (e.g. RDFa, RDF/XML, Turtle) as long as their deployment is compatible with the best practices recommended by the W3C [8].

B. Data Interlinking

During the data interlinking step, we generate links between the RDF graphs generated in the previous steps. This steps relies on the *Silk Link Discovery Engine*. It accesses the data sources that should be interlinked via the SPARQL protocol and can thus be used against local as well as remote SPARQL endpoints. This interlinking process is used to express heuristics for deciding whether there are semantic relationships between two related entities. In our approach, we generate a new set of RDF triples from this process and upload them into the *Fuseki* server instance. Then, we use the similarity measure *Levenshtein distance* to compare the labels (*rdfs: label*) of data in our statistical dataset. Therefore, we obtain semantic relationships of type *owl:sameAs* to express that two entities correspond to the same entity among our RDF graphs.

C. Data Storage

The data storage step aims to store the generated RDF triples which are pre-populated with metadata information. The Apache Jena Fuseki[12] tool is used to manage RDF store, which is a SPARQL server offering RDF storage capabilities and a SPARQL endpoint for querying RDF data.

D. Data Visualization

In the current development stage of our framework, we enhanced the data publishing process with visualization functionalities in order to facilitate decision-making and increase transparency. To this end, we use the *neo4j* datastore, which is the most popular and highly scalable graph database, to provide the user with graphical representations of the data and the relationships that exist among them.

[12]https://jena.apache.org/documentation/fuseki2/

E. Data Publishing

The *Pubby* server is used to publish RDF triplets on the web. This step allows to rewrite URI requests into SPARQL to describe queries against the underlying RDF store. It also provides a simple HTML view over the data stores [12]. Pubby provides a Linked Data interface to local or remote SPARQL protocol servers. Thus, a page is created for each resource on the Pubby server. Each page includes links to the corresponding resource and dataset as well as several mappings, which allow users to navigate between dataset resources through the properties used in their mappings.

IV. Illustrative Case Study

In this section, we present the evaluation efforts made to demonstrate the feasibility of our approach. To this end, we present at first a description about the real case study on which we have applied our Pub-LOGD (section IV-A). Then, we discuss the use of *Pub-LOGD* and its benefits (section IV-B).

A. Dataset Description

The dataset consists in a local Belgium catalogue[13] which consists in statistical data related to three data collection: (i) population, (ii) workers, and (iii) housing. These data exactly concern Wallonia region in Belgium where we aim to study the population statistics with taking into account a set of major facets per collection. Descriptions about these facets are presented below.

TABLE I: Dataset Description

Data	Category	Facets	Size
collection1	population	residence	139,000 rows
		sex	
		position household	
		size of the locality	
		labor market situation	
collection2	workers	place of residence	184,000 rows
		sex	
		position household	
		professional status	
		country of citizenship	
collection3	housing	geographical level	136,000 rows
		occupancy regime	
		type of building	

As shown in Table I, the catalogue includes three datasets: (i) collection1: referencing to the population category and including five facets; place of residence, sex, position in the household, size of the locality, and labor market situation (139,000 rows), (ii) collection2: denoting the workers' category involving four facets i.e., place of residence, sex, position in the household, and professional status and country of citizenship (184,000 rows), and (iii) collection3: Typical housing with taking into account three facets i.e., geographical level (municipality), occupancy regime and type of building (136,000 rows).

[13]https://statbel.fgov.be/en

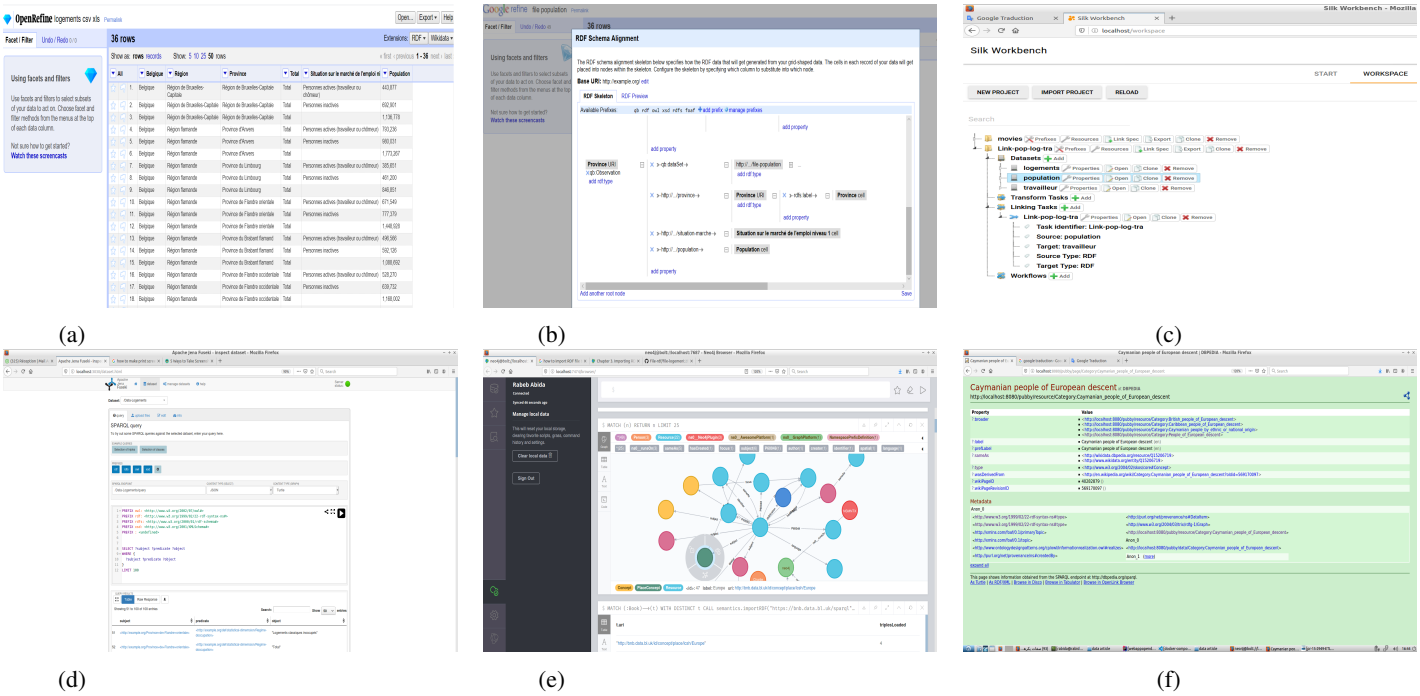

Fig. 2: User Interaction with the publishing Framework.

B. Evaluation Results

To evaluate our proposal, we apply the aforementioned catalogue on the *Pub-LOGD* framework in order to assist end users to achieve published data in the interlinked RDF format, in an easy and transparent manner. At the beginning, the user is provided with guidelines that explain the manner to utilize the framework and what tools are involved to ensure the publishing process. Besides, he is supplied with some reliable data portals, and tutorials to facilitate him tasks. Once the process is clarified, the user can launch the publishing steps.

As shown in Fig. 2, the user uses at first *Open Refine* to preview the data (Fig. 2 (a)), builds the RDF skeleton, clean and cluster statistical data according to its needs (Fig. 2 (b)). Then, an interlinking step, to improve knowledge with related external data sources, is followed through *SILK* (Fig. 2 (c)). Afterwards, the user stores his data via *Apache Jena Fuseki*, which provides him an easy user interface for server monitoring and administration (Fig. 2 (d)). Next, with the integration of *neo4j* (Fig. 2 (e)) the user may have a holistic and complete view about his data as well as the semantic relationships that exist among them. From this stage, the user can be sure if these data meet their requirements and expectations. If it is the case, he continues to properly publish the according data using *Pubby* (Fig. 2 (f)). We note that at each step the user is guided with instructions and is aware about the state of his advancement in the publishing process. Moreover, he couldn't proceed to the next step unless the previous has been correctly performed.

According to our data, the user which can be a public institution can obtain statistical data that he desires which improve his decision-making. In such context, multiple combinations of needs can be formulated via combining different facets of every data category. For instance, the public institution, in charge of helping jobless persons in finding jobs such as *ONEM* in Belgium, is interested to obtain statistics of workers per place of residence in order to strengthen its efforts in regions with low statistics more than other regions. Besides, the public organization that deals with social housing, such as *CPAS* in Belgium, is interested in knowing the proportion of beneficiaries in relation to the size of the population or the number of workers in the same family of these beneficiaries. The purpose is to adapt and regularize their decisions compared to statistics reflecting reality.

Our end-to-end framework includes a user-friendly interface that can be executed either from a web browser or through a standardized RESTful web API in order to enable easy integration into more complex workflows. *Pub-LOGD* permits to eliminate searches for ineffective tools and thus reduces the significant time it takes. Furthermore, it allows eliminating the problem of software tools dependencies and thus significantly reducing time in terms of tools research and combination. More details about our framework, its source code, and implementation aspects are available on https://github.com/123rabida123/WebAppOpenData.

V. Related Work

In the present section we discuss related work on two aspects: (i) Linked Open Data (LOD) technology, and (ii) Linked Open Government Data (LOGD)[14].

[14]https://www.w3.org/2011/gld/wiki/GLD-Life-cycle

TABLE II: Evaluation of related approaches

	Data Transforming	Data Interlinking	Data Storing	Data Visualizing	Data Publishing	User Assistance	Functional Completness	Interoperability	Operability
[13]	++	-	+	++	++	-	++	+	+
[7], [14]	++	++	++	-	++	-	++	+	++
Pub-LOGD	++	++	++	++	++	++	++	+	++

On the one hand, several challenges exist, in LOD context, in terms of the different data phases: extraction, storage, interlinking, enrichment, exploration and publishing aiming at an efficient exploitation of such data [15], [16]. The proposed works have separately deal with these steps. For instance, F. Benedetti et al. [17] have suggested an approach for automatic extraction of relevant knowledge sources from LOD sources. Freitas et al. in [18] present a survey of existing approaches tackling the heterogeneity of web-based datasets for searching and querying linked data. Hidalgo-Delgado et al. [19] have presented methodological guidelines to formalize the publishing process of LOD taking into account data quality. We also consider data quality in our work but with all the data phases forming its lifecycle. Whereas, some works have used the various data steps in different domains such as recommender systems [20] to propose solutions that leverage the data available within LOD datasets in order to recommend specific data to end users, or data indexing [21] to develop lookup indexes over Semantic web resources. Contrarily to our work, we aim at proposing a framework for governmental data.

On the other hand, several proposals have more focused on Linked Open Government Data (LOGD) in order to make such important data more transparent and fully exploited to public institutions as well as citizens. In [1], benefits and drawbacks about LOGD are outlined. Methodological guidelines for publishing Government Linked Data are also presented in [22] covering all the necessary data steps. In such context, diverse works have adopted a pipeline methodology to define their approaches. For instance, Maali et al. [8] have suggested an approach for converting raw government data into high-quality LOGD. While E. Tambouris [23] use such methodology to present multidimensional linked open data publication and reuse. Despite these works, there still some technical challenges for exploiting the full potential of LOGD and especially reusing related statistical data. To this end, our work aim at covering all the data phases from data extraction to data publishing and visualization in the context of government statistical data.

As shown in table II, we evaluate the approaches, that mainly publish LOGD, according to a set of criteria. These criteria represent (i) the different data processing steps defined in sectionII, and (ii) a set of qualitative metrics, specific to LOGD, defined in the literature [24]. The description of the qualitative metrics that we have adopted is presented in the table III. We note that contrarily to our work, the first work does not cover data interlinking and lacks taking into account the user assistance. Whereas the second work set does not offer to the user data visualizing step and does not highlight user guidelines.

TABLE III: Qualitative Metrics Description [24]

Qualitative Metric	Description
Functional Completness	The proposed platform has functionalities to browse, expand, analysis and visualisation, and is easy to use for beginners.
Interoperability	The platform is created interoperable. Considering data source on an interoperable format such as tabular CSV, and output data (Data Cubes, RDF and TTL).
Operability	If proper documentation, manual and also examples of usage are provided, the operation of platform is easy and transparent.

VI. Threads To Validity

Some potential threats to validity exist in our study. Firstly, our framework is based on the use of a set of predefined tools. We certainly need these tools yet the user is not confused neither in the choice among softwares ensuring same functionalities nor in the dependencies between them in order to achieve his goal. Besides, we have added the visualizing step that is fulfilled through *neo4j* which clarifies data knowledge. Additionally, we aim at extending each of these tools with novel functionalities that further customize the framework to the user requirements. Especially, we plan to add an automated framework that ensures the interoperability of LOGD at the beginning of our proposed process.

Secondly, we have shown the feasibility of our approach through a real case study from a local Government dataset. The work requests experimentations to further evaluate the effectiveness of our proposal. Indeed, qualitative and quantitative validation from a panel of users, including both experts and non experts, is being processed.

Finally, our work aims at achieving a semi-automated LOGD publishing. The manual intervention is important in this process to allow the citizens being aware of their data as well as giving him the access to eliminate unnecessary information, visualizing semantically-enriched LOGD, and exactly detecting the wrong step.

VII. Conclusions and Future work

In this paper, we presented an end-to-end framework (*Pub-LOGD*) to assist users in producing Linked Open Government Date (LOGD), by taking data quality into account. Our framework aims at shifting the burden of processing, converting, and publishing LOGD to data consumers rather than the government, in an easy and transparent way. It also allows to mitigate the problem of software tools dependencies.

The present work represents a first step towards producing and publishing statistical LOGD. In the near future, we intend

to further enrich our current framework. In particular, we are currently dealing with the heterogeneity issue of the data. We are also extending it by developing data exploration functionalities using *CubeViz*. As far as evaluation is concerned, we intend to better assess the quality of the published data, through a case study involving domain experts. Furthermore we are starting to consider larger and more diverse government datasets.

ACKNOWLEDGEMENTS

This work was carried on in the context of the European Regional Development Fund (ERDF) project IDEES, supported by the European Union and the Walloon region. The first author is supported by a CERUNA PhD fellowship at University of Namur.

REFERENCES

[1] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *IS Management*, vol. 29, no. 4, pp. 258–268, 2012. [Online]. Available: https://doi.org/10.1080/10580530.2012.716740

[2] "A systematic review of open government data initiatives," *Government Information Quarterly*, vol. 32, no. 4, pp. 399 – 418, 2015.

[3] T. Lebo and G. T. Williams, "Converting governmental datasets into linked data," in *Proceedings of the 6th International Conference on Semantic Systems*, ser. I-SEMANTICS '10. New York, NY, USA: ACM, 2010, pp. 38:1–38:3.

[4] A. Kohlhase, M. Kohlhase, and C. Lange, "Stex+: A system for flexible formalization of linked data," in *Proceedings of the 6th International Conference on Semantic Systems*, ser. I-SEMANTICS '10. New York, NY, USA: ACM, 2010, pp. 4:1–4:9. [Online]. Available: http://doi.acm.org/10.1145/1839707.1839712

[5] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia," *Semantic Web*, vol. 6, no. 2, pp. 167–195, 2015. [Online]. Available: https://doi.org/10.3233/SW-140134

[6] V. Presutti, F. Draicchio, e. T. A. Gangemi, Aldo", J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez, "Knowledge extraction based on discourse representation theory and linguistic frames," in *Knowledge Engineering and Knowledge Management*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 114–129.

[7] E. Kalampokis, E. Tambouris, and K. A. Tarabanis, "On publishing linked open government data," in *17th Panhellenic Conference on Informatics, PCI 2013, Thessaloniki, Greece - September 19 - 21, 2013*, 2013, pp. 25–32. [Online]. Available: https://doi.org/10.1145/2491845.2491869

[8] F. Maali, R. Cyganiak, and V. Peristeras, "A publishing pipeline for linked government data," in *The Semantic Web: Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings*, 2012, pp. 778–792. [Online]. Available: https://doi.org/10.1007/978-3-642-30284-8_59

[9] L. Ding, D. DiFranzo, A. Graves, J. R. Michaelis, X. Li, D. L. McGuinness, and J. A. Hendler, "Twc data-gov corpus: Incrementally generating linked government data from data.gov," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 1383–1386.

[10] K. Ham, "Openrefine (version 2.5). http://openrefine.org. free, open-source tool for cleaning and transforming data," *Journal of the Medical Library Association : JMLA*, 2013.

[11] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, ser. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers, 2011. [Online]. Available: https://doi.org/10.2200/S00334ED1V01Y201102WBE001

[12] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," 2009.

[13] E. Tambouris, E. Kalampokis, and K. A. Tarabanis, "Visualizing linked open statistical data to support public administration," in *Proceedings of the 18th Annual International Conference on Digital Government Research, DG.O 2017, Staten Island, NY, USA, June 7-9, 2017*, 2017, pp. 149–154. [Online]. Available: https://doi.org/10.1145/3085228.3085304

[14] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. Zheng, Z. Shangguan, J. Flores, D. L. McGuinness, and J. A. Hendler, "TWC LOGD: A portal for linked open government data ecosystems," *J. Web Semant.*, vol. 9, no. 3, pp. 325–333, 2011. [Online]. Available: https://doi.org/10.1016/j.websem.2011.06.002

[15] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," *Int. J. Semantic Web Inf. Syst.*, vol. 5, no. 3, pp. 1–22, 2009. [Online]. Available: https://doi.org/10.4018/jswis.2009081901

[16] N. Fayyaz, I. Ullah, and S. Khusro, "On the current state of linked open data: Issues, challenges, and future directions," *Int. J. Semantic Web Inf. Syst.*, vol. 14, no. 4, pp. 110–128, 2018. [Online]. Available: https://doi.org/10.4018/IJSWIS.2018100106

[17] F. Benedetti, S. Bergamaschi, and L. Po, "Online index extraction from linked open data sources," in *Proceedings of the Second International Workshop on Linked Data for Information Extraction (LD4IE 2014) co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 20, 2014.*, 2014, pp. 9–20. [Online]. Available: http://ceur-ws.org/Vol-1267/LD4IE2014_Benedetti.pdf

[18] A. Freitas, E. Curry, J. G. Oliveira, and S. O'Riain, "Querying heterogeneous datasets on the linked data web: Challenges, approaches, and trends," *IEEE Internet Computing*, vol. 16, no. 1, pp. 24–33, 2012. [Online]. Available: https://doi.org/10.1109/MIC.2011.141

[19] Y. Hidalgo-Delgado, R. Estrada-Nelson, B. Xu, B. Villazon-Terrazas, A. Leiva-Mederos, and A. Tello, "Methodological guidelines for publishing library data as linked data," in *2017 International Conference on Information Systems and Computer Science (INCISCOS)*, Nov 2017, pp. 241–246.

[20] T. D. Noia and P. Tomeo, "Recommender systems based on linked open data," in *Encyclopedia of Social Network Analysis and Mining, 2nd Edition*, 2018. [Online]. Available: https://doi.org/10.1007/978-1-4939-7131-2_110165

[21] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, H. Stenzhorn, and G. Tummarello, "Sindice.com: a document-oriented lookup index for open linked data," *IJMSO*, vol. 3, no. 1, pp. 37–52, 2008. [Online]. Available: https://doi.org/10.1504/IJMSO.2008.021204

[22] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, *Methodological Guidelines for Publishing Government Linked Data*. New York, NY: Springer New York, 2011, pp. 27–49.

[23] E. Tambouris, "Multidimensional open government data," 11 2016, p. 11.

[24] R. Matheus and M. Janssen, "An evaluation framework for linked open statistical data in government," in *Electronic Government - 16th IFIP WG 8.5 International Conference, EGOV 2017, St. Petersburg, Russia, September 4-7, 2017, Proceedings*, 2017, pp. 255–263. [Online]. Available: https://doi.org/10.1007/978-3-319-64677-0_21