# Producing linked open data to describe scientific activity from researchers of Ecuadorian universities

Elizabeth Cadme
Electronic and Computing Science Department
Universidad Técnica Particular de Loja
Loja, Ecuador
iecadme@utpl.edu.ec

Nelson Piedra
Electronic and Computing Science Department
Universidad Técnica Particular de Loja
Loja, Ecuador
nopiedra@utpl.edu.ec

*Abstract*—**Semantic interoperability can be reached using universal languages of knowledge representation and adopting open standards practices (such as URIs, RDF, OWL, and LOD) advantaging the content interchange. In this work, we present a life cycle process with the following stages: data collection, processing, transformation, enrichment, and publishing, of Linked Data into the domain of research activity produced by the university teaching staff in Ecuador. This article describes the processes for collecting data from repositories, creating a semantic data model, cleaning data, transforming, linking and publishing linked data. Special attention has been given to ontological design patterns, as well as tools to ensure the semantic interoperability of the converted data. Special attention has been given to ontological design patterns, as well as tools to ensure the semantic interoperability of the converted data. The model shows these heterogeneous data into an evenly and abstract way, which is crucial in the matter of providing "intelligent" applications for research activity.**

*Keywords*— *university; open data; linked data; interoperability, open linked data; ontology*

## I. INTRODUCTION

Universities worldwide are called to develop institutional processes of internal and external evaluation, accreditation, academic classification and quality assurance. In Ecuador, the government has established an accreditation and categorization process to be applied to all universities in the country, according to the Organic Law of Higher Education of Ecuador, known in Spanish as LOES [1].

Since a significant part of these processes are based on indicators and data from research activity of their teaching staff, the higher education institutions IES are required to pull and consolidate information from databases managed by its different departments. This fact involves IES to face several issues in the data sources such not digital information, and tacit knowledge about staff who has not been registered into any source, use of different code for the same entities, information duplicity, among others, which were found at Universidad Técnica Particular de Loja (UTPL). To solve these problems, currently, we use different automatic, semi-automatic and manual processes, which demand a considerable amount of institutional resources. This background opens the opportunity to apply the Semantic Web approach, allowing stored data to be processed by machines (information systems) with intelligence, providing user-friendly tools for searching, integration and analysis of the Web available information. For Semantic Web development it is used linked data technology, which consists of three main elements such as the languages that allow structuring and describing the Web content (XML, RDF); the use of URIs (Uniform Resource Identifiers) as unique and unmistakable identifiers of Web resources, and the use of ontologies (OWL) to formally object defining, object relationships and inference rules [2][3][4].

This research shows the result of a linked data production cycle under the domain of university scientific research activity. It is based on the legal regulations for quality accreditation of Ecuadorian universities. The main objective of this work is to help universities, improving aspects such as (a) assistance when incorporating teacher researchers production which is scattered into the network, fundamental when establishing scientific and academic research information metrics from individuals or groups at institutional level; (b) support on identifying and creation of scientific cooperation networks; (c) detection of priority potential fields where politicians can help creating policies for science and technology development; and, (d) increasing Web experience about searching information related to university teacher research activity, among others.

The first part of this document presents a review of related works that fit into the university linked data. Next, the analysis legal regulations documentation for university accreditation and data source. After, a section with the development process of an ontology network (Ontura-Net) to describe university teacher research curricula elements. Then, there is a summary of conversion and publishing of linked data. Finally, conclusions of this research work and others that may derive from this project.

## II. RELATED WORKS

In the context of university field there are many semantic initiatives which interrelate different data sets produced indoors, for example *The Open University*[1], has semantically described information about teacher research work, published material, titles, courses and audiovisual educational resources [7]; the

---

[1] http://www.open.ac.uk/

Technical Innovation and Development Team (TIDT) *iSolutions* and University of Southampton, have published linked data about academic courses, important dates, technological development, infrastructure, catering, educational resources, among others; the "*Linked Up*"[2] project that boosts data exploitation obtained from different educational institutions [8]; *Linked Universities*[3] comes from an alliance between European universities compromised to publish linked data through a common space into university domain; about topic semantic description, in [5] is described a case of search improvement and detection of open educational resources OER, and into [6][13] is shown a case of OERs semantic integration.

Reference research documents have developed different guides based on experience for each case, with different domains which have incorporated semantic technology of linked data. However, the view given on this work is oriented to provide a semantic model that relates curricular items of university teachers at the same time that helps to solve the problem of diversity of data sources through the use of linked data technology.

### III. LEGAL REGULATION ANALYSIS AND DATA SOURCES

#### A. Legal regulation analysis

The domain of university data is widely broad. Thus, it is necessary to make a selection of curricular elements seen by the Evaluation, Accreditation and Quality Assurance Council of Higher Education in Ecuador, CEAACES by its Spanish name, corresponding to the research criteria in the evaluation process of higher educational institutions IES.

Also, we have reviewed the legal documents that govern the Higher Education in Ecuador, such as Organic Law of Higher Education (LOES) [1], the regulations for the LOES and regulations about teaching and researching used inside the university where the study is held. TABLE I. , displays the elements which are part of the whole evaluation profile that a research teacher has to meet through the accreditation and categorization processes in Ecuador, and, how to locate the information at UTPL, able to be used for the linked data publishing process. It incorporates data from intellectual production, projects, professional upgrading, academic lessons, and work record.

TABLE I. MAPPING OF THE LAW FRAME AND DATA SOURCE PUBLISHING. CASE UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA (ECUADOR)

| Curricular Element | Responsible dependence of information managing |
|---|---|
| Researcher data | Human Resources General Direction |
| Projects | Research Results Transfer Office |
| Academic degree | Human Resources General Direction |
| Professional upgrading and training | Research Vicerrector |
| Intellectual production | Research Vice rector |
| Work experience | Human Resources General Direction |

Data from selected curricular elements have gone through an integration process and have been stored in the institutional database system called Scientific and Academic Information System SIAC. This database feeds from dependencies displayed in TABLE I.

In the document and file database review have been identified 84 tables. It was conducted a review of the established relationships between the existent entities, and also the main and foreign keys. In addition, to know and understand the meaning of each field and to identify the subdomains to which data belong to, it was analyzed the documents described in each table. TABLE II. , shows a segment of the mapping between physical storing entities inside the Scientific and Academic Information System SIAC and the specific curricular element "intellectual production" (published research works). This mapping was made using all curricular elements and tables from SIAC.

TABLE II. MAPPING BETWEEN CURRICULAR ELEMENT "INTELLECTUAL PRODUCTION" AND PHYSICAL STORING ENTITIES INTO SIAC

| Name of the table: published articles | Description: scientific and dissemination articles | |
|---|---|---|
| **PUBLISHED ARTICLES TABLE STRUCTURE** | | |
| *Name* | *Description* | *Reference Table* |
| id | Unique article identifier | |
| tipo_articulo | Type of article (scientific or dissemination). Related table items_catalogo. Type of articles catalogue. | items_catalogo |
| pais | Country where it was published. Related table items_catalogo. Country catalogue | items_catalogo |
| indice | Name of the index where it was published. Related table items_catalogo | items_catalogo |
| estado_pub_latin_index_id | State of the article publishing Latindex. Related table items_catalogo | items_catalogo |

The process was developed for 11 tables which have information about selected curricular items. The data review has allowed knowing certain conditions about data integrity as in the case of duplicates.

#### B. Definition of URIs

The creation of the URIs assignment scheme for the elements of the data set to be converted and published has been based on the guides provided by [9].

The selected URI base has developed taking into account the type of information that belongs to the curricular elements of a researcher teacher at UTPL: http://data.utpl.edu.ec/utpl/lod/. The structure of URIs for the ontology network has been defined according to the following scheme: http://data.utpl.edu.ec/utpl/lod/ontology/{concept|property} And finally, URIs used to define instances have been modeled like this: http://data.utpl.edu.ec/utpl/lod/resource/{r. type}/{r. name}

Also, it has been accomplished the definition of the URIs scheme to be used, the selection of the type of license under

---

which are going to be published the sets of generated data, it will be Creative Commons 3.0: Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)[4]. This license allows distribution, re-mixing, editing, and creation of data from generated data, always not having commercial benefits from them, assures the recognition of its precedence and fair conditions for the generated data publishing.

## IV. Ontology network Ontura-Net

Modelling curricular elements that belong to the university researcher teachers leads to having a model based on ontologies. It has been taken each curricular element as a subdomain. Ontura-Net, consists of a network with six ontologies which model the subdomains: (1) research teacher (person); (2) academic degree (title); (3) intellectual production (published work); (4) professional record (academic upgrading and training lessons); (5) developing project participation (projects); and, (6) academic experience (work record). The developing of the ontology network Ontura-Net (Ontology network to describe university research and academic resources) was created under methodology guides provided by NeOn[11].

Fig. 1, belongs to the conceptual model of the ontology network where we can find the relationship for each subdomain, which complements with TABLE III. , that describes the reused ontologies for the development and implementation of the ontology network.
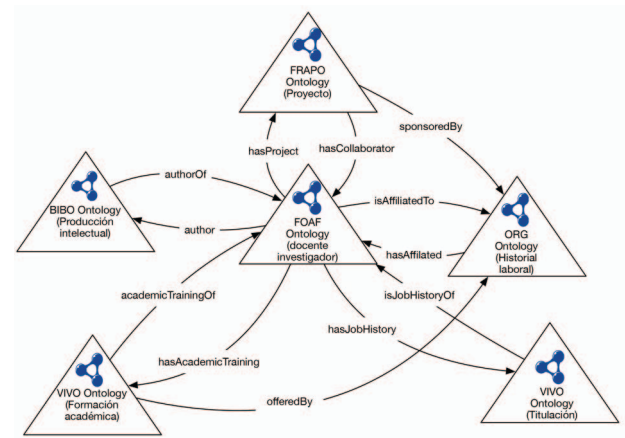


Fig. 1. Ontura-Net conceptual model

TABLE III. SELECCIÓN DE ONTOLOGÍAS REUTILIZABLES EN EL DESARROLLO DE ONTURA-NET

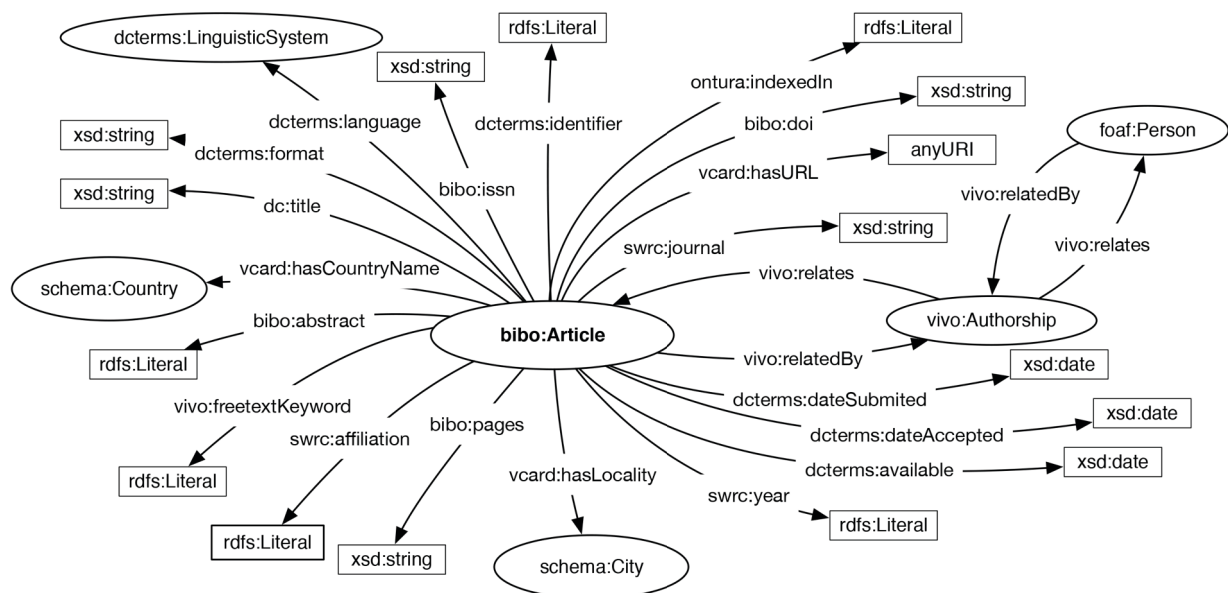| Subdominio | Ontología de dominio seleccionada |
|---|---|
| Persona | FOAF Friend of a Friend vocabulary |
| Titulación | VIVO Core Ontology |
| Publicaciones | BIBO The Bibliographic Ontology |
| Formación professional | DCTERMS The Bibliographic Ontology |
| Proyectos | FRAPO Funding, Research Administration and Projects Ontology |
| Historia laboral | ORG Core Organization Ontology |



Fig. 2. Ontologic model for the subdomain publicaciones-artículos (published articles)
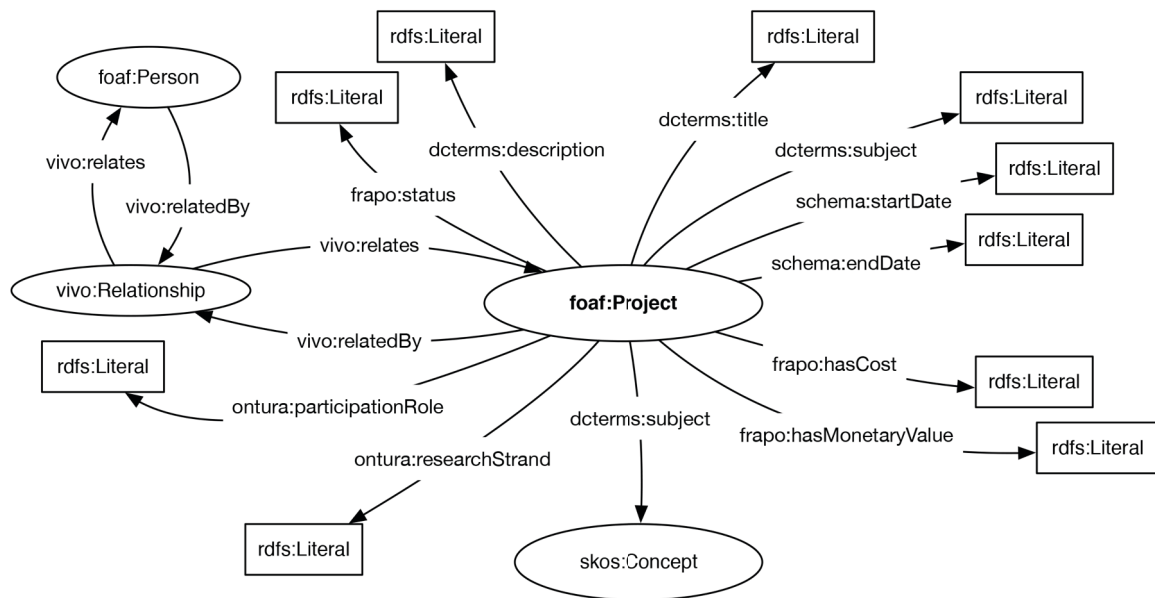
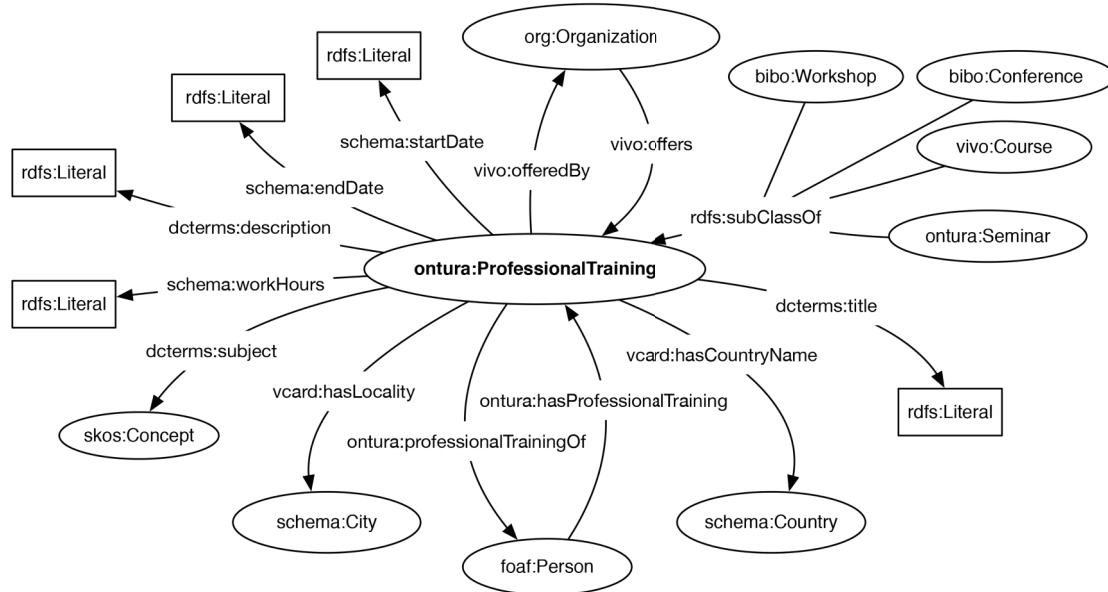Fig. 3. Ontologic model for the subdomain proyecto (project)



Fig. 4. Ontologic model for the subdomain formación profesional (professional record)

Next, it is shown an extract of the conceptual models from three of the modeled subdomains: Fig. 2, shows the conceptual model that belongs to articles. Fig. 3, displays the ontological model for the professional record subdomain, where it stands out a relationship with the organizations that extend the certificates of the professional training or upgrading lessons, type of course taken by the researcher. Fig. 4, shows the ontological model for project domain.

The developed ontology has been implemented in Protégé. It was made also a validation using the same developing tool, and an evaluation of the online tool OOPS! (OntOlogy Pitfall Scanner!)[5]

Ontura Net supports 19 classes and 56 properties, from which 5 classes and 12 properties (as shown in TABLE IV. ) belong to the developed ontology network, in other words, were created as new ones.

_____

[5] http://oops.linkeddata.es

| Ontura-Net Ontology | Type |
|---|---|
| ontura:ProfessionalTraining | Class |
| ontura:WorkHistory | Class |
| ontura:ProfessionalTraining | Class |
| ontura:Seminar | Class |
| ontura:WorkHistory | Class |
| ontura:hasProfessionalTraining | ObjectProperty |
| ontura:hasWorkHistory | ObjectProperty |
| ontura:hasProfessionalTraining | ObjectProperty |
| ontura:ProfessionalTrainingOf | ObjectProperty |
| ontura:hasWorkHistory | ObjectProperty |
| ontura:workHistoryOf | ObjectProperty |
| ontura:laboralStatus | DataProperty |
| ontura:maritalSatus | DataProperty |
| ontura:degreeLevel | DataProperty |
| ontura:imprint | DataProperty |
| ontura:indexedIn | DataProperty |
| ontura:researchStrand | DataProperty |

The Ontura-Net, OWL file, is currently published on http://data.utpl.edu.ec/vocabulary/ontura/ontura.owl, and we can find detailed information of this ontology network at http://data.utpl.edu.ec/vocabulary/ontura/

## V. GENERATION AND PUBLICATION OF RDF DATASETS

Data transformation has been held using a tool developed by the Knowledge-based Systems Research Group of UTPL[5] previously used to generate open educational resources datasets (Serendipity[6]) with academic structure (LOD UTPL[7]). This tool has allowed a flexible way to do the mapping for the Ontura-Net ontology, with information taken from the SIAC.

The result of this mapping made possible to verify the ontology elements (classes, properties), as well as to check the data type of the original source, the Scientific and Academic Information System SIAC, and the properties of Ontura-Net.

The generated triplets were stored into a triple store based on a free version of Virtuoso Universal Server. Previous to the load of the dataset it was verified that the correspondent namespaces the Ontura-Net ontology were already stored. Then, it was possible to load the triplet files divided by curricular element. Finally, we can access the stored data through:

SPARQL Endpoint: http://data.utpl.edu.ec/utpl/lod/sparql

Graph IRI: http://data.utpl.edu.ec/utpl/lod

Next step after datasets are stored and enabled the query service SPARQL [12], was to execute a linking process. TABLE V. , displays a list of the external data sources and the entities of the linked datasets: countries, cities, universities or organizations, languages; and, from a dataset already available for UTPL, it was possible to choose linking to areas, departments, and department sections.

Once datasets were linked, started an automatic and manual evaluation process of the links. The VAPOUR[8] tool was used for the automatic validation of results, a service that allows

verifying that linked data have been published in the right way, according to principles of linked data and under W3C recommendations.

| Entity | Data sources |
|---|---|
| Countries<br>Cities<br>University/Educational organizations | Dbpedia<br>Dbpedia in Spanish<br>Dbpedia Latin America |
| Language | Dbpedia<br>American Congress Library |
| Areas<br>Departments<br>Department sections | LOD UTPL |

Also, there were query tests using SPARQL for a simple of competence questions chosen according to each subdomain of the built model and where was possible to demonstrate the relationships between the different classes. A couple of query examples executed are:

***Teachers from the social-human area who have published scientific books or book chapters during 2013***

```
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX vivo: <http://vivoweb.org/ontology/core#>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
PREFIX dcterms: <http://purl.org/dc/terms/>

SELECT DISTINCT ?fname ?lname ?book ?year ?department
{ ?researcher a foaf:Person .
 ?researcher vcard:organization-unit
<http://data.utpl.edu.ec/utpl/lod/resource/Area_Socio_Humanistica> .
 ?researcher foaf:firstName ?fname.
 ?researcher foaf:lastName ?lname.
 ?researcher vcard:organization-unit ?department.
 filter regex (str(?department),"Department_")
 ?researcher vivo:relatedBy ?auths .
 ?auths vivo:relates ?doc .
 ?doc a bibo:Book .
 ?doc dcterms:title ?book .
 ?doc swrc:year ?year .
 filter regex(str(?year),"2013", "^i")
}
```

***Top 10 researcher teachers at UTPL who have published scientific articles***

```
PREFIX vivo: <http://vivoweb.org/ontology/core#>
PREFIX bibo: <http://purl.org/ontology/bibo/>

SELECT DISTINCT ?r ?names count(*) AS ?qty WHERE {?r a
foaf:Person ; foaf:firstName ?fName ; foaf:lastName ?lName .
```

---

[6] Serendipity (Linked OpenCourseWare Data Faceted Search) http://serendipity.utpl.edu.ec

[7] Linked Open Data UTPL http://data.utpl.edu.ec/?q=es/ejemplos_consultas_sparql
[8] http://linkeddata.uriburner.com:8000/vapour

BIND (CONCAT(?fName, " ", ?lName) AS ?names)
?r vivo:relatedBy ?authorship .
?authorship vivo:relates ?doc .
?doc a bibo:Article .

} GROUP BY ?r ?names ORDER BY DESC (?qty)
LIMIT 10

## CONCLUSIONS

It has been used semantic technology to represent university scientific and academic information based on the regulation institutions of higher education in Ecuador. It has been executed a practical case using information from UTPL. It was possible to integrate several heterogeneous information sources with university data through semantic technology and the use of ontologies.

About the requirement specification, it was important the analysis of curricular elements according to the university legal regulation in Ecuador. It was built an ontological model with curricular elements that correspond to a particular researcher assessment; and, on the university field still exist another area as part of it and are able to be modelled through an ontology network to complement the initial work, furthermore, it was carried out the definition of concepts, properties, relationships and restrictions over the curricular elements of a university teacher.

The reuse of ontological resources was key on the Ontura-Net ontology network design. It was not found an ontology/vocabulary that highly satisfies the specified requirements, the FOAF, VCARD, BIBO, VIVO, DCTERMS, and ORG, vocabularies were reused. It took place a satisfaction testing on modeled classes and relationships according to the Ecuadorian legal regulation, and under the agreement of assessment when data were generated.

Once developed the transformation and triplets storing process, we were able to validate and to execute queries on data from the curricular elements of a UTPL university teacher applying a SPARQL-endpoint to which is possible to access from any other Internet service point.

UTPL has scientific production datasets of its teachers, and it is linked to other datasets ready to cooperate with many applications developing processes (visualization, query, recommendation, enrichment, and others) that permit to improve its internal and external assessment processes, to create cooperation networks integrating other universities through linked data.

## FUTURE WORKS

This research has established the basis for future researches such as the extent of the scope of Ontura-Net, to complement the initial research; the application of publishing process to other institutions of higher education in Ecuador, a user interface implementation able to execute queries related to the researcher

performance indicators according to established parameters aligned to this goal. In addition, using published linked data as part of this work it will be possible to determine potential cooperation networks, the establishing of metrics to support decisions; and, the development of different kind of user-friendly visualizations.

## REFERENCES

[1] Ecuador. Registro Oficial, Ley Orgánica De Educación Superior. Ecuador, 2010, pp. 1–39.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Sci. Am., vol. 284, no. 5, pp. 28–37, 2001.

[3] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," Int. J. Semant. Web Inf. Syst., 2009.

[4] Silicon Valley Innovation Center (SVIC), "Web 3.0 and the semantic web: How the future internet will change everything," Technology, 2014. [Online]. Available: http://www.svicenter.com/technology/web-3-0-semantic-web-future-internet-will-change-everything/. [Accessed: 04-Aug-2015].

[5] N. Piedra, E. Tovar, R. Colomo-Palacios, J. López-Vargas, and J. A. Chicaiza, "Consuming and producing linked open data: the case of OpenCourseWare," Progr. Electron. Libr. Inf. Syst., vol. 48, no. 1, pp. 16–40, 2014.

[6] J. Chicaiza, N. Piedra, J. López, L. Quituisaca, F. Montaño-Sosoranga, P. Medina, and E. Tovar-Caro, "A contribution to encourage the dissemination of academic publishing: Finding diffusion media by means of a search engine based on semantic technologies," in 2016 IEEE Global Engineering Education Conference (EDUCON), 2016, pp. 854–859.

[7] E. Daga, M. d'Aquin, A. Adamou, and S. Brown, "The Open University Linked Data - data.open.ac.uk," Semant. Web J. – Interoperability, Usability, Appl. an IOS, no. November 2010, pp. 1–8, 2015.

[8] M. Guy, M. D'Aquin, S. Dietze, H. Drachsler, E. Herder, and E. Parodi, "LinkedUp : Linking Open Data for Education," 2014. [Online]. Available: http://www.ariadne.ac.uk/issue72/guy-et-al.

[9] B. Villazón-Terrazas, L. M. Vilches-Blázquez, O. Corcho, and A. Gómez-Pérez, "Methodological Guidelines for Publishing Government Linked Data," in Linking government data., New York: Springer, 2011, pp. 27–49.

[10] N. Piedra, J. Chicaiza, P. Quichimbo, V. Saquicela, E. Cadme, J. López, M. Espinoza, and E. Tovar, "Framework for the integration of digital resources based-on a Semantic Web approach," RISTI-Revista Ibérica Sist. e Tecnol. Inf., pp. 55–70, 2015.

[11] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "The NeOn methodology for ontology engineering," in Ontology engineering in a networked world, Berlin: Springer Berlin Heidelberg, 2012, pp. 9–34.

[12] C. Bizer, Silk Link Discovery Framework About Silk Silk Workbench, 1st ed. 2013.

[13] N. Piedra, "Marco de Trabajo de Interoperabilidad Semántica para la Educación Abierta: un enfoque para la Integración de Sistemas de Información de Recursos Educativos Abiertos autónomos, distribuidos y heterogéneos" (Doctoral dissertation, ETSI_Informatica), 2016.