Integration of Scientific Information through Linked Data - Preliminary Report

Gaston Michelan*[1], Germán Braun[1,2,3], Laura Cecchi[1], and Pablo Fillottrani[3,4]

[1] *Grupo de Investigación en Lenguajes e Inteligencia Artificial*
Departamento de Teoría de la Computación - Facultad de Informática
Universidad Nacional del Comahue
[2] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*
[3] *Laboratorio de I&D en Ingeniería de Software y Sistemas de Información*
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
[4] *Comisión de Investigaciones Científicas de la Prov. Bs.As (CIC)*

Abstract By implementing Linked Open Data principles throughout the scientific community, it is possible to make publications more visible and foster collaboration, both between universities, researching groups and partners. However, data should be curated and published, and a mature infrastructure needs to be provided to support it. In this work, we analyse the Linked Data's weaknesses and propose an application prototype to evaluate the state-of-the-art methodologies and tools. As case of study, data from the authors' researching groups will be publicly available in an attempt to integrate the scientific data of the Argentinian community which is an open issue. We plan to use this data to generate bottom-up methodological guidelines and thus enrich ontology-based conceptual models.

1 Introduction and Motivation

The goal of an information integration system is to answer queries that may require extracting and combining data from multiple sources, controlled by many people, under a common scheme. The aim is to gain independence of data sources, their locations, data models and syntax and semantic variations. Particularly in the Web, systems should also focus on data published in different formats such as XML, HTML, etc., where structure and semantics are sacrificed. Furthermore, the expressivity of current mark-up languages neither allow to describe entities in Web documents nor define relations between them. In this context, Tim Berners-Lee proposed a set of best practices known as Linked Data underpinning this evolution [1, 2, 3].

Linked Data does not only provide a mean to reach the Semantic Web [4, 5], but it also allows the integration of data from multiple, distributed and

* Research Student Scholarship. Secretaría de Ciencia y Técnica. Universidad Nacional del Comahue. Argentina.

heterogeneous sources across the Web. Linked Data refers to publish and connect structured data although semistructured and unstructured data could be also integrated to the Web as a "big" and distributed database. Users could retrieve this database from their applications and combine different data as simple as providing with HTTP dereferenceable URIs to structured data.

The most important initiative in this field is the Open Data movement whose aim is to bootstrap the Web of Data through the Linking Open Data project[1]. Currently, many datasets such as DBpedia [6], WordNet [7] and DBLP [8], between others, are available. The project includes more than 600 datasets according to the last update of the *LOD cloud diagram* [9]. With reference to scientific data, the need for improving the way in which scientific results are published has been addressed by Linked Science [10]. Linked Science is an approach to interlink all scientific assets and emphasise the need for transparency when results are being evaluated, getting into practice or being used to produce new knowledge on top of the existing one. More generally, Linked Data for Science and Education is being considering a central issue in the scientific and educational community [11]. In this respect, many existing approaches such as FacetedDBLP[2] make datasets publicly accessible, but they are not integrated to other data sources as researchers websites. In fact, there exist bigger challenges at this level concerning interoperability due to the heterogeneity of the data and vocabularies, limitations in infrastructure and the lack of shared conceptualisations about involved domains. At institutional level, the University of Münster [12] and the Open University [13], between others, which can be found in the *LOD cloud diagram*, have exposed data in various repositories and have made them available openly for reuse. Both approaches are motivated by the need to make the output of the Universities more visible and foster collaboration.

As recently discussed in [14], methodological guidelines have been developed for generating, publishing and consuming linked data, but have yet to be established for creating, using and enriching ontologies in the context of Linked Data. This can be done by means of bottom-up ontology development approaches starting from non-ontological data sources and intending to (semi-) automatically develop an ontology from that, as proposed in [15, 16]. From a practical point of view, bottom-up methodologies should take some "legacy" data and get ontology concepts by executing some processing. Data can be extracted from databases [17, 18] or other sources (conceptual models, between others) or to be generated as output of natural language processing [19] or machine learning [20]. Lastly, guidelines integrated with reasoning services have been proposed by the authors of this work in [23] although they are focused on a top-down approach for domain ontologies. The core of these methodologies is a set of pattern-based extension rules [24], which suggest to users new constraints along the ontology-based conceptual modelling process.

In this respect, we propose:

[1] `http://linkeddata.org/`
[2] `http://dblp.l3s.de/d2r/sparql`

1. To focus on the data integration by following the Linked Data principles and in the context of a Linked Data Life-Cycle [21, 26].
2. To study the underlying infrastructure to support Linked Data, making a diagnosis based on the development of a framework to curate, publish and access data. This approach should follow the 5-start scheme: (1) dataset is available on the Web, (2) as structured data, (3) in a standard format (RDF [25]), (4) uses URIs to denote things and (5) links to other datasets.
3. From this framework, to develop a Web application for browsing scientific publications in the field of our researching groups. Through this application, we hope to establish connections with other educational institutions and information providers.
4. To expand the methodological guidelines in [23] by suggesting new extension rules that can be defined taking into account the underlying structured data related to the universe of discourse. Thus, the aim is to integrate new Linked Data-based bottom-up guidelines with the defined top-down ones in a joint methodology for ontology development and specifically for ontology-based conceptual modelling.

Particularly, in this paper, we tackle the objectives proposed in (1), (2) and (3) by presenting the design of a framework to extract, interlink and publish data from our researching group, called *Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA)* at Universidad Nacional del Comahue (Argentina), which in turn we plan to interlink with scientific information from the *Laboratorio de I&D en Ingeniería de Software y Sistemas de Información (LISSI)* at Universidad Nacional del Sur (Argentina). For keeping complexity and consistency under control, we will use our own dataset.

This work is structured as follows. Section 2 introduces the details of the Linked Data life-cycle. Our proposal is presented in section 3, while a preliminary evaluation is done in 4. To conclude the paper, section 5 includes a comparison with related works and section 6 elaborates on final considerations and directions for future works.

2 Linked Data Life-Cycle

Linked Data paradigm has evolved as a powerful approach to reach the Web of Data and thus the Semantic Web. Nevertheless, Linked Data only refers to a set of best practices but does not take into account the underlying methodologies, tools and infrastructure required to implement it. In this context, Auer and Lehman have started to deal with improving the performance of RDF Data Management, interlinking, algorithms and tools, and maintenance of Linked Data. The result has been a Linked Data Life-Cycle that includes different stages to undertake these issues, depicted in Fig. 1 (picture by LOD2 project).

In order to structure data, information from unstructured and semi-structured sources must be extracted and mapped to RDF, which can be stored, queried and indexed in a RDF store. New structured information can be created or corrected

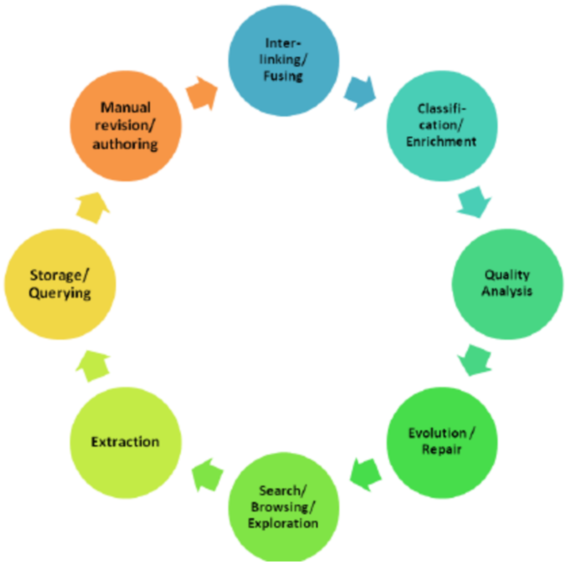

Figure 1: Linked Data Life-Cycle

by means of users. At the same time, this data must be linked to related data from different publishers, as required by Linked Data principles. Since Linked Data tries with raw data, approaches for classification and enrichment must be also provided in order to make efficient queries. Once problems are detected after analysing quality, strategies to repair them are required together with other strategies to support evolution. Finally, the life-cycle proposes to empower users by means of mechanisms to browse, search and explore its structured data in a friendly manner.

This life-cycle imposes the needs for sets of heterogeneous tools which support it and for techniques to exploit the orchestration of such tools, avoiding to tackle the life-cycle stages in isolation. In that way, the maturity of each existing tools is key to develop new applications based on the Linked Data paradigm. Aimed at speeding up this process, the LOD2 project proposed *LOD2 Stack* as an integrated set of state-of-the-art software components to employ Linked Data technologies. Its aim was to ease the installation of tools to support the life-cycle and to offer a look-and-feel to enhance the users' experience. The project ended in 2014 and the main software *LOD2 Stack* was deprecated in the very same year.

In an attempt to adopt the Linked Data life-cycle explained in this section, we propose an initial instantiation of this model for the scientific publications from the *Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA)* at Universidad Nacional del Comahue (Argentina). Nevertheless, the objective is not only get linked data, but also evaluate the current technologies to support the Web of Data and thus be used to develop mature Web Semantic applications.

Moreover, how this data can be mapped to ontologies and how can be used to enhance the conceptual level are other concerns considering a bottom-up methodological point of view.

3 A Linked Open Data proposal for Scientific Information

The first contribution is a framework covering initially the Linked Data life-cycle stages related to data extraction, storage and querying, and interlinking and fusion. The intention behind of this framework is study the existing infrastructure in order to linked data by fulfilling each life-cycle stage. Fig. 2 depicts a diagrammatic view of our proposal. Other two aspects related to authoring and search, browsing and exploration are partially covered and both are provided by the D2R tool [27]. The former facilitates users to create new information, correct it or extend it by manipulating data from the very same relational database and thus generating automatically the respective RDF triples. The latter enables users to explore the structure information through a simple Web interface allowing navigating the database's contents and gives users of the RDF data a "human-readable" preview. Moreover, D2R allows publishing metadata such as licensing and any other information to be attached to RDF documents, but it is out of the scope of this work. Lastly, despite the fact that every scientific publication is not necessarily indexed by databases like DBLP, this framework is a starting point to integrate scientific information and to establish connections with other researchers and institutions.

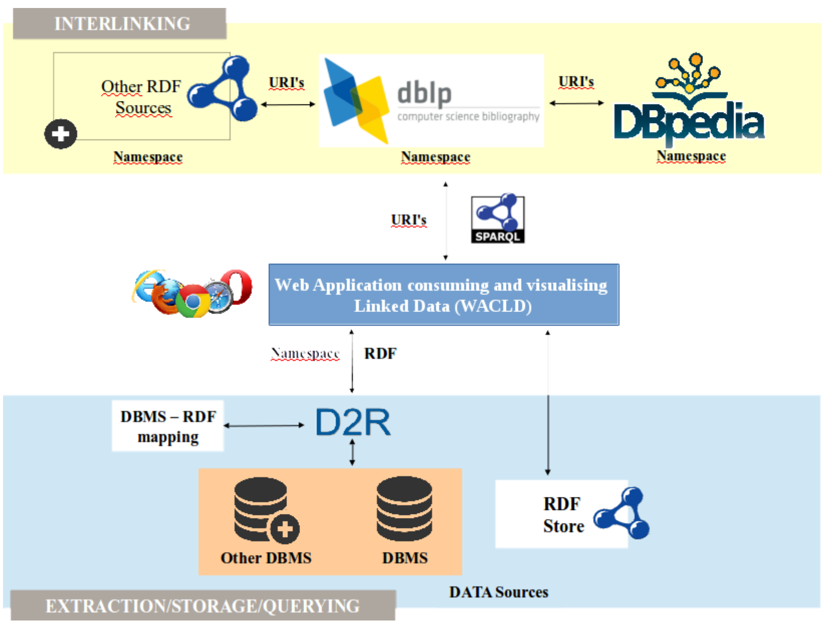

Figure 2: Overview of our proposal for Linked Data Applications.

As another contribution, we implement a prototype of a Web Application for consuming and visualising linked data (WACLD), which makes it easy for scientists to query dynamically their publications from Computer Sciences databases. It accesses to an up-to-date version of publications by means of resolvable URIs that can be queried instantly on the external SPARQL endpoints. Because the system is an initial version, data are only extracted from a SQL database. This prototype allows visualising linked data from both our own database and external sources in a standard Web application. The current implementation is not available yet so that we will write `<server>` in URIs to refer the domain where the system will be hosted.

The whole system relies on existing, open source software, especially the D2R and Apache servers. D2R also enables the tool to provide a back-end running at `http://<server>:2020/` to browse structured data, a SPARQL [28] endpoint at `http://<server>:2020/sparql` to be accessed from other applications, and a SPARQL explorer at `http://<server>:2020/snorql/` to query our own database in a friendly manner. The WACLD prototype is a PHP Web application following the Model-View-Controller (MVC) style and interfacing external SPARQL endpoints by means of the `sparqlib` PHP library.

Extraction/Storage/Querying. Data is extracted from a database located in a server from Universidad Nacional del Comahue. Until now, data from our groups only includes researchers' academic information (name, title, category, CONICET) to be published by following Five-Star open data. The extraction is based on mappings which can be generated by an automatic tool named `generate-mapping` provided by DR2 server. The mapping generator creates a customisable mapping scheme from each table to a new RDFS/OWL class together with the respective columns and written in turtle syntax. Fig 3 shows a reduced mapping for a specific table named `publications` from our database.

As we can see, the mapping makes use of namespaces as `foaf` and `dc` which have been manually added to the mapping file as well as `vocab` referencing the terms defined in our domain. Moreover, D2R helps to generate an URI scheme to represent concepts described on a particular table. The main namespace for describing data is `http://<server>/resource`. For example, for the researcher *Braun*, the URI `http://<server>/resource/researcher/Braun` will be generated. The RDF statements extracted from the `publications` table example are shown in Fig. 4.

In addition, local data transformed into RDF can be queried by means of a built-in SPARQL endpoint provided by the D2R tool. Requests are rewritten into SQL queries via a direct mapping eliminating the need for replicating the data. Anyway, in our approach, data continues being stored in the original database and no RDF storage has been installed.

Interlinking. Creating and maintaining links is initially done in a manual manner although tools for (semi-)automated links discovering are being considered but out of the scope of this work. FOAF [29] and Dublin Core [30]

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix vocab: <<server>:2020#>.
@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs:<http://www.w3.org/2000/01/rdf-schema#>.

# Table publication
map:publication a d2rq:ClassMap;
d2rq:dataStorage map:database;
d2rq:uriPattern "publication/@@publication.filenum@@";
d2rq:class vocab:publication;
d2rq:classDefinitionLabel "publication";
         .
map:publication_uri a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:publication;
d2rq:property vocab:uri;
d2rq:propertyDefinitionLabel "publication uri";
d2rq:column "publication.uri";
.
map:publication_description a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:publication;
d2rq:property dc:description;
d2rq:propertyDefinitionLabel "publication description";
d2rq:column "publication.description";
.
(...)
```

Figure 3: D2R mapping for the MySQL table `Publications` storing researcher's filenum and URI to its publications together with a description.

```
<http://<server>/resource/publication/1> vocab:filenum
                  <http://<server>/resource/researcher/Braun>
<http://<server>/resource/publication/1> vocab:uri
         "http://dblp.l3s.de/d2r/resource/authors/German_Alejandro_Braun"
<http://<server>/resource/publication/1> dc:description "DBLP"
<http://<server>/resource/publication/1> rdf:type vocab:publication
```

Figure 4: RDF tuples generated for a researcher by means of the mapping depicted in Fig. 3

between others namespaces are solved in the direct mapping defined while that links to researchers' URIs in other datasets are stored in the relational database.

Our system can be interlinked with external datasets following the Linked Data principles. In particular, this is done by means of SPARQL endpoints provided by these external sources such as Faceted DBLP. Every link to other third party dataset is maintained in a configuration file so that they can be extracted from there, supervised and tested for availability. A simple tool for SPARQL endpoints testing is located at `http://<server>/endpointTest`.

Lastly, an example for the usage of links is the combinations of data about publications from DBLP database and about researchers in GILIA, as depicted in Fig. 5. Notice that in this first prototype we are only considering the manual creation and maintenance of links since they are stored in our database.

```
SELECT ?s ?dblp
WHERE {
  ?s vocab:filenum <http://<server>/resource/researcher/1> .
  ?s vocab:uri ?dblp
}

SELECT ?paper ?title ?name ?pdf
WHERE {
  <http://<server>/resource/researcher/Cecchi> foaf:name ?name .
  ?paper foaf:maker
      <http://dblp.l3s.de/d2r/resource/authors/Laura_Andrea_Cecchi> .
  ?paper dc:title ?title .
  ?paper foaf:homepage ?pdf
}
```

Figure 5: SPARQL queries executed to retrieve publications from external DBLP datasets.

3.1 A Prototype Web Application consuming Linked Data.

As introduced above, the WACLD prototype is a Web application following the MVC style, which allows to hide the interaction with external datasets from how information is visualised. The SPARQL queries management is done by means of the `sparqllib` [31] PHP library under LGPL license. This library was selected by considering aspects as license, ease to expand and testing of SPARQL endpoints capabilities. New external datasets can be interlinked by adding a new model to interact with the respective SPARQL endpoint and a new related view, where data will be displayed according to final users preferences.

4 Preliminary Evaluation

The goal of exposing interlinked data is to make existing public data more accessible, reusable and exploitable. This can only be demonstrated through applications that make use of this data in innovative and/or cost-effective ways. The adoption of a life-cycle for this kind of applications implies to develop the corresponding methodologies and tools needed. Despite many of the proposed tools by *LOD2 Stack* are being maintained and various production systems are using this approach, the state-of-art tools and programming languages libraries are still limited. In particular, the `sparqllib` PHP library used in our prototype is part of the *Graphite PHP Linked Data Library*, but its last version was released in 2012. On the other hand, the quality of datasets, the SPARQL endpoints and their availability to be accessed and queried is other key concern. In this context, the W3C provides a list of links to alive SPARQL endopoints[3] whose last status checking was done in 2012. Another tool monitoring SPARQL endpoints statuses is SPARQLES[4].

With reference to the integration of scientific information, we have surveyed on similar approaches, but we have not found any concrete and implemented solution in the field of Argentinian community [32]. Generally, data from other researchers are maintained in common Web pages so that it should be extracted from there or make it public as linked open data. To overcome partially this issue, it is possible to interlink scientific publications from some public datasets as Faceted DBLP, ACM[5], IEEE[6] and Citeseer[7]. Nevertheless, an important number of publications are still being left out in this approach.

In the direction of bottom-up ontology design, the *LOD2 Stack* considers machine learning-based approaches to evolve with classified and enriched data into a knowledge base. However, we propose to use a stable and maintainable set of linked data to define new methodological guidelines for ontology engineering that can be integrated with top-down ones already published by authors of this work [23]. Because this is also an open issue, some related works highlighting this importance are compared below.

To sum up, issues related to the scientific information management in Argentina, the infrastructure needed for Linked Data and how the ontology engineering could make use of this data are clearly open and the proposed solutions are scarce. In order to start closing this gap, new applications should make use of linked open data. Despite this fact, such limitations should not override the power of the links in generating explicit knowledge through relationships between real-world data. Moreover, methodologies for ontology development should benefit from these valuable resources for deriving more knowledge.

---

[3] `https://www.w3.org/wiki/SparqlEndpoints`
[4] `http://sparqles.ai.wu.ac.at/`
[5] `http://acm.rkbexplorer.com/sparql/`
[6] `http://ieee.rkbexplorer.com/sparql/`
[7] `http://citeseer.rkbexplorer.com/sparql/`

5 Comparison with Related Works

In this work, we have surveyed some related works, ranging from linked data and its applications to the building of ontologies. Firstly and in order to strengthen our hypothesis, we have also analysed some results from a Linked Data survey published by the OCLC global cooperative[8]. The survey was conducted between 7 July and 15 August, 2014 and 76 out of 172 linked data project/services were described from 15 countries. 27 of them (36%) are not yet in production but only 24 project (32%) have been in production for more than two years. Moreover, 25 of the projects consume linked data, 4 publish linked data and 47 both consume and publish. The survey reports similar barriers to the ours ones such as volatility of data formats, lack of availability, unstable endopoints and outdated datasets, inconsistency in legacy data, immature software and integration with existing Web apps infrastructure. Finally, as other technical details, 20 project that publish linked data are not yet accessible and only 24 project give access to data through SPARQL endpoints. The complete survey can be accessed at `http://hangingtogether.org/?p=5206`.

The problem of the integration of scientific information has been already identified, however, proposed approaches partially solve it, mainly considering the Argentinian scientific community, where all of these topics have not been yet explored in depth. Recently, the *Instituto de Investigaciones Bibliotecnológicas* from Facultad de Filosofía y Letras at Universidad Nacional de Buenos Aires has proposed to elaborate a conceptual model about academic and scientific research by applying Linked Open Data [32]. This approach is close to our proposal from a conceptual viewpoint, however, we also propose to address it from both computational and technological aspects. Lastly, an important effort is being driven by the Linked Science movement[9] interconnecting scientific assets for enabling transdisciplinary research. This movement has grown collaboratively including international events where Linked Science aspects, vocabularies [33] and tools [34] are presented and discussed.

More specifically and with reference to the relation between ontology development and linked data, recent works have been published. Pattuelli, et.al. in [14] discuss the process of creating ontologies in the time of linked data. They propose a series of phases claiming that methodological guidelines have not been yet established for creating and using ontologies. However, although the building of ontologies is done in a bottom-up and incremental way, the guidelines consider ontology enrichment by extracting relevant information from other domains or linking vocabularies as FOAF, but do not consider the usage of logic-based reasoning leaving open some semantics aspects. Another approach from Hu and Janowicz [15] provide a workflow to mine bottom-up geographic knowledge from the LOD and thus enrich top-down ontologies. The workflow starts categorising data and classifying instances according to users. It continues by extracting common properties among the instances and finding similarities based on entropy

[8] `http://www.oclc.org`

[9] `http://linkedscience.org/`

and information gain methods. However, this work does not conclude about how this knowledge can be related to the top-down ontologies nor how can be used to enhance the initial conceptualisation.

A proposal combining data mining and ontology engineering is presented in [35], where data mining techniques make emerge implicit models and ontology engineering captures these models in reusable patterns. This position paper is close to our objectives, but it does not present any concrete implementation about how the generated patterns interoperate and converge in a common ontology. Finally and considering the ontology engineering activities, linked data could contribute in various techniques although one acts at the schema level and the other one at the instance level, respectively. An example of this can be found in [36].

6 Conclusions and Future Works

Along this work, we have identified three issues clearly open. Firstly, the need of more methodologies and technological infrastructure to support the Web of Data. On the other hand, the integration of scientific data, particularly in the Argentinian community. We have analysed the state of art tools, approaches and the OCLC survey which shows representative results. Finally, how the ontology engineering can make use of linked data to enrich models.

In this context, we have started by developing an easily scalable system to instantiate the Linked Data life-cycle together with a prototype for consuming structured data and thus promote its integration in the field of the researching groups, in which the authors of this work are involved. The dataset is not publicly available yet, but it could be explored and consumed by means of a SPARQL endpoint once released. Initially, we interlink to the external source Faceted DBLP and Citeseer in order to visualise the updated list of the members' publications in addition to other sources as DBpedia.

We plan to continue developing our prototype by interlinking more external datasets, extracting data from traditional Web pages where other publications can be taken and by integrating more aspects from the Linked Data life-cycle. From a technical point of view, we will identify new functionalities of *Graphite PHP Linked Data Library.* Issues will be also addressed. At the end, once the data has been integrated, we will work in bottom-up methodological guidelines to enrich a domain ontology and work with top-down approaches in a joint manner.

Acknowledgements

The authors would like to thank the anonymous referees for their comments and suggestions. This work is based upon research partially supported by the Universidad Nacional del Comahue (Project ID: 04/F006), the Universidad Nacional del Sur (Project ID: 24/N038), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and the Comisión de Investigaciones Científicas de la prov. de Buenos Aires (CIC).

References

1. Berners-Lee, T.: Linked Data (2006) `https://www.w3.org/DesignIssues/LinkedData.html`, accedida en Marzo de 2016.
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. (2009)
3. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. 1st edn. Morgan & Claypool (2011)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American (May 2001)
5. Yu, L.: A Developer's Guide to the Semantic Web, Second Edition. Springer (2014)
6. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web Journal (2015)
7. Miller, G.A.: Wordnet: A lexical database for english. Commun. ACM (1995)
8. Ley, M.: Dblp: Some lessons learned. Proc. VLDB Endow. (2009)
9. Max Schmachtenberg, Christian Bizer, A.J., Cyganiak, R.: Linking Open Data cloud diagram (2014) `http://lod-cloud.net/`, accedida en Marzo de 2016.
10. Kauppinen, T., de Espindola, G.M.: Linked Open Science-Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers. Procedia Computer Science (2011) Proceedings of the International Conference on Computational Science, ICCS 2011.
11. Keßler, C., d'Aquin, M., Dietze, S.: Linked Data for Science and Education. Semant. web (2013)
12. Keßler, C., Kauppinen, T.: Linked Open Data University of Münster—Infrastructure and Applications. In: Demos ESWC, Heraklion, Crete, Greece (May 2012)
13. Daga, E., d'Aquin, M., Adamou, A., Brown, S.: The Open University Linked Data - data.open.ac.uk. Semantic Web Journal (2015)
14. Pattuelli, M.C., Provo, A., Thorsen, H.: Ontology building for linked open data: A pragmatic perspective. Journal of Library Metadata **15**(3-4) (2015) 265–294
15. Hu, Y., Janowicz, K.: Enriching Top-down Geo-ontologies Using Bottom-up Knowledge Mined from Linked Data. In: Advancing Geographic Information Science: the past and next twenty years. Global Spatial Data Infrastructure (GSDI) Association (2016)
16. Keet, C.M.: Transforming semi-structured life science diagrams into meaningful domain ontologies with DiDOn. Journal of Biomedical Informatics **45**(3) (2012) 482 – 494
17. Hainaut, J.L., Chandelon, M., Tonneau, C., Joris, M.: Contribution to a theory of database reverse engineering. In: Reverse Engineering, 1993., Proceedings of Working Conference on. (1993)
18. Lubyte, L., Tessaris, S.: Automatic extraction of ontologies wrapping relational data sources. In: Proceedings of the 20th International Conference on Database and Expert Systems Applications. DEXA '09, Springer-Verlag (2009)

19. Alexopoulou, D., Wächter, T., Pickersgill, L., Eyre, C., Schroeder, M.: Terminologies for text-mining; an experiment in the lipoprotein metabolism domain. BMC Bioinformatics (2008)
20. Lehmann, J., Bühmann, L.: ORE - a tool for repairing and enriching knowledge bases. In: Proceedings of the 9th International Semantic Web Conference (ISWC2010). Lecture Notes in Computer Science, Springer (2010)
21. Auer, S., Lehmann, J.: Making the Web a Data Washing Machine—Creating Knowledge out of Interlinked Data. Semantic Web Journal (2010)
22. Baader, F., Küsters, R.: Nonstandard inferences in description logics: The story so far. In Gabbay, D., Goncharov, S., Zakharyaschev, M., eds.: Mathematical Problems from Applied Logic I. Volume 4 of International Mathematical Series. Springer-Verlag (2006) 1–75
23. Braun, G., Cecchi, L., Fillottrani, P.: Integrating Graphical Support with Reasoning in a Methodology for Ontology Evolution. In: Proc. of the 9th Int. Workshop on Modular Ontologies WoMO 15 IJCAI 15. CEUR Workshop Proceedings (2015)
24. Braun, G., Cecchi, L.: Extension Rules for Ontology Evolution within a Conceptual Modelling Tool. In: Proc. of the 1st Simposio Argentino de Ontologías y sus Aplicaciones SAOA 15 JAIIO 15. CEUR Workshop Proceedings (2015)
25. World Wide Web Consortium: RDF 1.1 Primer. Available at `https://www.w3.org/TR/rdf11-primer/`, last accessed May, 2016.
26. Auer, S., Lehmann, J., Ngomo, A.C.N.: Introduction to linked data and its lifecycle on the web. In: Proceedings of the 7th International Conference on Reasoning Web: Semantic Technologies for the Web of Data. RW'11, Berlin, Heidelberg, Springer-Verlag (2011) 1–75
27. Bizer, C., Cyganiak, R.: D2R Server – Publishing Relational Databases on the Semantic Web. Poster at the 5th ISWC (2006)
28. World Wide Web Consortium: SPARQL 1.1 Query Language. Available at `http://www.w3.org/TR/sparql11-query/`, last accessed May, 2016.
29. Brickley, D., Miller, L.: The Friend Of A Friend (FOAF) vocabulary specification (November 2007) `http://xmlns.com/foaf/spec/`, last accessed May, 2016.
30. Powell, A., Nilsson, M., Naeve, A., Johnston, P.: Dublin Core Metadata Initiative - Abstract Model (2005) White Paper.
31. University of Southampton: Graphite PHP Linked Data Library. Available at `http://graphite.ecs.soton.ac.uk/`, last accessed May, 2016.
32. Barber, E.E., Pisano, S., Romagnoli, S., Parsiale, V., Pedro, G.d., Gregui, C., Blanco, N., Mostaccio, M.R.: Aplicación de Linked Open Data para la realización de un modelo conceptual que permita diseñar un mapa de las investigaciones académicas y científicas de la Argentina. Información, cultura y sociedad (2015)
33. Baglatzi, A., Kauppinen, T., Keßler, C.: Linked science core vocabulary specification. Technical report, LinkedScience.org (2011)
34. Alonen, M., Kauppinen, T., Suominen, O., Hyvönen, E.: Exploring the linked university data with visualization tools. In Cimiano, P., Fernández, M., Lopez, V., Schlobach, S., Völker, J., eds.: The Semantic Web: ESWC 2013 Satellite Events. Volume 7955., Montpellier, France, Springer Berlin Heidelberg (2013) 204–208
35. d'Aquin, M., Kronberger, G., Suárez-Figueroa, M.C.: Combining data mining and ontology engineering to enrich ontologies and linked data. In: Proceedings of the First International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, Heraklion, Greece, May 27, 2012. (2012)
36. Scharffe, F., Euzenat, J.: Linked data meets ontology matching - enhancing data linking through ontology alignments. In: Proceedings of the International Conference on Knowledge Engineering and Ontology Development (IC3K 2011). (2011)