Linked Data Apps: Lessons Learned

Valentina Janev[1], Vuk Mijović[1], Uroš Milošević[1], Sanja Vraneš[1]

[1] Mihajlo Pupin Institute, University of Belgrade, Serbia
{valentina.janev, vuk.mijovic, uros.milosevic, sanja.vranes}@pupin.rs

Abstract. The Linked Data approach, based on principles defined back in 2006 and best practices for publishing and connecting structured data on the Web, can play an important role for opening the government resources, making new citizen and business applications possible, which, in turn, creates (both tech and non-tech) business and job opportunities. This paper explores the challenges software engineers face when selecting and adopting technologies and tools for building Linked Data Apps development. Two different innovative applications: (1) the mobile supplier chain dashboard (SCD) tool and (2) the Geospatial-semantic Exploration on the Move (GEM) tool are presented in details. This paper contributes to the understanding of the Linked Data technologies and points to lessons learnt from recent EU research projects.

Keywords: Linked Data, Scenarios, Applications, Piloting methodology.

1 Introduction

The Linked Data approach is based on principles defined back in 2006 (Berners-Lee, 2006) and best practices for publishing and connecting structured data on the Web elaborated by ICT experts [1-3].

Linked Data enables datasets to be linked together through references to common concepts. HTTP URIs (*Uniform Resource Identifier*) are used to identify any entity or concept (people, locations, products, events etc.), so that data consumers are provided with more information when accessing data, including *relationships*, i.e. links to other related URIs. The standard for the representation of the information that describes those entities and concepts, and is returned by dereferencing the URIs, is RDF, see https://www.w3.org/RDF/. Through the use of Linked Data principles, RDF provides a mechanism for data publishing on the web which supports easy discovery and cross-linking of published data. RDF simple, but flexible representation uses binary predicates to describe resources. Tim Berners-Lee (2006) [4] proposed also a five star model for rating the Linked Open Data, where higher star ratings indicate more added value in the used approach. According to this rating, a five stars dataset is linked to other publishers' data to provide context.

In the last decade, Linked Data best practices have 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 Linked Open Data cloud, http://lod-cloud.net/. The cloud has been enlarged from 12 datasets in 2007 to 1,139 in January 2017.

Opening up i.e. making the data public provides citizens with easier access to services, greater transparency and understanding of services, and improved communication through feedback loops, which immediately results in greater understanding and insight for the planning and delivery of community resources and societal support. Moreover, publishing the data in its original form (raw data) makes new applications possible, which, in turn, creates (both tech and non-tech) business and job opportunities.

In this paper we point to scenarios for innovative use of open-source tools from the Linked Data Stack[1] [5], based on our experiences in recent EU projects. In the period 2011-2016, the Mihajlo Pupin Institute (PUPIN) was involved in three projects (LOD2, GeoKnow and SHARE-PSI) that besides the Linked Data Stack delivered also the SHARE-PSI Best Practices[2] [6] as recomendations for Directive for re-use of Public Sector Information [7].

The project LOD2 was crucial for the adoption of Linked Data technologies in government, as well as in an enterprise context. As part of the Linked Data Stack development, the PUPIN team designed and implemented several open-source tools and evaluated the Linked Data approach in several different scenarios with public and private datasets. Based on the experience gained in LOD2 and GeoKnow, in Section 3, we are proposing a methodology for piloting innovative apps based on Open Data.

Further on, in Sections 4 we showcase how Open Data can be used for enriching private datasets and supporting innovative applications in a supplier value chain, and in Section 5 we demonstrate the use of Open Data in touristic scenarios through the GEM tool. Section 6 concludes the paper.

2 Related Work

2.1 Linked Data Stack and Linked Data Methodologies

The Linked Data Stack [2,3,8] is a distribution platform for software components which support one or more aspects of the Linked Data life cycle composed of data storing, integration, linking to publishing, analysing and data visualization. The Linked Data Stack comprises a number of pre-configured tools for managing the life-cycle of Linked Data and thus simplifies the deployment and distribution of the tools. The pre-configuration ensures that the deployed components are able to interact with each other. Moreover, it eases the information flow between components to enhance the end-user experience while harmonising the look and feel.

The Linked Data methodology followed in the LOD2 (2010-2014) was mainly devoted to the publishing process, opening the data in a machine readable format and establishing the prerequisite tools and technologies for interlinking and integration of heterogeneous sources in general (government and enterprise context). The subse-

[1] http://stack.linkeddata.org/
[2] https://www.w3.org/2013/share-psi/bp/

quent GeoKnow (2013-2015) project has contributed additional tools for bringing geospatial data to the Web of Data and showcased the benefits of Linked Data in enterprise context. The delivered open-source frameworks were made available for further use via the in the Debian repository of the Linked Data Stack, see also the GitHub repository (https://github.com/geoknow). However, commercial partners used the opportunity to initiate new data-driven business model and continued to develop the prototype tools toward fully operable and enterprise-ready solutions.

Having in mind the position and needs of Open Data consumers, in our opinion, there is a space for further development of Linked Data methodologies toward a standard software engineering methodology. Often the Linked Data methodologies that can be found in literature [9] are related mainly to the publishing process. In [10], Jovanović proposed a new Linked Data methodology with a focus on reuse. It provides guidance to data publishers on defining reusable components in the form of tools, schemas and services, for the given domain.

2.2 Linked Data Scenarios

Figure 1 presents different scenarios for opening and consuming Open Data that were tested in the Mihajlo Pupin Institute in the period 2011-2016. Further details are provided in Table 1.

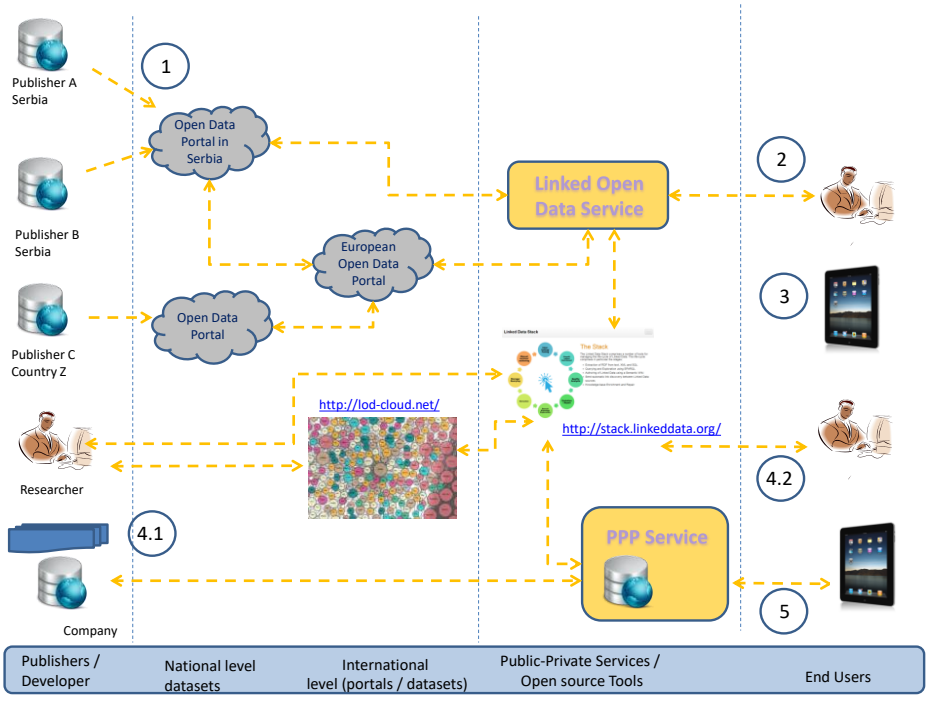


Fig. 1. Open Data Scenarios.

Scenario 1 aims at sharing public sector information in machine readable format and, via a national portal (e.g. the Serbian CKAN [11]), and integration of datasets on the European Open Data Portal. This scenario covers the publishing process for statistical data. Once available under a public license, Open Data from different data sources can be integrating and analyzed / visualized using Linked Data tools e.g. the ESTA-LD[3] [12] tool in Scenario 2.

Scenario 2 is devoted to the consuming operation and aims at re-using the available government sector data in innovative applications.

In this paper we will discuss in detail the Case 3 i.e. leveraging available *crowd-sourced datasets such as* DBpedia, http://wiki.dbpedia.org/, and LinkedGeoData, http://linkedgeodata.org/ in tourist scenarios, see Section 5.

Scenario 4 and Scenario 5 illustrates the activities of a company interested to use public data for commercial purposes. Section 4 discusses the case in detail.

Table 1. Application scenarios and the corresponding PUPIN tools.

Data type	Scenarios	Tool	Features
Scenario 1 and 2: Statistical data (public)	Automation of the statistical production process (Publishing)	Statistical Workbench (in cooperation with LOD2 partners) [5]	Import / Export / Interlinking / Exploration / Visualization / Publishing
	Monitoring socio-economic indicators (Analytics)	Exploratory Spatio-Temporal Analysis of Linked Data, 2015 [12]	Transformation / Exploration / Visualization
Scenario 3: Geospatial data (public)	Route planning Motive-based search	Geospatial-semantic Exploration on the Move [13]	Authoring / Exploration /
Scenario 4: Unstructured data (private / public)	Improved management of unstructured data	Multilingual and multifunctional Natural Language Processing (NLP) and Linked Data tool, 2014 [14]	Dictionary management / Text enrichment / Document similarity search
S5: Structured / Unstructured (private)	Supply-chain management	Mobile Supply-chain dashboard, 2015 [15]	Exploration / Visualization

3 Piloting Linked Data Apps

The process of developing a new pilot application using open-source Linked Data tools (e.g. from the Linked Data Stack) can be divided into three phases, namely Ini-

[3] http://geoknow.imp.bg.ac.rs/ESTA-LD

tialization, Innovation and Validation (see Figure 2). The Innovation phase is dedicated to further development of selected Linked Data components (vocabularies, taxonomies, open-source tools) and their customization to the target application. Often this phase include integration with existing enterprise systems and adoption of proven technologies for the benefits of the end-user organization. At the end of the Validation phase, the fully operable and enterprise-ready tools are put on market.

3.1 Phase I - Initialization

- **Business Objectives & Requirements:** Requirement Specification, Technical Characterization and setting up of 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;

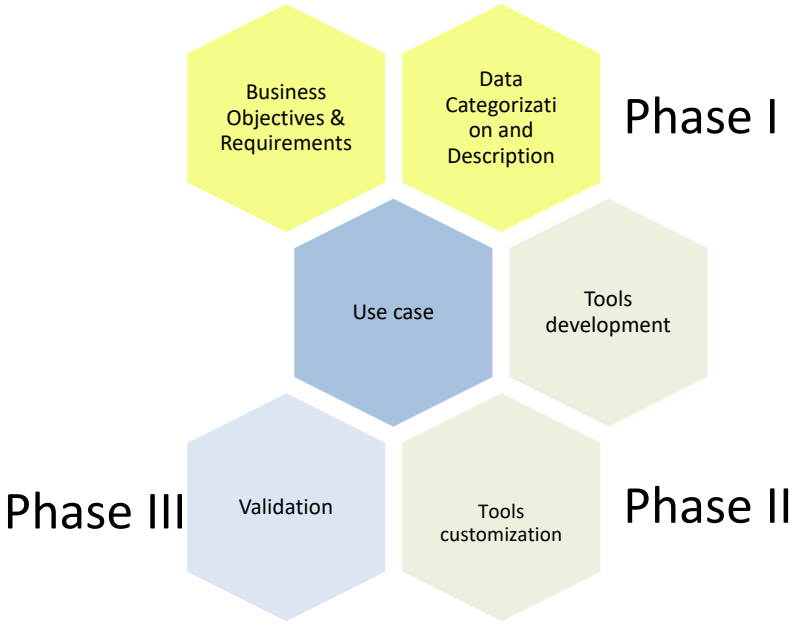


Fig. 2. Piloting Methodology.

3.2 Phase II - Innovation

- **Components selection and Tools development**: Data access, transformation and enrichment, Integration of security measures in order to deal with possible communication threats;
- **Tools customization for the Pilot Applications**: Customization of Linked Data components for use in the targeted domain (statistical data publication, navigation and situational awareness, etc.).

3.3 Phase III - Validation

- **Continuous validation** of the open-source tools that have been re-used, providing feedback for improving the solution components; tests with imperfect data;
- **Testing and replication** with business users and citizens.

4 Mobile Supplier Chain Dashboard - Tool

4.1 Leveraging Open Data for More Informed Decision Making

The goal of this target application was to prove the applicability of Open Data in business scenarios i.e. development of Mobile supply chain dashboard that provides comfortable user experience on mobile devices and allow the user to retrieve information from private data sources (suppliers, orders and shipments) and interlink with public information (news, weather data). The goal was achieved by the development of the mobile supply chain dashboard, a mobile application that delivers a mobile friendly view of the whole supply chain that keeps decision makers informed and allow them to make decisions on the go.

4.2 Applying the Piloting Methodology

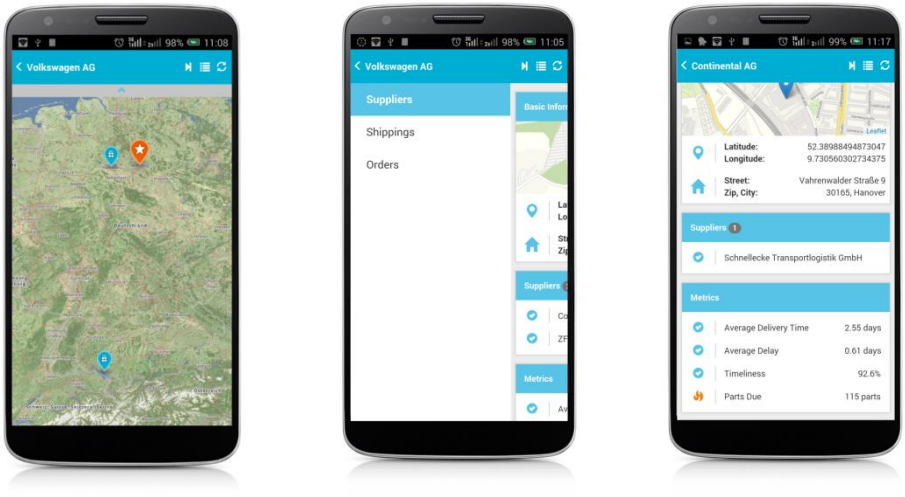
In a realistic scenario a set of tools are needed for real time access to message flow in a supply chain network, pre-processing and linking with external background knowledge and visualization in real time.

In *Phase I* data sources were identified and vocabularies selected for extraction of order and shipment information from EDI messages and weather and news information in RDF format [16].

In *Phase II* reusable components were leveraged for automating the data processing flow in the background. Additionally, mobile supply chain dashboard was developed and customized to allow easier navigation of the dashboard functionalities on smartphones thus enabling supply chain managers to receive notifications, gain insights into the state of the supply chain management, and react on the go [15].

In *Phase III* the application was tested in laboratory environment [15], see Figure 3.

Fig. 3. Navigation elements: a) menu; b) supplier map view; c) supplier metrics.



5 Geospatial Exploration on the Move - Tool

5.1 Leveraging Open Data for Navigation and Exploration

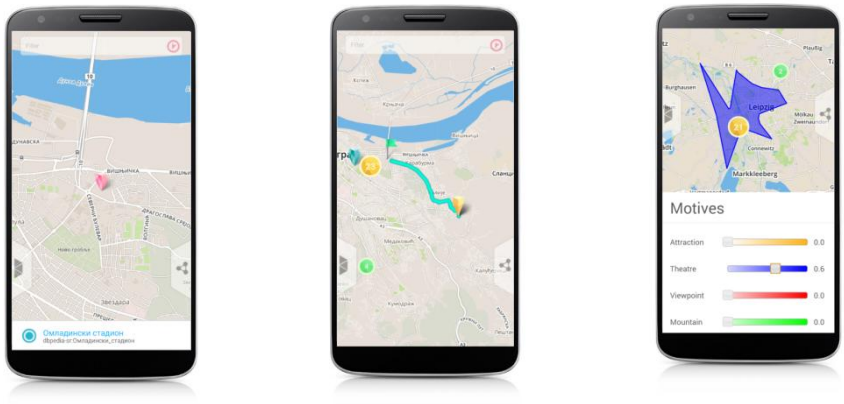
The goal of the Geospatial Exploration on the Move (GEM) application [13] was to provide highly customizable and information rich slippy map to the geospatial data consumers on the move. An example scenario of the use of GEM is as follows: end-user (e.g. tourist, see Figure 6) searches and retrieves touristic information efficiently by leveraging available Open Data. The application answers specific questions, such as ”I am in Belgrade. Retrieve the location of sculptures by Meštrović” (*Semantic filtering functionality*); ”I am at the School of Electrical Engineering and on my way to Kalemegdan. Retrieve the most interesting monuments in the neighbourhood as well as the restaurants along with the opinions of the visitors” (*Semantic routing functionality*); ”I am in Metropol hotel. Display area with shops where I can buy cheapest cloths for my kid” (*Motive-based search functionality*);

5.2 Applying the Piloting Methodology

In *Phase I* data sources were identified that can be used for retrieving public data resources such as DBpedia, http://wiki.dbpedia.org/, and LinkedGeoData, http://linkedgeodata.org/ as architecture of the future GEM component was defined. The data is exchanged in RDF and retrieved with SPARQL queries from SPARQL end-points. All data format heterogeneity is hidden from the user.

In *Phase II* reusable components were selected for building the application: LeafletJS[4], a modern open-source JavaScript library for mobile-friendly interactive maps; Leaflet Routing Machine[5], an easy, flexible and extensible plugin that adds routing to a Leaflet map; JASSA libraries and its modules [17] for querying the underlying data layer; FACETE for faceted filtering, Mappify for authoring.

In *Phase III* the application was tested in laboratory environment, see https://www.youtube.com/watch?v=PTOeY0bc6dE and Figure 4.



a) Semantic filtering; *b) Semantic routing;* *c) Motive-based search*

Fig. 4. GEM used in different scenarios.

6 Conclusions and Lessons Learned

In December 2011, The European Commission (EU) launched an Open Data Strategy for Europe, which is expected to deliver a €40 billion boost to the EU's economy each year[6]. Later in July 2013, the Commission put into force the Directive on the re-use of Public Sector Information (known as the 'PSI Directive', 2013/37/EU). This Directive revised the Directive 2003/98/EC and provides a common legal framework for a European market for government-held data (public sector information). It focuses on the economic aspects of re-use of information rather than on the accessibility of information to citizens. As enabling technology for the re-use of public data, the Commission has recognized the Linked Data paradigm. However, the economic effects of using Open Data for commercial purposes is not easy measurable. Open Data, even when freely available, is not free to use since so much time has to be spent cleaning it up, converting it, integrating and maintaining it. There is a marked difference in approach between government and business. One provides long term investment and

[4] http://leafletjs.com
[5] http://www.liedman.net/leaflet-routing-machine/
[6] http://europa.eu/rapid/press-release_IP-11-1524_en.htm?locale=en

slow innovation, but the opposite is true for business, especially activists. Hence, many startups have appeared that foresee the opportunities of the Linked Data approach, Open Data and Big Data strategies currently supported by the European Commission [18].

In the last few years, participating in few EU projects, the PUPIN team had an opportunity to validate the Linked Data approach with Open Data from Serbia and abroad (i.e. Germany). We contributed to the development of the Linked Data Stack and developed several open-source tools that can be further re-used for building innovative applications in government and business sector (tourism, supply-chain management), see http://linkeddata.rs/Products.

Future work will include further development of the developed resources (Serbian code lists, prototype applications and interfaces to government institutions) and guidelines summarized in the Open Data Handbook, see http://linkeddata.rs/OpenDataHandbook.

Acknowledgments. The research presented in this paper is partly financed by the European Union (FP7 GeoKnow, Pr. No: 318159; CIP SHARE-PSI 2.0, Pr. No: 621012), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. In Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 5(3), Pages 1-22. DOI: 10.4018/jswis.2009081901, http://linkeddata.org/docs/ijswis-special-issue (2009).
2. Auer, S., Lehmann, J.: Making the web a data washing machine - creating knowledge out of interlinked data. Semantic Web Journal. vol. 1. no. 12. pp. 97-104. IOS Press (2010).
3. Auer, S., Bryl, V., Tramp, S. (eds): Linked Open Data -- Creating Knowledge Out of Interlinked Data. Lecture Notes in Computer Science vol. 8661. Springer International Publishing. ISBN: 978-3-319-09845-6 (Print) 978-3-319-09846-3 (Online) (2014).
4. Berners-Lee, T.: Design Issues: Linked Data (2006), http://www.w3.org/DesignIssues/LinkedData.html, last accessed 2017/07/31.
5. Van Nuffelen et all.: Supporting the Linked Data Life Cycle Using an Integrated Tool Stack. In S. Auer, V. Bryl, S. Tramp (eds) Linked Open Data -- Creating Knowledge Out of Interlinked Data. Lecture Notes in Computer Science vol. 8661, pp. 108–129. Springer International Publishing. ISBN: 978-3-319-09846-3 (2014).
6. Pop, D.et all (Ed.): Deliverable 7.2 Stable Version of the Share-PSI 2.0 Best Practices (https://www.w3.org/2013/share-psi/bp/) (2016).
7. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF (2003).

8. Lehmann, J.: The GeoKnow Handbook, GeoKnow Project, https://svn.aksw.org/projects/GeoKnow/Public/GeoKnow-Handbook.pdf, last accessed 2017/07/31/.
9. Villazón-Terrazas, B., Vilches-Blázquez, L.M., Corcho, O., Gómez-Pérez, A.: Methodological Guidelines for Publishing Government Linked Data. In: Wood, D. (ed.) Linking Government Data, ch. 2. Springer (2011).
10. Jovanovic, M.: Linked Data Application Development Methodology, PhD Thesis. Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Macedonia, November 2016.
11. Janev, V., Mijović, V., Vraneš, S.: Proposal for Implementing the EU PSI Directive in Serbia, In Kő, A., and Francesconi, E. (eds), Proceedings of the 5th International Conference on Electronic Government and the Information Systems Perspective (EGOVIS 2016, September 5 - 8, 2016, in Porto, Portugal). Lecture Notes in Computer Science, pp 16-30. Springer International Publishing. http://link.springer.com/chapter/10.1007%2F978-3-319-44159-7_2 (2016).
12. Mijović, V. Janev, V., Paunović, D., Vraneš, S.: Exploratory Spatio-Temporal Analysis of Linked Statistical Data, Journal of Web Semantics, Web Semantics: Science, Services and Agents on the World Wide Web 41C (2016) pp. 1-8 DOI information: 10.1016/j.websem.2016.10.002 (2016).
13. Milošević, U., Stadler, C.: Mobile Semantic Geospatial Visualization and Exploration, In Zdravković, M., Trajanović, M., Konjović, Z. (eds) Proceedings of the 5th International Conference on Information Society Technology (March 8-March 11, 2015, Kopaonik, Serbia), Society for Information Systems and Computer Networks (2015).
14. Hellmann, S. et all.: Knowledge Base Creation, Enrichment and Repair, in Linked Open Data -- Creating Knowledge Out of Interlinked Data, Volume 8661 of the series Lecture Notes in Computer Science, pp 45-69 (2014).
15. Mijović, V., Janev, V., Paunović, D., Vraneš, S.: Deliverable 5.6.1 Release of the Mobile Supply Chain Consolidated View and Application (2015).
16. Isele, R., Pietzsch, R.: Deliverable 5.3.1 GeoKnow Generator Release for Data Web Background Knowledge Provisioning (2015).
17. Stadler, C., Jassa: JAvascript Suite for Sparql Access, http://aksw.org/Projects/Jassa.html, https://github.com/GeoKnow/Jassa-UI-Angular/tree/master/jassa-ui-angular-core, last accessed 2017/07/31/.
18. Loutas, N.: Business models for Linked Government Data: what lies beneath? European Data Forum 2014, https://2014.data-forum.eu/person/nikolaos-loutas-0.html, last accessed 2017/07/31/.