Linked Open Data Literacy for Librarians

Jasmin Hügi and René Schneider

Haute école de gestion de Genève, HES-SO, University of Applied Sciences Western Switzerland, rte de Drize 7, 1227 Carouge, Switzerland
{jasmin.hugi,rene.schneider}@hesge.ch

Abstract. Linked Data has become an important issue, not only for the process of building the web of data, but also for the mutual knowledge transfer between libraries and the web of data. Based on a study concerning Linked Open Data applications in libraries and the qualifications of future librarians involved in the development of such applications, we built a one day training program for academic librarians. The overall goal was to make librarians literate and heighten awareness concerning the Linked Open Data technology on a single day. With the help of that training program, the librarians should become cognitively able to reflect on the integration of Linked Data in their current working environment. We considered it therefore necessary to couple the training program with the development of cognitive strategies. In this paper we will give an in-depth description of the didactical approach, the theoretical and practical components as well as their possible combinations.

Keywords: Linked open data, data literacy, training, cognitive strategies.

1 Introduction

Since the rise of the World Wide Web, libraries have always seen themselves in need of responding adequately to the change, be it in terms of benefits, or be it in terms of challenges that the web has brought with it. They tried to connect the attitudes and technologies that were part of this change to the services that they offered, namely the supply and mediation of information. The aspect of mediation has been strengthened by the special significance that was assigned to information literacy and the role that libraries could play in this context; this aspect of information has been enhanced by the new interest in data, especially research data or in the case of this paper Linked Open Data (hereafter LOD). At the present time, quite a large number of catalogues or data bases are transformed in new interoperable data sheets based upon the Resource Description Framework (hereafter RDF). In addition to this, further library applications emerge that enrich the panoply of services offered by libraries. Thus, every library has to consider whether it shall adapt this new technology and participate in building the web of data.

As a consequence, the librarians have to ask themselves how they should become literate with this rather complex matter since it will have an impact on their future work considerably. In this context, we developed a training for Linked Data Literacy, initiated by a mandate from the Cantonal and University Library of Lausanne

S. Kurbanoğlu et al. (Eds.): ECIL 2014, CCIS 492, pp. 109–117, 2014.

(Bibliothèque cantonale et universitaire de Lausanne, hereafter BCU Lausanne) that besides being an academic library, has - as the name indicates - the role of a cantonal library for the canton of Vaud, being the third largest Swiss canton by population. Things started informally when the library directorship asked for the possibility of a one day training on LOD in libraries that would serve to all different types of librarians to raise awareness to the matter, with no further constraints or requirements given. Shortly after this, the library of the Swiss Federal Institute of Technology in Lausanne (in French: Ecole polytechnique fédérale de Lausanne, hereafter EPFL) wished to have a similar training by adding the constraints of integrating the evolution of library standards into the workshop. Finally, a third workshop was given, once more at the BCU Lausanne. During these workshops the authors gained rich experiences and precious insights concerning the teaching of Linked Open Data to librarians that will be described in the next section of this article.

In this context, our goal was not only to give an introduction to the main aspect of LOD, namely how to link data that was free of copyrights into the world of libraries [1], but also to strengthen the librarian's capacity to respond actively to the challenges mentioned above, (i.e. the development of cognitive strategies) [2]. They should thus acquire a maximal autonomy during a one day training by a) studying trustworthy sources about LOD, b) the identification of relevant websites about standards and formats, and last but not least c) the decision making process and organization of an LOD project. So beside the constraint of becoming literate to the topic of linked open data in a minimum of time and with a maximum of practical exercises. It was important for us that the participants gained the capacity to develop autonomously their own strategies and to be able to undertake first steps for the implementation of LOD-projects of any size within the institution in which they work.

2 Construction of the Training Session

Before starting with the development of the presentation content, we defined learning objectives in order to have a frame to orient our efforts. Based on these objectives we identified the best methods and structure of the training workshop with which we could best achieve the set goals. This methodology is explained in further detail in the next two paragraphs.

2.1 Learning Objectives

In order to construct the training, we defined the overall goal of the training day based on the informal directive given by the client. We considered it wise to balance between essential factual knowledge on the one hand and procedural knowledge on the other hand, with the factual knowledge more prevalent in the first half of the training and the procedural knowledge located in the center of the second half. Besides that, we determined that the training should start with a very general and invitational introduction in order to reduce objections und resistances, it should furthermore end in an open and reassuring atmosphere with all the participants having a clear idea about the next steps to be done in order to realize own projects, no matter their size and efforts.

Thanks to working in a university which trains future library and information specialists, we could derive which competencies would be present among the learners and which knowledge in regard to LOD were lacking. We concluded that the learners wouldn't be necessarily very strong in technical matters. As LOD is a very technical topic, we wanted to give a gentle introduction without alienating some of the participants. Taking into consideration the typical curriculum for the initial training librarians, their general technological knowledge and the experience gained in LOD studies, the following learning objectives were defined:

- being able to describe the concept of LOD;
- being able to name at least three examples of LOD applications in libraries;
- being able to cite at least four benefits of LOD for libraries;
- being able to find relevant information regarding vocabularies, ontologies, RDF and its serializations on the web.

2.2 Didactical Approach

In order to achieve these objectives, we chose the cognitivist approach to be most adequate for our undertaking, mainly due to two reasons. First, a purely behavioristic approach seemed to be inappropriate and a merely constructive approach seemed unreasonable. Most participants would not have any further experience which LOD so an introduction with the essential elements was necessary. We decided therefore to start with an instruction on LOD, RDF, and Turtle before opening towards more discussion and the creation of own ideas, after an intermediate sequence where the participants had the time to understand, practice and reflect upon what they had learned in the first section.

As mentioned earlier, special attention was given to the integration of cognitive strategies into the serialized training modules [2]. Since we do understand cognitivism as “mind in action”, we tried to break up the usual behavioral pattern of a ‘disciple’ sitting still and quiet behind a table as much as possible. We thus tried to create a participative and collaborative environment from the beginning and forced participants to move, by creating and mixing groups and thus motivating participants to leave their chairs and to speak, either by inviting them to present results to the rest of the class or by inspecting the group work of the others by moving around and discussing openly the different approaches they followed from group to group.

The cognitivist approach can also be justified by the connective and collaborative nature of the web that was strengthened and completed with the rise of the social web and the web of data; we hence tried to foster an approach that focuses on collaboration and connectedness in any possible way. Our intention was that, the participants should become as autonomous as possible within minimum time and undergo a change from receptive beginners to reflective newcomers and finally to self-governed beginners who understand the matter discussed, and able to develop their own ideas and to discuss the proposals of others.

Finally, we decided to have computers in class, which were needed a) to give access to the digital course material on the web besides the paper bound material handed out at the beginning of the course, b) to follow examples being accessible on the web, and c) to use either web based or desktop based tools like an RDF syntax

validator and protégé. This does not necessarily mean that the training could not take place without the use of computers. The experience of teaching three times the training has shown us that computers might also cause distraction and that interaction between the participants and team work is much more important.

Based on successful experiences gained through other training workshops, we developed two didactical constraints: firstly that all lessons were consequently taught as an alternation between theoretical and either practical exercises or recreation breaks, and secondly that every theoretical intervention should ideally last 20 minutes but never longer than 40 minutes. This was mainly done to avoid monotony, assure hands-on experience and reduce loss of attention that is likely to happen during intensive courses [3]. No time limits were given concerning the duration of practical exercises.

These decisions led to the creation of short to middle length teaching units for the mediation of Linked Data essentials and short to long practical exercises that were performed alone or in small groups of two to four participants. These modules were then arranged in a logical sequence according to our constraints. Though theoretical and practical modules could be different in numbers, the overall duration of all theoretical and practical parts should be similar; preferably with more time spent with a practical experience of the matter taught to give hands-on experience.

3 Development of Theoretical and Practical Training Modules

The content to be taught was granulated in different modules according to the beforehand defined learning objectives and the didactical constraints. The modules' content ranged from the evolution of the web and the consequences for libraries and cataloguing over introductory units to Linked Open Data, RDF & Turtle, Linked Data Vocabularies and their added values to the understanding of ontologies and the future of cataloguing. The practical exercises started with a) the simple analysis of documents in Turtle and RDF/XML and were followed by b) designing RDF graphs and c) describing and validating them in Turtle to d) code essential elements of an ontology. The training program ended with an exercise where all participants were asked to discuss and model the integration of the learned contents into future projects. All practical exercises were designed to give concrete examples of bibliographical data represented as Linked Data.

The actual program consists of nine theoretical and six practical modules that can be combined differently with respect to the needs of the targeted audience. The teaching units and exercises tried to cover all aspects that were considered important by experts working in the domain and being involved in the development of state-of-the-art projects. For a thorough understanding of the program we will present shortly all theoretical and practical parts:

3.1 Theoretical Modules

1a. **The Evolution of the Web and Its Impact on Libraries:** This module is considered to be a general introduction to the topic. It describes the development

from the principles of the World Wide Web (web 1.0) to the Social web (web 2.0) and finally the web of data (web 3.0) with projections of the future web (web 4.0). Every stage of the web was linked to the libraries and their development with respect to the different technological peculiarities.

1b. **Evolution of Library Standards:** This module is a different starting point for the training in comparison to module 1a and oriented toward the future of cataloguing and the specific contribution of Linked Open Data. It gives general knowledge about ISBD, AACR, MARC, FRBR, BIBFRAME and RDA. The major goal of this module is to make the participants familiar with the terms and acronyms of the domain of cataloguing as well as to demonstrate the evolution taking place in cataloguing standards which may well lead to a substitution of some well-known standards.

2. **Introduction to LOD:** This is the first of four modules concerning Linked Open Data, in which the principles and consequences of linkedness and openness of data as formulated by Tim Berners-Lee [4] are presented together with a first link to the realm of open bibliographical data [5]. At the end, a first informal description of triples is given by illustrating the evolution of Wikipedia to DBpedia including a demo of the "influenced / influenced by" links between authors using the application "relfinder"[1].

3. **Introduction to RDF:** This module gives an overview of the basic elements of RDF (Resource Description Framework) [6], namely the notions of triples, subject – predicate – object, their decentralized realizations in graphs using URIs and the HTTP protocol. It ends with an example that shows the transfer of a bibliographical relational data base to a database with RDF-representation.

4. **Introduction to Turtle:** This module contains a short review of the RDF principles and a concise set of 9 principal rules that are necessary to know in order to read and write Turtle documents. At the end, participants should be able to fully interpret documents written in the syntax Turtle.

5a. **Introduction to Ontologies:** After an historical introduction to the matter of ontologies, this module summarises the main concepts (instances, (sub-) classes, properties, domain & range) of ontologies and the necessary steps to create them. This part was mainly based on Noy and McGuiness [7] which is a very helpful and essential document that helps lay people to understand this complex topic.

5b. **Introduction to Vocabularies and Ontologies:** Similar to the former module, this alternative module kept the historical introduction and the main concepts of ontologies, but neglected the steps to create them in favor of an overview and short presentation of the main vocabularies that are used for the description of bibliographical data in a Linked Data format, mainly based on Klee [8].

6. **The Added Value of LOD for Libraries:** This module summarised the findings of a study [9] whose aim was to give an overview of the applications that so far

[1] `http://www.visualdataweb.org/relfinder/relfinder.php`

have been created to bring Linked Open Bibliographical Data closer to the user and to distinguish between the mere data layer and the necessary work to make them usable and re-usable in potential applications.

7. **The Future of Cataloguing:** This module presents the outcomes of a study [10] that focused on the skills that future librarians should have whenever they are involved in LOD projects. At the same time this module was used to show which of these skills were taught or mentioned during the training and to what degree these competencies were covered on that day.

3.2 Practical Modules

i. **Analysis of a Turtle Document:** In this first practical exercise one or two Turtle-Documents were given to the participants with the tasks of first reading them and then trying to detect some elements that had meaning for them. In one case this was a document describing a person with FOAF, in the other case a document describing a bibliographical record. This was done with the certitude that almost all participants had no prior knowledge of neither LOD nor RDF or Turtle and that most parts of the documents remained undecipherable for them. Nevertheless most of them were able to detect the literals and the general meaning of the document(s). At the end of the exercise that was done in small teams of neighboring participants and for which a maximum of 10-15 minutes were reserved in the schedule, the findings of all participants were collected in a plenary session.. The goal of this exercise is that even if the learners don't understand Turtle yet, they are able to recognize a bibliographical record. It serves as well as an example of the learning progress of each participant, toward being able to read the document by the end of the training day.

ii. **Comparison of RDF/XML and Turtle Documents:** In this second exercise two documents, one in RDF/XML and the other one in Turtle were handed out to the participants in order for the participants to compare them. Both documents described the same bibliographical record. Thus the participants were able to find out (still without any knowledge of RDF and Turtle) that a) in RDF there are different syntaxes to describe the same object or instance and b) that Turtle was much easier to understand. Similar to the prior exercise this exercise was done in neighboring groups with two to four members and the findings were discussed afterwards in plenary.

iii. **Designing an RDF Graph:** For this exercise, new groups were built in order to create new group dynamics. The groups were given a scientific article, paper prints of the Dublin Core specification[2] and the FaBiO vocabulary[3] and a blank A3 paper sheet. The learners were given the task of designing an RDF graph for the scientific article on the paper sheet by using predicates of the vocabularies and linking them with elements of a bibliographical nature as they would do while cataloguing. After 20 minutes, the groups were invited to hang their graphs on the

[2] `http://dublincore.org/documents/dcmi-terms/`
[3] `http://purl.org/spar/fabio`

wall of the room for a "vernissage" enabling the learners to walk around and look at the results of the other groups. The overall duration of this exercise, being the first profound active examination of RDF, was about 40 to 50 minutes.

iv. **Writing and Validating a Document in Turtle:** This exercise with a duration of about 30 minutes was designed to make use of the computer and a web browser. Participants were asked to transform the graph designed into a Turtle document and to validate the correctness of the syntax with the help of a Turtle validator (that unfortunately is no longer active but was to be found under http://www.rdfabout.com/demo/validator/): it mainly served to make the participants use vocabularies and define triples correctly with respect to the syntax and the semantics (although the latter was validated by the trainers). Implicitly it served also to show the participants the learning progress during the first half of the workshop.

v.i **Coding Dublin Core with Protégé:** The goal of this exercise was to make people familiar with the creation of ontologies and existing tools such as Protégé. Due to the complexity of ontologies and the "heaviness" of Protégé, a tutorial similar to the well-known tutorial for the pizza ontology was handed out to all participants so that they could individually create classes, instances, relations and define the domain and range of bibliographical data. This exercise, for which a duration of between 40 and 50 minutes was reserved, was without doubt by far the one that was most behavioristic in its nature because of its step by step guide. It mainly served to give an insight into the large and complex realm of ontologies.

v.ii **LOD @ Your Institution:** In this alternative to exercise vA the participants were once more asked to find themselves together in small teams (either those built for analyzing the documents or for designing RDF graphs and writing Turtle documents) to discuss the relevancy of the lessons learned during the training day for their institution. This was intended for them to define potential projects that could be realized within the horizon of the next one or two years. Concretely this meant for them to see which a) in house databases could or should be connected to the catalogue to build a linked data environment b) how far they could or should be connected to parts of the linked data cloud, and c) to what extend copyright issues, i.e. the openness of the data could be a matter of which to take special care. The groups were also asked to break down the project in those steps that were necessary for the conversion of MARC data into Linked Open Data. All groups were asked to briefly present their results. For this exercise 50 to 60 minutes were foreseen, depending on the number of people participating in the training.

4 Compilation and Serialization

Table 1 illustrates the programs of the three training sessions given so far with the practical modules written in italics. The whole one day program finally consisted of four sessions each of about one and half hours.

Table 1. Overview of the serialized training modules

First training: BCUL I	*Second Training: EPFL*	*Third training: BCUL II*
1a. The evolution of the Web and libraries	1b. Evolution of cataloguing standards	1a. The evolution of the Web and libraries
i. Analysis of a Turtle Document		
2. Introduction to LOD		
Short Break		
ii. Comparison of RDF/XML and Turtle Documents		
3. Introduction to RDF		
iii. Designing an RDF Graph		
Long break		
4. Introduction to Turtle		
iv. Writing and Validating a Document in Turtle		
6. The added value of LOD for libraries	5a. Introduction to Ontologies	5b. Vocabularies and Ontologies
Short Break		
5a. Introduction to Ontologies	6. The Added Value of LOD for Libraries	
v.i Coding Dublin Core with Protégé	*v.ii LOD @ your institution*	
7. The Future of Cataloguing		

From a first glance at this synopsis it becomes immediately clear that only the modules of the first and of the last training session were prone to change and that all training programs consisted of a rather stable series of theoretical input with preceding or following hands-on experiences. The modular approach proved on the other hand to be flexible enough to allow changes as a reply to either requirements of the client or as a consequence of lessons learned in foregoing trainings.

All training workshops ended with an oral evaluation of the content where all participants were asked to discuss their experiences and invited to make proposals that could be helpful to improve the program. Due to the long and somewhat exhausting training day this feedback session was rather short but generally positive. The participants were also invited to give written feedbacks via mail or the e-learning platform where all the teaching materials handed out in paperbound documents were available in a digital version.

5 Conclusions

In this paper we reported on the conception and realization of a one day training program to teach librarians the essential knowledge about Linked Open Data with special regard to challenges and changes that they face in their working environment. The training was based on different serializations of basic theoretical and practical teaching modules that were consequently improved and modified after every execution. Experience has shown us that for each training session there needs to be

slight modifications according to the needs of the institution that ordered the on-the-job training, but that the core modules and practical exercises remained unchanged. After having taught three times the program described in this paper, we can rely on a consistent and stable core while being open and flexible in the arrangement of the introductory and closing sessions of the module chain. While we continue to teach (and modify) our training program, first reflections were done to integrate this training into the Bachelor studies in Library and Information Science in which the compilation or transformation of thesauri in Linked Open Data will also play an important role. The feedback of participants will help us to respond to the needs of the professionals working in academic libraries and assure the adequate preparation of future librarians.

References

1. Danowski, P., Pohl, A.: (Open) Linked Data in Bibliotheken. De Gruyter, Berlin (2013)
2. Perkins, D.N., Simmons, R., Tishman, S.: Teaching Cognitive and Metacognitive Strategies. Journal of Structural Learning 10(4), 285–292 (1990)
3. Middendorf, J., Kalish, A.: The "Change-up" in Lectures. TRC Newsletter 8, 1 (1996)
4. Berners-Lee, T.: Linked Data - Design Issues (2006), http://www.w3.org/DesignIssues/LinkedData.html
5. Pohl, A.: Launch of the Principles on Open Bibliographic Data (2011), http://blog.okfn.org/2011/01/18/launch-of-the-principles-on-open-bibliographic-data
6. Lahaye, P.: Introduction à RDF. In: xmlfr (2004), http://xmlfr.org/documentations/tutoriels/041015-0001
7. Noy, N.F., McGuiness, D.L.: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 (March 2001)
8. Klee, C.: Vokabulare für Bibliographische Daten. In: Danowski, P., Pohl, A. (eds.) (Open) Linked Data in Bibliotheken, pp. 45–63. De Gruyter, Berlin (2013)
9. Prongué, N., Hügi, J.: Les Applications Basées sur les LOD en Bibliothèque: Un Tour d'Horizon. Arbido 3, 15–16 (2013)
10. Hügi, J., Prongué, N.: Linked Open Data: Quelles Nouvelles Compétences pour les Professionnels de l'Information. Arbido 3, 7–9 (2013)
11. Zapounidou, S., Sfakakis, M., Papatheodorou, C.: Highlights of Library Data Models in the Era of Linked Open Data. In: Garoufallou, E., Greenberg, J. (eds.) MTSR 2013. CCIS, vol. 390, pp. 396–407. Springer, Heidelberg (2013)