LINKDALE: A LIGHTWEIGHT LEARNING ENVIRONMENT FOR (GEOSPATIAL) LINKED DATA

S. Ronzhin [a] [*], G. Bosch [b], E. Folmer [c], R. Lemmens [a]

[a] University of Twente, Faculty of Geo-Information and Earth Observation (ITC), PO BOX 217 7500 AE Enschede, The Netherlands – (s.ronzhin, r.l.g.lemmens)@utwente.nl
[b] Trypli, The Netherlands, gerwinbosch@chello.nl
[c] University of Twente, Faculty of Behavioural Management and Social Sciences (BMS), PO 217 7500 AE Enschede, The Netherlands – Erwin.folmer@utwente.nl

Commission IV, WG IV/4

KEY WORDS: Linked Data, education, learning environment, reactjs, visual linking,

ABSTRACT:

Modern software tools for managing Linked Data are often designed for skilled users. Therefore, they cannot be used for education purposes because they require substantial a priori knowledge about the Resource Description Framework and the SPARQL query language. LinkDaLe is a single page application designed to teach students the concept of Linked Data and work with linked data at the same time. In the paper we showcase the interface and functionality of LinkDaLe by triplifying data on Geo4All member organizations. The application was built and evaluated within The Business Process Integration Lab, a master programme course in 2016 and 2017 years. Positive feedback from both students and teachers proved the relevance of the proposed design consideration. LinkDaLe showed usability working with domain specific data e.g. geospatial and logistic data.

1. INTRODUCTION

Linked Data(LD) has a steep learning curve. On the one hand this stems from the need to master a wide range of diverse skills and topics from the domains of knowledge representation, information management and retrieval. To understand the mere LD design rules formulated by Berners-Lee (Berners-Lee, 2006), a student must know the Hypertext Transfer Protocol (HTTP), the Resource Description Framework (RDF) and finally, SPARQL, a query language of the Semantic Web.

On the other hand, despite of a great number of tools developed for LD so far, they are not meant for lay users or beginners. Available tools are built with a skilled user in mind, not a novice. Those two factors create a chicken and egg problem – learning LD requires tools, and tools require understanding of LD. The less the IT background of a student the more significant this problem is. In the case of Geospatial Linked Data, the picture is even more complicated by introducing concepts from the GIS world like coordinates, projections and geometries.

The Faculty of Behavioral Management and Social sciences (BMS) of the University of Twente (UT) (1) together with the Dutch Cadaster (Kadaster) (1) developed LinkDaLe (1) (Linked Data learning environment) (1), a learning environment to facilitate Linked Data assignments. The objective was to design, develop and evaluate a lightweight one-page application that would provide functionality to support practicals and workshops on Linked Data and Geospatial Linked Data in different stages of

its lifecycle (Ngomo et al., 2014) for a non-technical or mixed audience.

The following section gives details on the educational environment where LinkDaLe was used. After that, Section 3 introduces a set of design considerations taken for the development of the application. Section 4 presents the interface and functionality of the application and explains how the design consideration was reflected by the implementation. In Section 5 we present and discuss results of evaluation followed by Section 6 where conclusions are drawn.

2. BUSINESS PROCESS INTEGRATION LAB

The Business Process Integration Lab (BPIL) (1) is a course for master students at BMS UT aimed at understanding key concepts, methods and tools for integration of business processes from both an organizational and technological point of view.

The Linked Open Data approach to integration is one of the central topics of this course. During the assignments students model and generate their Linked Data and used it together with the data from official governmental registres mainteined by Kadaster. This data has a strong spatial component that was used for linking the data. The students had mixed educational background – business administration and business information technology. They lacked any knowladge about GIS concepts like projection and coordinates.

---

[*] Corresponding author
https://www.utwente.nl/en/bms/
https://www.kadaster.nl/.
http://linkdale.org
https://github.com/PDOK/LinkDaLe
https://goo.gl/9pmn3C
https://www.getpostman.com/

In the past, students used multiple tools such as OpenRefine with the RDF extension (Verlic, 2012), LinDA (Thellmann et al., 2014) and Postman (1) to perform the assignments. The tools provided extensive functionality supporting different stages of the Linked Data lifecycle. One shortcoming was that tools required installation and setting up in class which was time-consuming. Even though all the assignment had step-by-step instructions on how to operate the software, many students experiences difficulties especially in the beginning of the course.

3. DESIGN CONSIDERATIONS

Here we present seven design considerations formulated based on the experience of running BPIL course. They are as follows:

1. For students, LD usually begins with four canonical design rules (Berners-Lee, 2006). For this reason, the interface should be designed in such a way that a student clearly sees how implementation of those rules influences their data.
2. Linking between data items should be performed by interacting with a network visualization. This promotes data modeling skills and does not require knowledge of the RDF syntax.
3. Students are forced to follow best practices of Linked Data. For example, *rdf:type* and *rdfs:label* should be obligatory properties for every data instance. This allows students generating LD in a quick and dirty manner with enough semantics to fix the problems with data later if needed.
4. The application follows the Linked Data Visualization Model (Brunetti et al., 2013) by providing an appropriate visualization for specific datatypes – a map interface for geo features, a table view for tabular data and a network visualization to show class and entity hierarchies.
5. Users should be assisted with providing input such as Uniform Resource Identifiers (URIs) classes and properties to avoid syntax and grammar mistakes.
6. Main instructions should be embedded into the pages. Therefore, students can read about the interface while performing exercises.
7. Teachers need to publish and maintain assignments and tutorial scripts with ease in a user-friendly way.

4. INTERFACE AND FUNCTIONALITY

LinkDaLe is a lightweight one-page application built with the React framework (1) and served via GitHub pages (1). The interface is built with the Google's material design UI (1) components to achieve a recognizable and familiar look and feel of the application.

From a landing page (Figure 1), a user can go into one of the four sections of the application. Under the "Create Linked Data" section, users can upload their data in the Comma-Separate Values (CSV) format, generate Linked Data from it and publish it in a triple store. "Browse data" allows browsing through datasets published via the tool. "Query Data" provides a SPARQL interface to facilitate federated querying. Section "Tutorial" is self-explanatory and contains assignments and tutorials.
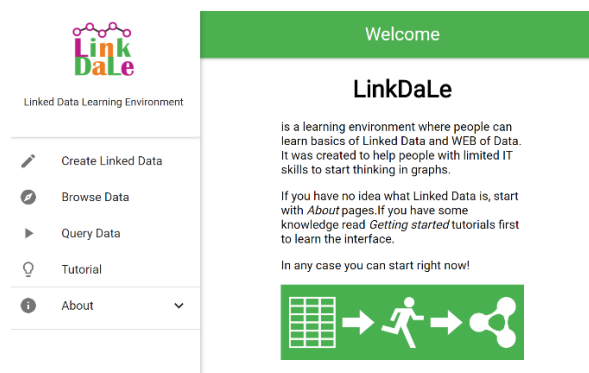

Figure 1. Landing page of LinkDaLe.

4.1 From tables to networks

The Create Linked Data section of the application helps users to make an LD representation of a simple table using ontologies. A stepper element guides through this process in four steps: upload, classification, linking, publishing. At the last step users can either download the results or to publish them into a triple store.

The data conversion process introduces the rules of linked data consequently starting with the first two rules. This is done in line with the first design consideration. The rules are as follows:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names

In practice for those who study link data these rules can be translated into two questions to be answered:

1. What things should be named?
2. What URI strategy should be used?

These two questions require understanding of the subject-predicate-object model of RDF and URI strategies. LinkDaLe helps with the latter by providing proper base URIs and dereferencing functionality. Therefore, users can focus on the first question and learn basics of RDF by analysing what data items can be used as subjects and therefore deserve URIs.

As a running example, let us create a linked data representation of the information about Geo4All member organizations (OSGeo, 2018). The source data features information about 125 organizations. Table 1 presents a structure of the data and an example record.

Field name	Example record
Laboratory name and institution	University of Nottingham
Country	UK
Lat	52.831497
Long	-1.250296
Contact names	Stuart Marsh
Contact emails	Stuart.Marsh@nottingham.ac.uk

Table 1. Example data for triplification describing a name and location of organization (fields: Laboratory name and

---

https://reactjs.org/
https://pages.github.com/

https://material.io

institution, Country, Lat, Long), as well as names and contact of contact person (fields: Contact names, Contact emails)

In Table 1, the data have 6 fields with information about a name of organization, its location and a name and contacts of a contact person.

These data were uploaded into LinkDaLe as a CSV table. Figure 2 provides a view of the classification screen where users can see the structure of the uploaded data. If a column from a source data contains things that can be named, then a user puts a tick in the related checkbox "Is it a URI?".

In Figure 2, three field are checked - Laboratory name and institution, Country and Contact names. These fields containe unique values thereofore it is allowed to use them for generating URIs.

Once a source field for URIs is identified users are promt to apply the third rule of LD wich is as follows:

> 3. When someone looks up a URI, provide useful information, using the standards (RDF and SPARQL)

In practice for novices this can be read as:

> 3. Provide types and labels for things, so people can understand your data.

Figure 2. View of the classification step where users are asked to define what data item deserve URIs and what can be expressed as literals.

LinkDaLe assists users in searching relevant classes using the Link Open Vocabularies (LOV) (1) service (Vandenbussche et al., 2017). Figure 3 provides an example search for classes that have "spatialthing" as part of the class name. *rdfs:label* is inferred for every URI using values from the original data.

4.2 Visual linking

Once classification is done a user is prompt to the Link Data view. In this view user visually connects items identified in the classification screen by interacting with a network visualization. Figure 4 provides an example of such visualization with circles representing classes, rectangles literal values and arrows showing relation between items.

As can be seen from Figure 4(A) there are three links which were inferred by software – *rdfs:label*. Figure 4(B) shows the final version of the network. User has manually connected relevant

http://lov.okfn.org/

items and provided proper relations between them. Search for relations is implemented in similar way as the select class dialogue using LOV search.
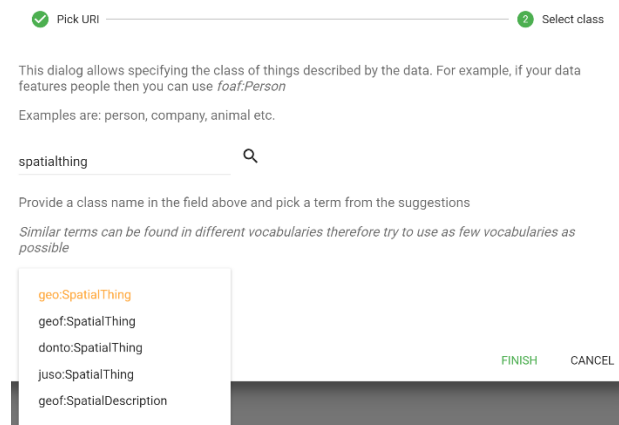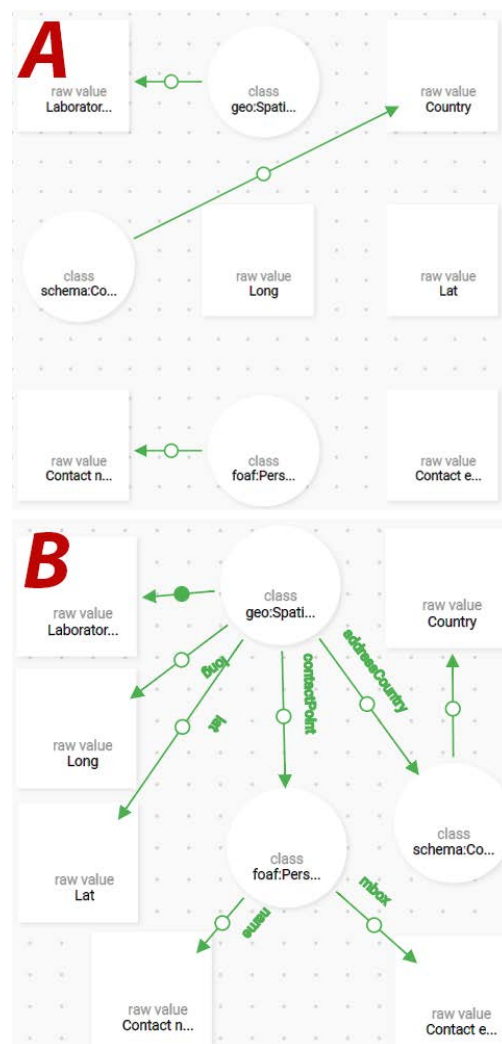

Figure 3. The Select Class dialog uses the Link Open Vocabularies service to search for relevant classes

Figure 4. Network visualisation of data after classification (A) and (B) with all identified relations

At the latest step of the Create Linked Data process users fill a small metadata form submitting the name and description of the dataset. LinkDaLe inserts all the datasets as named graphs into a remote triple store and can be accessed via a public SPARQL endpoint (http://almere.pilod.nl/sparql).

4.3 Browsing and querying results

The Browse Date section gives access to the created data. As can be seen from Figure 5, users can select a dataset from the list (upper part of the screen) and explore it using different views on the data: tabular, data graph and class graph view.

Users can either query the data in the query interface of LinkDaLe or using any other query interface (e.g. YASGUI) connected to the public SPARQL endpoint.
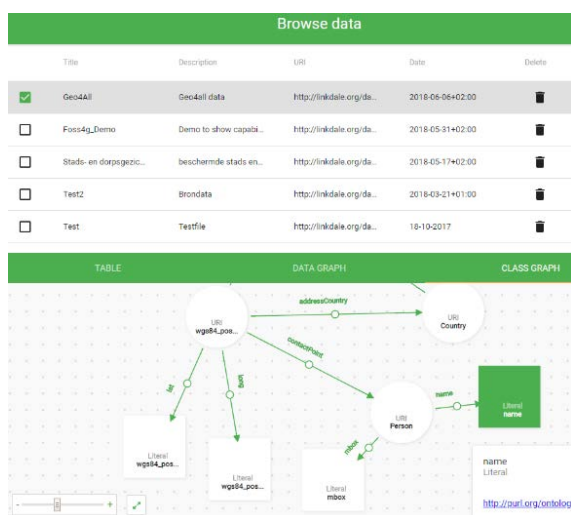

Figure 5. The view of the Browse Data screen

5. EVALUATION

LinkDaLe was evaluated within the Business Process Integration Lab (BPIL) in 2016 and 2017 years. The course included 3 practicals on LD. In both years students were asked to perform the same assignments where they created Linked Data from their sources, enriched it with data from Kadaster, and query it together with other resources.

In 2016, students chained multiple tools to perform the assignments and in 2017, they used only LinkDaLe. The main difference between these years was in the time needed for performing an assignment. With LinkDaLe, the assignments took only half of the allocated time. Students were two time faster with LinkDaLe than with the chain of tools used in 2016.

Another difference was in the time needed to assess the performed assignments. In 2016, a teacher had to collect results from students using email or blackboard environment. With LinkDaLe it was possible to set up SPARQL queries to evaluate quality of student works. This shortens the time needed for assessment with a factor 10. In general, LinkDaLe was appreciated by both students and staff.

6. CONCLUSION AND FUTURE WORK

LinkDaLe is a one-page application where students can learn LD principles and create their own data at the same time. By providing users with proper URIs and search functionality for classes and relations the software decreases the need for a priori knowledge required to start working with LD. In addition, interactive network visualization fosters modelling skills and does not require knowing syntax. All of these allows novices creating divers LD e.g. as was shown by creating LD description of the Geo4All member organisations.

Since all the data is available via a SPARQL endpoint, teachers can setup queries to automate evaluation of assignments. This significantly decreased time teachers spent on assessment of student performance.

For the next academic year, the tool will be improved with a map interface in the "Browse data" section. The map will depict each class of features in a dataset as a separate map layer to allow mash upping them on a map. The usability of the application will be further researched.

7. REFERENCES

Berners-Lee, T., 2006. Linked Data - Design Issues. Retrieved October 1, 2014 from www.w3.org/DesignIssues/LinkedData.html

Brunetti, J. M., Auer, S., García, R., Klímek, J., & Nečaský, M., 2013. Formal linked data visualization model. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services* (p. 309). ACM.

Ngomo, A. C. N., Auer, S., Lehmann, J., & Zaveri, A., 2014. Introduction to linked data and its lifecycle on the web. In *Reasoning Web International Summer School* (pp. 1-99). Springer, Cham.

OSGeo, 2018. Edu current initiatives, Current members of the Geo for All Labs Network. Open Source Geospatial Foundation https://wiki.osgeo.org/wiki/Edu_current_initiatives (1 June 2018)

Thellmann, K., Orlandi, F. and Auer, S., 2014. LinDA-visualising and exploring linked data. In *Proceedings of the Posters and Demos Track of 10th International Conference on Semantic Systems-SEMANTiCS2014, Leipzig, Germany*(Vol. 9).

Vandenbussche, P.Y., Atemezing, G.A., Poveda-Villalón, M. and Vatant, B., 2017. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. *Semantic Web*, *8*(3), pp.437-452.

Verlic, M., 2012. LODGrefine-LOD-enabled Google Refine in Action. In *I-SEMANTICS (Posters & Demos)* (pp. 31-37).