Linking Open Drug Data: Lessons Learned

Guma Lakshen[1], Valentina Janev[2][0000-0002-9794-8505] and Sanja Vraneš[2]

[1] School of Electrical Engineering, University of Belgrade, Serbia
[2] Mihajlo Pupin Institute, University of Belgrade, Serbia
jlackshen65@yahoo.com valentina.janev@institutepupin.com
sanja.vranes@institutepupin.com

Abstract. Linked Open Data illustrates the concept that provides an optimum solution for information and dissemination of data, through the representation of the data in an open machine-readable format and to interlink it from diverse repositories to enable diverse usage scenarios for both humans and machines. The pharmaceutical/drug industry was among the first that validated the applicability of the approach for interlinking and publishing open linked data. This paper examines in detail the process of building Linked Data application taking into consideration the possibility of reusing recently published datasets and tools. Main conclusions derived from this study are that making drug datasets accessible and publish it in an open manner in linkable format adds great value by integration to other notable datasets. Yet, open issues arose clearly when trying to apply the approach to datasets coded in languages other than English, for instance, in Arabic languages.

Keywords: Linked Data; Drugs Application: Arabic Datasets; Quality; Lessons Learned.

1 Introduction

The World Wide Web has emerged in 1989 with an objective to tackle the widely spread difficulties to exchange information between different systems that arose in the 1980s, such as incompatible networks, disk/data formats, as well as character-encoding schemes [1]. The power the web possesses today lies in the amount of information it holds that represents a goldmine for data-driven applications and services. Its weakness, however, is caused by a potential design flaw: it was envisioned as a virtual documentation system. Storing the documents in unstructured form makes extracting information a manual and often tedious task. To obtain information, perform detailed analysis, create new products or additional documents, an efficient approach is needed to code the data and describe the data sources. More precisely, we need a standard, machine-readable format that allows for large-scale integration of, and reasoning on, data on the Web.

1.1 About Linked Data Approach

In 2001, Sir Tim Berners-Lee, the Director of the Wide Web Consortium outlined his vision for the Semantic Web as an extension of the conventional Web and as a world-wide distributed architecture where data and services easily interoperate [2]. Additionally, in 2006, Berners-Lee proposed the basic principles for interlinking linking datasets on the Web through references to common concepts known as Linked Data principles [3].

Thus Linked Open Data (LOD) movement was initiated for organizations to make their existing data available in a machine-readable format. In the last decade, the Linked Data approach has been adopted by an increasing number of data providers leading to the creation of a global data space that contains many billions of assertions—the LOD cloud, http://lod-cloud.net/. The cloud has increased from 12 datasets in 2007 to 1,229 with 16,113 links (as of April 2019, https://www.lod-cloud.net/).

1.2 Motivation

Due to the standardization and development of semantic web technologies, data being published on the web as linked data added tremendous value to institutions, research centers, and enterprises. In the drug industry, the rapidly increasing amount of data on the web opens new opportunities for integrating and enhancing drug knowledge on a global scale. The pharmaceutical/drug industry was leading others in expressing interest in validating the approach for publishing and integrating open data. Linked Open Drug Data LODD endpoint was created in 2011, https://www.w3.org/wiki/HCLSIG/LODD [4], which is a set of linked datasets related to Drug Discovery. It includes data from several datasets including DrugBank (https://www.drugbank.ca/), DailyMed, LinkedCT, SIDER, ClinicalTrials.gov, RxNorm, and NCBI Entrez Gene, a detailed comparison of the LODD datasets can be accessed at https://www.w3.org/wiki/HCLSIG/LODD/Data, notably this page was last updated on 28th December 2012. Later, in 2014, the 3rd release of Bio2RDF (http://bio2rdf.org/ or https://github.com/bio2rdf/) was published as the largest network of Linked Data for the Life Sciences (35 datasets).

This motivates the authors to examine the use of available Linked Data tools for interlinking local drug datasets (from one or more countries) with datasets integrated in the LOD cloud. In this paper the authors proposes improvements in the Linked Data lifecycle in tasks related to quality assessment in the process of consuming the open data. The framework enhances the existing methodology of integrating (transforming / interlinking) open data and proposes integration of quality assessment services for improving the overall performance of the Linked Drug Data application.

The paper is structured as follows. Section II introduces the business objectives, gives a review on Linked Data methodologies and proposes an approach for building a Linked Data application. Section III presents in detail the process of transforming the Arabic drug datasets as a linked data and discusses the results and benefits for end-users. Section IV points to lessons learned in relation to the flexibility of using the created graph and challenges with quality assessment issues. Section V concludes the paper with hints for future work.

2 Building Linked Data Apps

2.1 Pharmaceutical / Drug Industry: Business Objectives

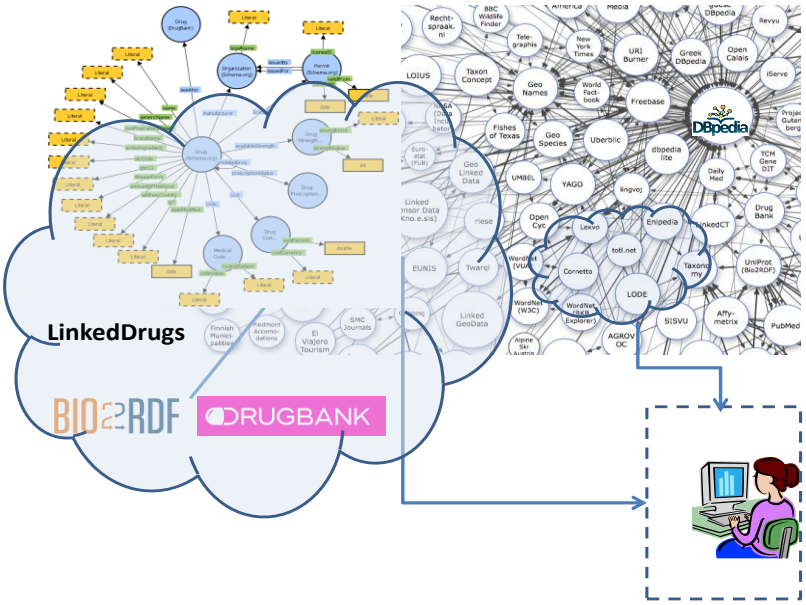

Fig. 1. Integrating public and private datasets.

The goal of the target application is to enable end-users to answer inquiries about drug availability in the open datasets (e.g. DrugBank, DBpedia, see Table 1) and enrichment of local data with information from the Web. The end-user will benefit from the interlinking of private datasets with public data and enrichment of local data with information from the Web.

Examples of key business queries are:

1. For a particular drug, retrieve relative information in the Arabic language (if it exists) from other identified datasets, such as DrugBank and DBpedia.
2. For a particular drug, retrieve equivalent drugs, and compare their active ingredients, contradictions, and prices.
3. For a particular drug, retrieve valuable information about equivalent drugs with different commercial names, manufacturers, strengths, forms, prices, etc.
4. For a particular drug, retrieve its reference information to highlight possible contradiction, e.g., in combination with other drugs, allergies, or special cases (e.g., pregnancy).
5. For a particular active ingredient, retrieve advanced clinical information, i.e., pharmacological action, pharmacokinetics, etc.
6. For a particular drug, retrieve its cost, manufacturer, and country.

Table 1. Linked datasets - examples

DataSet	Description
DrugBank	DrugBank is a web-enabled database containing comprehensive molecular information about drugs, their mechanisms, their interactions, and their targets. First described in 2006, DrugBank has continued to evolve over the past 12 years in response to marked improvements to web standards and changing needs for drug research and development. See https://www.drugbank.ca/
DBpedia	DBpedia is an ongoing project designed to extract structured data from Wikipedia. It contains RDF data about 2.49 million things out of which is 218 million triples describing 2300 drugs. DBpedia is updated every three months. See https://www.DBpedia.org

2.2 Survey of Linked Data Methodologies

In literature, not many papers dealt with Linked Data methodologies i.e. the process of generating, linking, publishing and using Linked Data, to name a few: *Best Practices for Publishing Linked Data* (W3C-Government Linked Data Working Group, 2014) [5], *A Cookbook for Publishing Linked Government Data on the Web* (Hyland et al., 2011) [6], *Linked Data Life Cycles* (Hausenblas et al., 2016) [7], *Guidelines for Publishing Government Linked Data* (Villazón-Terrazas et al., 2011) [8], *Managing the Life-Cycle of Linked Data with the LOD2 Stack* (Auer, et al., 2012) [9], and *Methodological guidelines for consolidating drug data* (Jovanović and Trajanov; 2017) [10], see table 2 below for a brief comparison. One of the first Linked Data methodologies was developed in the European research project LOD2 (2010-2014)[9] that was mainly dedicated to the publishing process, i.e. opening the data in a machine-readable format and establishing the prerequisite tools and technologies for interlinking and integration of heterogeneous data sources in general. Jovanović and Trajanov [10] proposed a new Linked Data methodology with a focus on reuse which provides guidelines to data publishers on defining reusable components in the form of tools and schemas/services for the given domain (i.e. drug management).

Table 2. Survey on linked data methodologies

Authors	Title / Steps	
W3C Government Linked Data Working Group (2014) [5]	***Best Practices for Publishing Linked Data*:**	
	(1) Prepare stakeholders, (2) Select a dataset, (3) Model the data, (4) Specify an appropriate license, (5) Good URIs for linked data, (6) Use standard vocabularies,	*Initialization*
	(7) Convert data, (8) Provide machine access to data,	*Innovation*
	(9) Announce new data sets, (10) Recognize the social contract	*Validation &Maintenance*
Hyland et al. (2011) [6]	***A Cookbook for Publishing Linked Government Data on the Web*:**	
	(1) Identify, (2) Model, (3) Name, (4) Describe,	*Initialization*
	(5) Convert, (6) Publish,	*Innovation*
	(7) Maintain	*Validation*

		&Maintenance
Hausenblas et al. (2016) [7]	***Linked Data Life Cycles*:**	
	(1) Data awareness, (2) Modeling,	*Initialization*
	(3) Publishing, (4) Discovery, (5) Integration,	*Innovation*
	(6) Use-cases	*Validation &Maintenance*
Villazón-Terrazas et al. (2011) [8]	***Guidelines for Publishing Government Linked Data*:**	
	(1) Specify, (2) Model,	*Initialization*
	(3) Generate, (4) Publish,	*Innovation*
	(5) Exploit	*Validation &Maintenance*
Auer, et all. (2012) [9]	***Managing the Life-Cycle of Linked Data with the LOD2 Stack*:**	
	(1) Extraction,	*Initialization*
	(2) Storage, (3) Authoring, (4) Interlinking, (5) Classification,	*Innovation*
	(6) Quality, (7) Evolution/Repair, (8) Search/ Browsing/ Exploration	*Validation &Maintenance*
Jovanovik and Trajanov (2017) [10]	***Methodological guidelines for consolidating drug data*:**	
	(1) Domain and Data Knowledge, (2) Data Modeling and Alignment,	*Initialization*
	(3) Transformation into 5-star Linked Data, (4) Publishing the Linked Data Dataset on the Web,	*Innovation*
	(5) Use-cases, Applications and Services	*Validation &Maintenance*

2.3 Proposal for a Piloting Methodology

Taking into consideration that end-use organization might be interested to implement innovations in existing drug data management, the authors propose to split the implementation of the Linked Data application development into three phases as follows [11]:

Phase I: INITIALIZATION

Business objectives and requirements: Requirement specification, technical characterization, and setting up of the demo site; establishing acceptance (success) criteria for pilot applications validation based on performance characteristics, usability, as well as EU and national regulations (e.g., related to data access and security measures);

Data categorization and description: Analysis of the datasets to be published in linked data format and selection of vocabularies and development other specifications for metadata description.

Phase II: INNOVATION

Integrating datasets in the form of a knowledge graph: Data access, transformation, and enrichment.

Generic component selection and tool customization for the pilot applications: Customization of linked data components for use in the targeted domain.

Specific tools development: Integration of security measures to deal with possible communication threats.

Phase III: VALIDATION

Continuous validation of the open-source tools that have been reused, providing feedback for improving the solution components, and testing for imperfect data.

3 Testing the Piloting Methodology in the Drug Domain

The authors tested the proposed methodology for development of the Linked Drug Data Application with datasets from the pharmaceutical/drug industry from the Arabic region. The Arab world also known as the Arab nation currently consists of the 22 Arab countries of the Arab League and has a combined population of around 422 million inhabitants. The Arabic language content in the World Wide Web is less than 3%; the situation is even worse regarding Arabic open data, Arabic linked data, and Arabic drug open linked data. As far as medical data available in the Arab region, there are only a handful of Arabic drug applications such as Webteb [12], Altibbi [13], and 123esaaf [14], which provide their services in Arabic and English, but unfortunately, the data is not open and most are not free. Arabic language content on the web is less than 3%. The situation is even worse regarding Arabic open data, linked data, and open linked data on drugs. This limitation of Arabic content encourages the researcher to enrich the Arabic user experience by utilizing semantic web technologies to interlink their data with other languages, including English.

3.1 Phase I - Data Categorization and Description

As a use case scenario, the authors selected four drug data files from four different Arabic countries, Iraq, Saudi Arabia, Syria, and Lebanon as shown in Table 3. Most of the open published files in the Arab region are either in PDF or XLS format. The reasons for choosing XLS format were data fidelity, ability to source from a wider range of public sector domains, and to have increased value that comes from many information linkages. The authors believe that for many years to come, more drug data will be published in XLS format in the Arab countries.

The selected datasets are open data published by health ministries or equivalent bodies in the respected governments. They are regularly updated, usually after a two-year period. As it can be noticed from the difference in the number of columns, the structure of the datasets is not unified, which makes the unification and mapping of data necessary.

The data quality of the selected files is too low; most of XLS documents do not represent the generic name or their ATC code which makes the data almost unusable for further transformation.

Table 3. Selected Arabic open drug datasets

Country	No. Of Tuples	No. of Columns
Iraq	9090	9
Lebanon	5822	15
Saudi Arabia	6386	10
Syria	9375	7

3.2 Phase II - Integrating Datasets in a form of a Knowledge Graph

In what follows we will describe the steps required in the process of transforming, and linking the Arabic drug data in a form of a knowledge graph.

Data Cleaning

OpenRefine (http://openrefine.orgVersion 2.6-rc1) was used to clean the selected data in order to make it coherent and ready for further operations according to the methodology. A well-organized cleaning operation minimizes inconsistencies and ensures data standardization among a verity of data sources.

Ontology Definition and Data Mapping Schema

Some of the ontologies and vocabularies which a data publisher needs to have in mind biomedical ontologies. The schema comprises classes and properties are Schema.org (https://schema.org/), DBpedia Ontology UMBEL (http://umbel.org/), DICOM (https://www.dicomstandard.org/), and the DrugBank Ontology used, as well as other from the Schema.org vocabulary: the *schema:Drug* class (https://health-lifesci.schema.org/Drug), along with a large set of properties which instances of the class can have, such as generic *drug name, code, active substances, non-proprietary Name, strength value, cost per unit, manufacturer, related drug, description, URL, license, etc.* Additionally, in order to align the drug data with generic drugs from DrugBank, properties *brandName, genericName, atcCode and dosageForm* from the DrugBank Ontology were used. The relation rdfs:seeAlso can be used to annotate the links which the drug product entities will have to generic drug entities from the LOD Cloud dataset. The nodes are linked according to the relations these classes, tables or groups have between them. There exist a few tools for ontology and vocabulary discovery which should be used in this operation such as Linked Open Vocabularies (LOV, http://lov.okfn.org/) and DERI Vocabularies (http://datahub.io).

Data Conversion

Create RDF dataset: The previously mapped schema can produce an RDF graph by using RDF-extension of LODRefine tool. This step transforms raw data into RDF dataset based on a serialization format. Transformation process can be executed in many different ways, and with various software tools, e.g. OpenRefine (which the authors used), RDF Mapping Language (http://rml.io/spec.html), and XLWrap (http://xlwrap.sourceforge.net/), among others.

Interlinking

LODRefine was used for reconciliation in interlinking the data. In this case, columns atcCode, genericName1, activeSubstance1, activeSubstance2 and activeSubstance3 reconciled with DBpedia. This operation enables interoperability between organization data and the Web through establishing semantic links between the source dataset (organization data) with related datasets on the Web. Link discovery can be performed in manual, semi-automated, or fully-automated modes to help discover links between the source and target datasets. Since the manual mode is tedious, error-prone, and time-consuming, and the fully-automated mode is currently unavailable, the semi-automated mode is preferred and reliable. Link generation yields links in RDF format using *rdfs:seeAlso* or *owl:sameAs* predicates. The activities of link discovery and link generation are performed sequentially for each data source. The last activity within the interlinking stage is the generation of overall link statistics which showcase the total number of links generated between the source and target data sources.

Storage/SPARQL Endpoints

OpenLink virtuoso server (version 06.01.3127, https://virtuoso.openlinksw.com/) on Linux (x86_64-pc-Linux-gnu), Single Server Edition was used to run the *SPARQL endpoint.*

Publication

RDF graph can be accessed on the following link: http://aldda.b1.finki.ukim.mk/. For publishing linked data on the web, a linked data API is needed, which makes a connection with the database to answer specific queries. The HTTP endpoint is a webpage that forms the interface. A REST API is used to make a web application. It makes it possible to give the linked data back to the user in various formats, depending on the user's requirements. The linked data can be made visible in HTML on a website as HTTP links or as RDF data in a browser or a graphic visualization in a web application, which would be the most user-friendly.

3.3 Phase II - Specific tools development

The authors have made experiments and tested the quality of the public datasets that are used for enrichment, in particular the DBpedia. The DBpedia knowledge base contains information on many different domains that is automatically extracted from Wikipedia, based on infoboxes. The automatic extraction has obvious advantages in terms of speed and quantity of data (ensures wide coverage), but it also poses some quality issues. When it comes to quality assessment of the DBpedia Arabic Chapter, there are problems specific to the Arabic language that result in:

1. Presentation of characters as symbols via web browsers due to errors during the extraction process.
2. Wrong values in numerical data, due to the use of Hindu numerals in some Arabic sources.

3. Occurrence of different names for the same attribute, for instance, the birth date attribute appears in various infoboxes by different names: one time as "الم يلاد تـ اريـ خ" [birth date], another time as "الـ ولادة تـ اريـ خ" [delivery date], the third time as "الـ م يلاد" [birth].
4. Inconsistency of names between the infobox and its template; for instance, there is a template called "مديـ نة" [city] while the infobox name is called "مديـ نة مـ عـ لومات" [city information].
5. Geo-names templates formatting problems when placed in the infobox.
6. Errors in *owl:sameAs* relations and problems in identifying the *owl:sameAs* relations due to heterogeneity in different data sources.

However, some of the problems present in other DBpedia chapters are also identified in the Arabic chapter, specifically:

7. Wrong Wikipedia Infobox information; for example, the height of minaret of the grand mosque in Mecca (the most valuable mosque for all Muslims) is given as 1.89 m, where the correct height is 89 m.
8. Mapping problems from Wikipedia, such as unavailability of infoboxes for many Arabic articles; for example, "Man-made river in Libya الـ نهر الـ صـ ناعي,"which is considered as the biggest water pipeline project in the world, or not containing all the desired information.
9. Object values incompletely or incorrectly extracted.
10. Data type incorrectly extracted.
11. Some templates may be more abstract, thus cannot map to a specific class.
12. Some templates are unused or missing inside the articles.

Hence, what is needed when working with open data is tools for quality assessment of datasets prior to interlinking with the private datasets.

3.4 Phase III – Validation of Scenarios for Querying Pharmaceutical / Drug Datasets

Visualization and Querying.

After publishing the data in Linked Data format it becomes available to other web applications for retrieving and visualization. The use of standard vocabularies for modeling the data allows the end-users to use different visualization opportunities e.g. freely available libraries can be used that offer diverse types of visualization such as a table or in a diagram formatted in different ways. Custom visualization applications can be used to enable a user to interact with data.

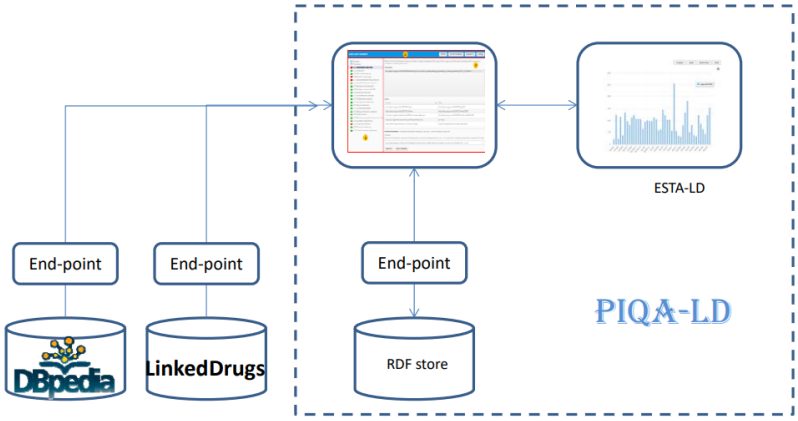

Fig. 2. Knowledge Graph Visualization and Querying.

4 Lessons Learned

4.1 About the Linked Data Format and the Knowledge Graph

The linked data approach, based on principles defined back in 2006 and on best practices for publishing and connecting structured data on the web (as elaborated by ICT experts), can play an important role in the domain of semantic interoperability. Web data, including drug data, is published most often in a two-star format data, i.e., PDF or XLS format (see Berners-Lee's categorization of open data). However, the authors decided to use the RDF, because it is recommended by W3C, and it has advantages, such as an extensible schema, self-describing data, de-referenceable URIs. Further, since RDF links are typed, it enables good structure, interoperability, and safely linking different datasets.

Before converting XLS data to RDF, the authors selected a target ontology to describe the drugs contained in the drug availability dataset. Authors selected the LinkedDrugs ontology [10], Schema.org vocabulary, and DBpedia, as they have the needed properties and provide easier interlinking possibilities for further transformation. The Web Ontology Language allows complex logical reasoning and consistency checking for RDF and OWL resources. These reasoning capabilities helped the authors to harmonize the heterogeneous data structures found in the input datasets.

4.2 About the Quality of Open Data

Many authors have pointed out issues such as the completeness, conciseness, and consistency of open data. In 2014, Kontostas et al. [15] provided several automatic quality tests on LOD datasets based on patterns modeling various error cases, and they detected 63 million errors among 817 million triples. At the same time, Zaveri et al. [16], conducted a user-driven quality evaluation which stated that DBpedia indeed has quality problems (e.g., around 12% of the evaluated triples had issues). They can

be summarized as incorrect or missing values, incorrect data types, and incorrect links. Based on the survey, the authors developed a comprehensive quality assessment framework based on 18 quality dimensions and 69 metrics. Based on the work of Zaveri et al. [17] and the ISO 25012 DQ model, Radulović et al. [18] developed a linked data quality model and tested the model with DBpedia with a special focus on accessibility quality characteristics.

Based on the analysis of quality issues with DBpedia [19] and the problems identified, the authors conclude that most important dimensions to be taken into consideration are Accuracy (triple incorrectly extracted; Data type problems; errors in implicit relationship between attributes); Consistency (Representation of number values) and Relevancy (Irrelevant information extracted).

5 Conclusions and Future Work

Most of the available drug datasets nowadays are still provided in 2-star format and in English language due to the fact that the English language is widespread among physicians and pharmacists and also a predominant language in communications between physicians and pharmacists. In order to showcase the possibilities for large-scale integration of drug data, the authors proposed a piloting methodology and tested the approach with datasets from Arabic countries. The authors presented the transformation process of 2-star drug data into a 5-star Linked Open Data with DrugBank and DBpedia. The OpenLink virtuoso server (version 06.01.3127) on Linux (x86_64-pc-linux-gnu), Single Server Edition is used to run our SPARQL endpoint (see http://aldda.b1.finki.ukim.mk/sparql).

The paper showcases the benefits from the Linked Data approach and for the first time discusses the issues with drug data from Arabic countries (authors selected four drug data files from four different Arabic countries, Iraq, Syria, Saudi Arabia, and Lebanon).

Taking into consideration the issues identified with quality of the open data (in particular, the issues with drug data from Arabic countries), the future work will include implementation of a stable and open-source web applications that will allow the end-user to fully explore and assess the quality of the consolidated dataset, and if possible, to repair the errors observed in the Arabic Linked Drug dataset.

Acknowledgments. The research presented in this paper is partly financed by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010) and partly by the EU project LAMBDA (GA No. 809965).

References

1. Halpin, H.: Social Semantics: The Search for Meaning on the Web. Semantic Web and Beyond 13. Springer, New York (2013).
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001).

3. Berners-Lee, T.: Design issues: Linked Data, http://www.w3.org/DesignIssues/LinkedData.html, last accessed 2019/05/01.
4. Jentzsch A., et al.: Linking Open Drug Data. Triplification Challenge of the International Conference on Semantic Systems. 2009.
5. W3C, Best Practices for Publishing Linked Data, http://www.w3.org/TR/ld-bp/ (2016), last accessed 2019/05/01.
6. Hyland, B., Wood D.: The Joy of Data: A Cookbook for Publishing Linked Government Data on the Web. In: Linking Government Data, New York: Springer New York;. 2011, pp. 3–26.
7. Hausenblas, M.: Linked Data Life Cycles, 2016. http://www.slideshare.net/mediasemanticweb/linked-data-life-cycles.
8. Villazón-Terrazas, et al.: Methodological Guidelines for Publishing Government Linked Data. In: Wood, D. (ed.) Linking Government Data, Ch. 2, 2011.
9. Auer, S., et al.: Managing the Life-Cycle of Linked Data with the LOD2 Stack. In: The Semantic Web-ISWC 2012. Boston: Springer Berlin Heidelberg, pp. 1–16, 2012.
10. Jovanovik, M., Trajanov D.: Consolidating drug data on a global scale using linked data. Journal of Biomedical Semantics, 8(3), 2017.
11. Janev, V., Mijović, V., Milosević, U., Vraneš, S.: Linked Data Apps: Lessons Learned. In: Dimitar Trajanov, Verica Bakeva (Eds) ICT Innovations 2017 Web proceedings ISSN 1865-0937, Communications in Computer and Information Science book series (CCIS, volume 778)
12. WebTeb, https://www.webteb.com/aboutusen
13. Altibbi, https://www.altibbi.com/
14. 123esaaf, https://www.123esaaf.com/
15. Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R.:Test driven Evaluation of Linked Data Quality. In Proceeding of the 23rd International Conference on World Wide Web. New York, NY, USA, 2014, pp. 747–758. DOI: http://dx.doi.org/10.1145/2566486.2568002 (2014)
16. Zaveri, A., Kontokostas, D., Sherif, M. A., Bühmann, L., Morsey, M., Auer, S., Lehmann, J.: User-driven Quality Evaluation of DBpedia. In Proceedings of the 9th International Conference on Semantic Systems. New York, NY, USA, 2013, pp. 97–104 (2013)
17. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. Auer, S.: Quality assessment for linked data: A survey. Semantic Web – Interoperability, Usability, Applicability, Vol. 7, No. 1, 63-93. DOI: http://dx.doi.org/10.3233/SW-150175 (2016)
18. Radulović, F., Mihindukulasooriya, N., García-Castro, R., Gómez-Pérez, A.: A Comprehensive Quality Model for Linked Data. Semantic Web – Interoperability, Usability, Applicability, Vol. 9, No. 1 (2018), (2018), 3-24, Special issue on Quality Management of Semantic Web Assets (Data, Services and Systems). DOI: https://doi.org/10.3233/SW-170267 (2018)
19. Lackshen, G., Janev, V., Vraneš, S. Quality Assessment of Arabic DBpedia. In Proc. of 8th International Conference on Web Intelligence, Mining and Semantics. June 25 – 27 2018, Novi Sad, Serbia. ACM New York, NY, USA DOI: https://doi.org/10.1145/3227609.3227675 (2018)