LinkedPipes Visualization: Simple Useful Linked Data Visualization Use Cases

Jakub Klímek(✉), Jiří Helmich, and Martin Nečaský

Faculty of Mathematics and Physics, Charles University in Prague,
Malostranské nám. 25, 118 00 Praha 1, Czech Republic
{klimek,helmich,necasky}@ksi.mff.cuni.cz

Abstract. There is a need for being able to effectively demonstrate the benefits of publishing Linked Data. There are already many datasets and they are no longer limited to research based data sources. Governments and even companies start publishing Linked Data as well. However, a tool, which would be able to immediately demonstrate the Linked Data benefits to those, who still need convincing, was missing. In this paper, we demonstrate LinkedPipes Visualization, a tool based on our previous work, the Linked Data Visualization Model. Using this tool, we show four simple use cases that immediately demonstrate the Linked Data benefits. We demonstrate the value of providing dereferenceable IRIs and using vocabularies standardized as W3C Recommendations on use cases based on SKOS and the RDF Data Cube Vocabulary, providing data visualizations on one click. LinkedPipes Visualization can be extended to support other vocabularies through additional visualization components.

Keywords: Linked Data · RDF · Visualization · Discovery

1 Introduction

The research in the area of Linked Data shifts from Linked Data preparation towards Linked Data consumption and visualization. There are many tools that allow the users to somehow visualize RDF data, however, the process tends to be more complicated than necessary. The potential of simplification is thanks to the Linked Data principles and standardized vocabularies (W3C Recommendations) such as SKOS[1], the RDF Data Cube Vocabulary[2] and others. One of the promises used to convince data publishers to use Linked Data is, that by using these standardized vocabularies, users and Linked Data enabled applications around the world will easily understand the published data and will be able to reuse it. An easy way of showing the publishers the added value of Linked Data would be to show them their own data used in those various applications

This work was supported in part by the Czech Science Foundation (GAČR), grant number 16-09713 and in part by the project SVV-2016-260331.

[1] https://www.w3.org/TR/skos-reference/.
[2] https://www.w3.org/TR/vocab-data-cube/.

H. Sack et al. (Eds.): ESWC 2016 Satellite Events, LNCS 9989, pp. 112–117, 2016.
DOI: 10.1007/978-3-319-47602-5_23

immediately after publishing it. This is because many data publishers still feel that publishing data is not enough, that a user also needs a visualization and that they are the ones best qualified to develop it. Unfortunately, this is the moment when the Linked Data arguments stop being convincing and shrink to a simple statement of the fact that once the data is published as Linked Data, others can link to it. The cause of this problem is the lack of simple and ready to use applications that could immediately show the benefits of publishing data as Linked Data.

In this paper, we introduce LinkedPipes Visualization (LP-VIZ), a Linked Data visualization tool built on top of our Linked Data Visualization Model (LDVM) [6], which we developed and experimented with earlier. This time, we focus on the demonstration of benefits of usage of standardized vocabularies and the ability to dereference an IRI, to access it to get additional data.

The rest of the paper is structured as follows. In Sect. 2 we survey the related work. In Sect. 3 we briefly mention our Linked Data Visualization Model (LDVM) as the theoretical basis and we introduce its implementation, Linked-Pipes Visualization, which we demonstrate on use cases in Sect. 4. In Sect. 5 we state our future efforts and we conclude.

2 Related Work

There are many RDF visualization approaches, but most of them are not easy to use and require the user to provide unnecessary information about the data that could be automatically discovered from it instead. For their overview, see [6] and [5]. With LinkedPipes Visualization, we focus on providing simple visualizations with minimal user effort and we try to get as much information from the data itself as possible. From this point of view, the most related is LodLive [3], which enables users to quickly browse their Linked Data provided that their IRIs are dereferencable. It also parses labels of properties and classes based on the RDF Schema (RDFS)[3] vocabulary. However, LodLive does not provide vocabulary specific visualizations and even though it provides a map visualization, it does not recognize the Schema.org GeoCoordinates class. Another tool in early prototype stage heading in the direction of analyzing data based on Linked Data vocabularies and principles is LDVizWiz [1], which detects categories of data in a SPARQL endpoint.

3 Linked Data Visualization Model and LinkedPipes Visualization

In our previous work, we defined the Linked Data Visualization Model (LDVM) [2]. It consists of components of four types. The *data sources* represent the source RDF data, the *analyzers* extract data from data sources, the *transformers* manipulate the data, e.g. transform it between formats according

[3] https://www.w3.org/TR/rdf-schema/.

to different vocabularies, and, finally, the *visualizers* provide the visualization itself. A key concept of LDVM is the *component compatibility checking*, which allowed us to automatically compose a visualization pipeline out of the components registered in a LDVM instance, based on their capabilities described by RDF data samples, SPARQL input descriptors and the source data.

Based on LDVM, we showed advanced use cases [6] where we were able to automatically compose visualization pipelines combining multiple data sources. In [4] we show how RDF data cubes can be visualized using LDVM and our Data cube visualizer.

LinkedPipes Visualization[4] is our current LDVM implementation. Besides being able to support the advanced use cases from [6], it now supports much simpler and straightforward use cases, on which we can more directly demonstrate the benefits of Linked Data and which we cover in this paper.

4 Demonstrated Use Cases

In this section, we demonstrate the usability of LinkedPipes Visualization on four use cases. We focus mainly on simple demonstrations of Linked Data benefits regarding use of vocabularies and providing dereferencable IRIs.

4.1 Validation

An important feature to have in a data consumption tool is *validation*. This means that instead of seeing an error or no result when trying to visualize data, the user has an option to have his data validated and see what is wrong. So far, we have two validators to choose from, SKOS for taxonomies and Data Cube for statistical data cubes. There already are validators for SKOS[5] and RDF Data Cubes[6]. However, we focus on validation of specific common errors that break our visualizations.

Because we assume that the data is wrong, first we need to know, what the data is expected to be. Therefore, in this use case, we start with selecting the validator we want to use. For the demonstration, we will use the simpler SKOS validator, which checks mainly for missing SKOS concept schemes, links from concepts to the concept schemes via the `skos:inScheme` property and missing target concepts of `skos:broader` and `skos:narrower` links and their transitive variants.

First, the user inputs a broken SKOS concept scheme[7]. Second, the validator detects, that the object of one of the `skos:broader`, one of the `skos:narrower` and one of the `skos:inScheme` properties is missing in the data. This could be caused by omitting an entity or a simple misspell during manual creation of the concept scheme.

4 http://visualization.linkedpipes.com.
5 https://www.w3.org/2001/sw/wiki/SKOS/Validation.
6 https://github.com/yyz1989/NoSPA-RDF-Data-Cube-Validator.
7 http://visualization.linkedpipes.com/example/broken-skos.ttl.

4.2 One Click Visualization

One of the promised advantages of using Linked Data is that when the data is modeled according to standardized vocabularies, it can be easily used in tools that support such vocabularies. When applied to visualizations, the user expects that when he models his data correctly, he gets a visualization specific for used vocabularies, easily. From the workflow perspective, this means that the user has to only provide the right data to get a vocabulary-specific visualization, nothing more.

In this use case we demonstrate, how LinkedPipes Visualization can achieve just that, a *one click user data visualization*. This is given that the data is a SKOS concept scheme, a Data Cube or a set of points on a map according to Schema.org[8], for which we currently have visualization components installed in our demo instance[9]. More vocabularies (or their combinations) can be supported by adding additional visualization components with corresponding input descriptors.

First, the user inputs a URL of an RDF file with a SKOS Concept scheme. This can be e.g. some of the newly published Named Authority Lists (NALs)[10] in the European Metadata Registry, such as the EU programmes[11] and clicks on *Visualize*. Second, the concept scheme is detected and visualized using our library of hierarchy visualizers from the D3.js[12] library. For the EU programmes, the Tree visualization is the most appropriate, even though it is really a code list, not a hierarchy.

4.3 Enhancing Visualization with Dereferencing of IRIs

Another one of the key principles of Linked Data states that the data publisher should link to other data to provide context. This is almost mandatory in the case of publishing data cubes, as they consist of observations identified by dimension values and those values typically come from a code list represented as a SKOS concept scheme. This concept scheme is usually shared by dimensions of multiple data cubes, and possibly even other types of datasets, and therefore, it is almost necessarily a different dataset. This is especially true for highly reused concepts such as the ones for sex[13], time periods[14] or geographical locations. Therefore, when the user wants to visualize a data cube, which is using such dimensions, all the values for those dimensions will almost certainly not be available in the same file or even the same SPARQL endpoint as the actual data cube data. Instead, they will be simple links to those other datasets. For conventional data cube

[8] http://schema.org.
[9] http://demo.visualization.linkedpipes.com.
[10] http://publications.europa.eu/mdr/authority/.
[11] http://publications.europa.eu/mdr/resource/authority/eu-programme/skos/eu-programme-skos.rdf.
[12] https://d3js.org/.
[13] http://purl.org/linked-data/sdmx/2009/code#sex.
[14] https://datahub.io/dataset/data-gov-uk-time-intervals.

visualization methods, this usually means that the dimension values will not be displayed correctly, because the human readable labels are not available in the same data source as the observations, and only their IRIs will be visible.

In this use case we demonstrate, how our Data Cube visualizer can handle this situation naturally by *dereferencing the IRIs* of dimension values, for which there are no labels in the original data source and which can be dereferenced. First, the user inputs the data cube[15] and its Data Structure Definition (DSD)[16]. LP-VIZ detects that there in fact is a data cube in the data and displays it. Next, for dimension values, labels are displayed even though they are not in the source data, because they are obtained through dereferencing of their IRIs.

4.4 Embedding Visualizations

In this short use case, we will create a visualization in LinkedPipes Visualization and show how it can be easily embedded into another website as an HTML *iframe*. This can help data publishers to easily develop simple visualizations with little extra cost, again thanks to the usage of Linked Data. We will start with the visualization from Subsect. 4.3 and pre–select some dimension values. Next, we will use the *embed link* to add this visualization as a resource to a CKAN data catalog.

5 Future Work and Conclusions

Currently, we are working on extending our library of visualizers with ones allowing us to demonstrate the Linked Data benefits on a wider range of well–known vocabularies.

In this paper, we demonstrated LinkedPipes Visualization (LP-VIZ), a tool implementing our Linked Data Visualization Model (LDVM). On four use cases we showed how the Linked Data benefits promised to data publishers can be immediately exploited by validating the data, producing vocabulary specific visualizations on one click, enhancing them with dereferenceable IRIs and embedding them in websites.

References

1. Atemezing, G.A., Troncy, R.: Towards a Linked-Data based visualization wizard. In: Hartig, O., Hogan, A., Sequeda, J. (eds.) Proceedings of the 5th International Workshop on Consuming Linked Data (COLD 2014) Co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, 20 October 2014, vol. 1264. CEUR Workshop Proceedings, CEUR-WS.org (2014)
2. Brunetti, J.M., Auer, S., García, R., Klímek, J., Nečaský, M.: Formal Linked Data visualization model. In: Proceedings of the 15th International Conference on Information Integration and Web-based Applications & Services (IIWAS 2013), pp. 309–318 (2013)

[15] http://visualization.linkedpipes.com/example/datacube.ttl.
[16] http://visualization.linkedpipes.com/example/dsd.ttl.

3. Camarda, D.V., Mazzini, S., Antonuccio, A.: LodLive, exploring the web of data. In: Presutti, V., Pinto, H.S. (eds.) 8th International Conference on Semantic Systems, I-SEMANTICS 2012, Graz, Austria, 5–7 September 2012, pp. 197–200. ACM (2012)
4. Helmich, J., Klímek, J., Nečaský, M.: Visualizing RDF data cubes using the Linked Data visualization model. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8798, pp. 368–373. Springer, Heidelberg (2014). doi:10.1007/978-3-319-11955-7_50
5. Klímek, J., Helmich, J., Nečaský, M.: Application of the Linked Data visualization model on real world data from the Czech LOD cloud. In: Bizer, C., Heath, T., Auer, S., Berners-Lee, T. (eds.) Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014), Seoul, Korea, 8 April 2014, vol. 1184. CEUR Workshop Proceedings. CEUR-WS.org (2014)
6. Klímek, J., Helmich, J., Nečaský, M.: Use cases for Linked Data visualization model. In: Bizer, C., Auer, S., Berners-Lee, T., Heath, T. (eds.) Proceedings of the Workshop on Linked Data on the Web, LDOW 2015, Co-located with the 24th International World Wide Web Conference (WWW 2015), Florence, Italy, 19 May 2015, vol. 1409. CEUR Workshop Proceedings, CEUR-WS.org (2015)