

Ontological Modeling of POHUA: An Institutional Knowledge Repository

Esteban Castillo¹, Ofelia Cervantes¹, Ma. Auxilio Medina², José Luis Zechinelli-Martini¹

¹ Universidad de las Américas Puebla,
Department of Computer Science, Electronics and Mechatronics,
Mexico

² Universidad Politécnica de Puebla,
Postgraduate Department,
Mexico

{esteban.castillojz, ofelia.cervantes, joseluis.zechinelli}@udlap.mx, maria.medina@uppuebla.edu.mx

Abstract. This paper presents an ontology-based proposal for a digital repository called POHUA. The relevance of this paper relies on the theoretical and practical description of the performed steps of an ontological design process: gathering domain specific metadata, discovering relevant classes, modeling relationships among these classes and creating instances that show the semantic expressiveness of a digital repository. For each step, all the aspects involved in a knowledge representation tool (Protégé) are presented and discussed pointing out main highlights. The paper's main contribution includes a characterization of a specific university repository, the extraction and analysis of concepts and the specification of properties, restrictions and rules that represent all the key aspects associated to the deployment of an institutional repository.

Keywords. Open access repositories, ontologies, Protégé, OWL, RDF, metadata, classes, relationships, instances.

1 Introduction

On the web today, there are multiple **digital libraries** [7] sites that support tasks to collect, preserve, manage and retrieve electronic documents in different formats. An **open access repository** [29] can be seen as a domain digital library that comprises documents of two main types: *thematic documents* related to a specific knowledge area

and *institutional documents* which are associated to the cultural and scientific production of a university or research facility.

In Mexico, CONACyT^{1,2} (National Council of Science and Technology), created a public national repository³ to store, manage, preserve and disseminate scientific, cultural and technological knowledge that is derived from research, academic products and technological deployments from Mexican institutions. Following CONACyT's initiative and sponsorship, the Universidad de las Américas Puebla⁴ (UDLAP) launched the creation of POHUA⁵, its own digital institutional repository⁶ to compile the knowledge generated by its academic community through the years (bachelor and postgraduate studies) contained in thesis, dissertations, scientific articles, magazines, among others.

The creation of a domain-based ontology that represents the main aspects of POHUA started in parallel with the aim of constructing a richer semantic representation of an open-access

¹<https://www.conacyt.gob.mx/>

²Mexico government department that supports the scientific research through grants for students or institutions.

³<https://www.repositorionacionalcti.mx/>

⁴<http://www.udlap.mx/inicio.aspx?idioma=2>

⁵Náhuatl word that means read, tell, narrate among other interpretations.

⁶<http://repositorio.udlap.mx/xmlui/>

repository. The ontology helps to modelate the main actors, elements and interactions typically present in a university and therefore, helps to create this kind of knowledge base accurately according to the institution necessities. Additionally, it is important to mentioned that one of the main advantages of using this kind of modeling is that this representation can be applied to other universities as an standard for a specific archive. Finally, this paper details the methodology followed to create the classes, relationships, instances and rules/inferences using a well know ontology modeling software (Protégé).

The remainder of this paper is structured as follows: in Section 2 existing digital repositories associated to the modeling of institutional environments are presented. Section 3 provides details on design and implementation of the ontology created for POHUA. In Section 4 a discussion about the relevance of the ontology created is presented. Finally, implications and conclusions derived from this work so far are included in Section 5.

2 Related Work

Digital repositories have a rich background related to information management on distinct topics and domains [31, 55, 60]. These archives have become an increasingly complex landscape for public and paid articles around the web [2, 5, 43].

Among the most popular engines for finding information related to digital repositories, the citation/integration⁷ databases [23, 38, 42] have emerged as a suitable option due to different factors such as obtaining metadata⁸ associated to articles, retrieving specific metrics (like citations or number of readers) and getting access to full papers. Examples of this kind of websites can be found in Table 1.

In the case of digital archives that are widely used without many citation/integration features, the institutional repositories [37] have been gaining terrain in recent years [50, 63, 81] thank to the efficient storage, management and browsing of different types of documents related to academic

⁷Websites that enable the interaction between a user and a digital repository.

⁸Data that provides semantic information about an article.

Table 1. Citation database/integration examples

Citation engine	A	B	C	Subject area
ACM [4]	x	✓	✓	Computer science
AMiner [45]	✓	✓	✓	Networking information
CiteSeerx [46]	✓	✓	x	Scientific documents
EBSCO [20]	x	x	✓	Different subjects
Google Scholar [3, 47]	✓	✓	x	
IEEE xplore [24]	x	✓	✓	Engineering
LA referencia [32]	✓	✓	✓	Different subjects
Microsoft Academic [48]	x	x	✓	
ResearchGate [34]	✓	✓	✓	
Proquest [52]	x	x	✓	
Scopus [58]	x	✓	✓	Scientific documents
Springer [62]	x	✓	✓	Science and business

A: Free or restricted access (✓/x)

B: Metadata access (✓/x)

C: Full text-access (✓/x)

or research topics. For the case of Mexican institutional repositories, there has been a parallel growth too, but these efforts have not been enough for creating widely open sites that can handle different types of documents [1, 21]. Considering the above, Table 2 shows a comparison of international and Mexican repositories created so far to publish scientific and academic information.

From tables 1 and 2 it can be observed that there are multiple options for finding knowledge associated to scientific and cultural production of different topics, domains and even languages (most of them on English). The citation/integration engines provide tools for mining valuable metadata related to articles, books and even magazines but in most of the cases these tools do not support full-text access to documents. In contrast, the institutional repositories have been contributing to the management, curation and dissemination of full documents (in most of the cases) linked to academic or research institutions around the world.

Most of the institutional repositories manage a wide range of documents depending of the production rate of their researchers, faculty members or students. Examples of documents included in these repositories are thesis, dissertations, articles, books or essays. One of the major advantages of these repositories is the compilation of all the institutional knowledge created so far which facilitates the access and dissemination of information while the major disadvantage is

Table 2. Examples of institutional repositories

International Repositories									
Repository	Documents language	Documents type				Documents search type			Dspace software
		Thesis	Articles	Books	Others	Basic	Medium	Advanced	
B-Columbia [6]		✓	X	X	X	✓	✓	X	X
Caltech [9]	English	X	✓	X	X	✓	✓	X	X
Cambridge [10]		✓	✓	✓	✓	✓	✓	✓	✓
Columbia [11, 13]		X	✓	X	✓	✓	✓	X	X
Dialnet [18]	Spanish	X	✓	X	✓	✓	X	X	X
Harvard [22]		X	✓	X	X	✓	✓	✓	✓
MIT [35, 40]	English	✓	✓	X	✓	✓	✓	✓	✓
Oxford [8, 49]		✓	✓	✓	✓	✓	✓	X	X
PolyU [51]	Chinese and English	X	✓	X	✓	✓	✓	✓	X
RABCI [54]	Portuguese	X	✓	X	✓	✓	X	X	X
Scielo [59]	Spanish and English	X	✓	X	✓	✓	X	X	✓
UPM [80]	Portuguese	✓	✓	✓	✓	✓	X	X	X
UNESP [79]	Spanish	✓	✓	X	✓	✓	X	X	✓
Yale [82]	English	X	✓	X	X	✓	✓	✓	✓
Mexican Repositories									
Repository	Documents language	Documents type				Documents search type			Dspace software
		Thesis	Articles	Books	Others	Basic	Medium	Advanced	
COLPOS [12]		X	✓	X	X	✓	X	X	X
CUDI [17]		X	✓	X	X	✓	X	X	X
INSP [25]		✓	X	X	X	✓	X	X	X
IPN [26]		✓	✓	X	X	✓	✓	X	✓
ITESM [27]		X	✓	X	✓	✓	X	X	✓
ITESO [28]		✓	✓	X	X	✓	X	X	✓
UASLP [68]	Spanish and English	✓	✓	X	✓	✓	X	X	✓
Redalyc [56]		X	✓	X	✓	✓	✓	X	✓
Remeri [57]		✓	✓	✓	✓	✓	✓	X	✓
UACJ [64]		✓	✓	✓	✓	✓	X	✓	✓
UAEH [65]		✓	✓	✓	✓	✓	X	X	✓
UAM [66]		X	✓	X	✓	✓	X	X	✓
UANL [67]		X	✓	X	✓	✓	X	X	X
UGuadalajara [77]		✓	✓	X	X	X	✓	X	X
UNAM [78]	✓	✓	✓	✓	✓	✓	X	✓	
UDLAP POHUA <i>Proposed ontology</i>		✓	✓	✓	✓	✓	✓	X	✓

the difficulty to keep track of all the documents produced through the years.

In Mexico, institutional repositories have increased their visibility as a viable option to storage information until recent years. Most of these repositories have their own policies to produce, save and manage digital documents which make difficult the possible interaction among distinct archives.

Additionally, the use of specific metadata to describe each type of document complicates also the automatic use and analysis of documents. In this sense, CONACyT has made several efforts to integrate these repositories [15] using the same interoperability-standards, procedures and types of documents into a single national repository that links each academic and scientific document produced in each institution around the country

by means of open access policies. Finally, it is important to notice that independently of the type of repository analyzed, most of them are based on the Dspace platform [19, 33, 83] which provides an easy to use platform focused on the long-term storage, access and preservation of digital content through the interaction of users and communities. Additionally, it is important to remark that most of repositories rely on the implemented tools provided by Dspace for searching documents using classical techniques [36], but few of them apply semantic approaches to implement an advanced search that can expand queries based on the content and meaning of terms [44, 61].

3 Ontology Modeling and Implementation

In this section, key aspects associated to the creation of the POHUA⁹ ontology using Protégé are provided. In particular, the description of classes found in UDLAP's scientific and cultural production, the relationships and restrictions among these classes and the creation of instances that show the semantic expressiveness and functionalities of the ontology are discussed. It is important to remark that the ontology implemented is based on a previously created one called **Onto4AIR** [39], which models the basic functionality of a repository in a university context according to the general and technical requirements of CONACyT Call 2016 [14].

3.1 Protégé: Ontology Modeling Tool

In order to create a formal and explicit description of an institutional repository (ontology creation), the open source tool called Protégé [41] was used. This tool provides an easy to use system for creating domain models and knowledge-based applications [16, 30]. Among the different features offer by this tool, some of the most relevant for the creation of an ontology are the following:

⁹POHUA is defined in the Spanish language and its available in <https://github.com/estebancj/Notebooks>

- A friendly easy-to-use IDE¹⁰ that allows the implementation of different ontological specifications.
- Integration of different standard languages for creating ontologies like RDF¹¹ or OWL¹².
- Implementation of different interfaces for adding classes, relationships, restrictions and instances to create domain specific models.
- Usage of different visualization tools like Ontograph¹³ that allows users to interact easily with an ontological model depending of the application needs.
- Employment of multiple reasoners like Pellet¹⁴ or Fact¹⁵ for inferring knowledge based on classes, relationships and restrictions (previously created) and for supporting an automatic logical consistency validation.
- Utilization of distinct ontological query languages like Sparql¹⁶ for extracting or inferring semantic information on specific classes or instances.

From those functionalities, it can be observed the advantages of Protégé as a tool for creating domain specific models that not only stores information about a topic but also enables users to discover and infer knowledge based on semantic aspects of information.

3.2 Classes or Concepts Definition

In order to create an ontology using Protégé, the first step performed was the **extraction and definition of classes that represent concrete concepts associated to the digital repositories domain**.

¹⁰integrated development environment (IDE).

¹¹The Resource Description Framework (RDF) is a family of specifications for modeling and describing information related to a specific domain.

¹²The Ontology Web Language (OWL) is a family of knowledge representation languages for authoring ontologies.

¹³<https://github.com/elsiklab/ontograph>

¹⁴<https://github.com/stardog-union/pellet>

¹⁵<http://owl.cs.manchester.ac.uk/tools/fact/>

¹⁶<https://www.w3.org/TR/rdf-sparql-query/>

In this sense, after analyzing the scientific and cultural production of the university, in Table 3 the selected sources for obtaining relevant information considering the importance and impact of the publications are presented.

Table 3. UDLAP's scientific and cultural information sources

Source	A	B	C	Type
Thesis [71]	✓	X	✓	Postgraduate dissertations
Entorno [76]	X	X	✓	Magazines
Contexto [75]	X	X	✓	Magazines
Editorial UDLAP [73]	X	X	X	Books
Articles	X	X	X	Published documents
Datasets [69, 70, 74]	✓	✓	N/A	Collection of information
Other sources	A	B	C	Type
CVU único [72]	X	✓	✓	Professor's profile
Scopus [58]	X	✓	✓	Articles metadata
LA referencia [32]	✓	✓	✓	Articles metadata

A: Free or restricted access (✓/X)

B: Metadata access (✓/X)

C: Full text-access (✓/X)

From Table 3 it can be observed that UDLAP's contributions are mainly associated to the distribution of information in four major fields: the creation of different graduate and postgraduate thesis, the dissemination of multipurpose magazines/books, the creation of different data sources (datasets) that comprise scientific and cultural information and the publication of scientific papers. Additionally, the use of other sources for enriching the information related to publications and authors have a major role for the correct use of the information.

Considering the distinct elements available in UDLAP's sources of information, the following main features were implemented for the creation of classes that accurately represent entities in the ontology repository:

- Create a class in Protégé for each representative (and indivisible) element in the data sources or frameworks used. For example the classes constructed the elements: *collection* and *community* from the Dspace software or the elements: *file* or *institution* which model specific aspects of UDLAP's documents.
- Assign different names for each class created in the ontology. This in turn, helps

to distinguish one element from others avoiding ambiguity.

- Permit the use of class names (or abbreviations) in other languages besides Spanish, considering that much of the terminology used in Dspace and in the cultural and scientific document production is based on English. As examples, consider terms *DOI* (Digital Object Identifier) or *ISBN* (International Standard Book Number).
- Generate a concrete description of the classes using Protégé's *Comment* attribute which adds a semantic description that helps users to understand and manage ontology elements.
- Organize representative classes in a hierarchy/taxonomy where components that share similar information can be understood, accessed and manipulated efficiently. It is important to remark that Protégé implements a root class called *Thing* from which other classes inherit main characteristics.
- Add parallel classes that share the same meaning using Protégé's *equivalent to* attribute which helps to disambiguate classes that do not share the same name but are similar in terms of their role in the ontology. For example the classes *Student*, *SchoolBoy* and *Disciple* that do not have the same name but the same role and actions.
- Append special restrictions for classes that are mutually different using Protégé's *disjoint* attribute. As example consider the classes *Women* and *Men* from which no example can be an instance of both classes.

According to the main features showed below, some representative examples of the hierarchy created using Protégé are presented in Figure 1, taking into account that some classes group together others that represent specific entities in the ontology domain.

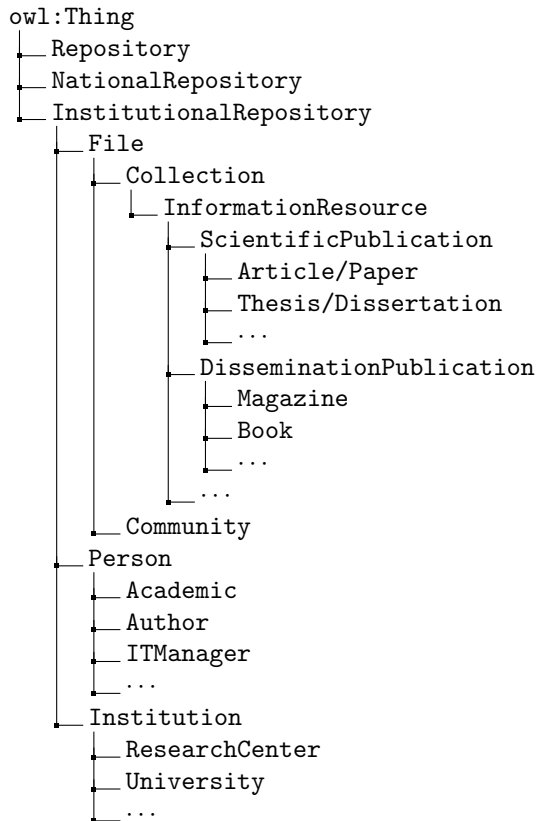


Fig. 1. Protégé's class hierarchy excerpt

3.3 Properties Definition

The second step performed for the creation of an ontology was the **creation of data properties that characterize the classes previously created for adding information related to the nature of entities**. These properties are a valuable asset for representing special qualities that classes exhibits in the context of an institutional repository and makes them relevant for the understanding and use of the proposed ontology.

Like in Section 3.2, after analyzing the different sources found in Table 3, distinct relevant data properties were observed, considering this fact, the following main features were specified for creating properties in the ontology.

- Create several data properties in Protégé for each class previously implemented, having

in mind that these properties must have a quantifying nature for obtaining discrete or continuous values. As examples of data properties created consider the creation of *Author* or *Title* that characterize information related to the journal publication class.

- Assign different names for each data property created in the ontology to distinguish one from others avoiding ambiguity.
- Generate a description of data properties created using Protégé's *Comment* attribute which adds a semantic description of the property's purpose.
- Group together representative data properties in a hierarchy/taxonomy where components that share similar information can be understood, accessed and manipulated easily. Like in the case of classes, Protégé implement a root class called *TopDataProperty* from which other data properties inherit main characteristics.
- Add a scope and a value type to each data property implemented using Protégé's *Domain* and *Range* attributes, where the first one indicates the classes attached to an specific data property and the second one indicates the type of values that a property can have (integer, float, string, etc.). For example the property *Title* which is used for several classes like *Thesis* or *Book* and has a string value. On the other hand, the data property *EmbargoEndingDate* which has a numeric/date nature is assigned to be an exclusive property of published journals.
- Apply a constraint to avoid the use of multiple values types in a data property using Protégé's *Functional* attribute.

Keeping in mind the main features presented, some representative examples of the hierarchy implemented using Protégé are presented in Figure 2, where it can be observed the value or range (\rightarrow) of some properties created as well as the scope or domain of the classes they belong to (\Rightarrow).

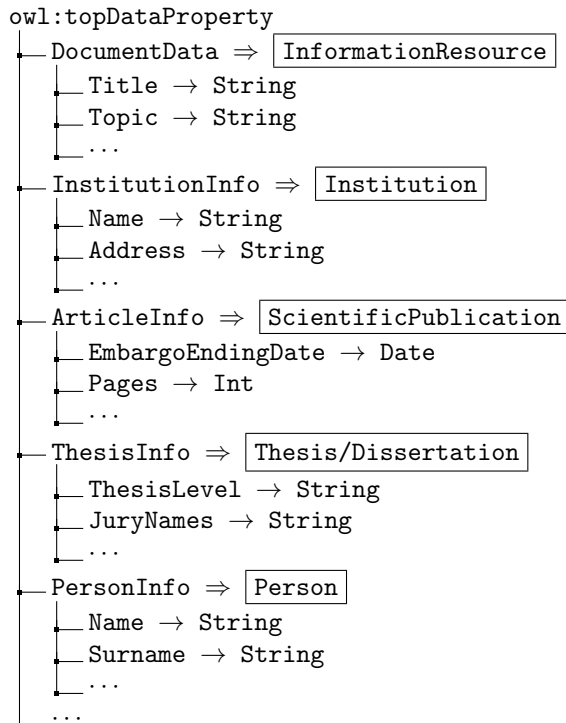


Fig. 2. Protégé's data properties hierarchy excerpt

3.4 Relationships Definition

The third step associated to the creation of an ontology is the **definition of relationships among instances to group and infer more complex information related to the elements in an institutional repository**. One of the major differences between data properties and relationships is that data properties describe the characteristics of classes while the relationships model the actions/associations among them and their instances.

After analyzing the classes and their corresponding data properties, the following guidelines were used for the creation of the relationships of the proposed ontology.

- Create multiple relationships in Protégé for each possible interaction among two instances in the context of an institutional repository. Consider as example, the creation of the relationship *ContentBy* over instances

of the classes *File* and *InstitutionalRepository*, where the goal is to capture the notion that an institutional repository stores multiple files.

- Assign different names for each relationship created in the ontology to distinguish the different kinds of relations that instances have among them.
- Generate a description of the relationships created using Protégé's *Comment* attribute for adding a semantic description of the connection between instances.
- Group together representative relationships in a hierarchy/taxonomy. Like in the case of classes and data properties, Protégé implements a root class called *TopObjectProperty* from which other relationships inherit main characteristics.
- Add a scope and a value type to each relationship implemented using Protégé's *Domain* and *Range* attributes, where the first one specifies the origin classes and the second one the destiny classes (like in a mathematical function). As example, consider the relationship *AuthorOf*, where the classes *Academic* or *Student* are the origin classes and the class *InformationResource* is the destiny class, this relationship can be read as following: an academic or student are authors of an information resource.
- Implement inverse relationships that exchange the scope and value of previously created ones. Take as example the relationship *WrittenBy* which inverse form is *AuthorOf*.
- Add Protégé's *functional*, *symmetric* and *transitive* attributes for ensuring that all relationships have the ability to deduce new information based on the connections formed with others.

Figure 3 shows a subset of the relationships defined at the Protégés root level, showing the name of the relationship and the name of the classes it relates (\mapsto). The instances of those classes constitute the domain and range of the declared relationship which is materialized when

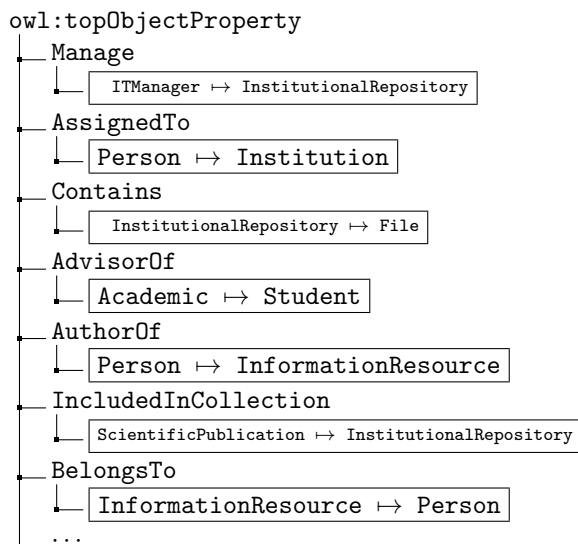


Fig. 3. Protégé's relationships hierarchy excerpt

specific individuals in the ontology are linked together.

3.5 Instances Creation

The final step related to the creation of an ontology is the **implementation of instances or representative examples that demonstrate how an institutional repository works using all the semantic expressiveness of ontology formal languages**. In this sense, different actions were followed for creating instances that represent elements in the repository.

- Create multiple instances of elements involved on the daily functionality of an institutional repository. Consider as example, the creation of specific instances of *students*, *professors*, *articles* and *books* that emulates potential users and documents classically found in a repository.
- Associate each instance to its class category previously created using Protégé's *Types* attribute. Take as example the instance *Article/Paper1* which is associated to a *ScientificPublication* class or the instance *Person1* who is related to the *Student* class.

- Assign values to the different data properties of the instances created using Protégé's *Data property assertion* attribute. For example, consider the instance *UDLAPRepository* which has a number of properties available like *RepositoryName*, *Description*, *NumberOfFiles*, etc.
- Append distinct relationships to the instances created, using Protégé's *Object property assertion* attribute. Take as example the instance *Person1* who is author of *Article/Paper1* by means of the relationship *AuthorOf*.
- Apply one of the reasoners provided by Protégé to infer new knowledge associated to each instance considering the class type, properties used and the relationships implemented. As example consider the instance *Person1* from which it is infer that also belongs to the classes *Student* and *Disciple*.

3.6 Ontology Main Features

Table 4 shows a summary of the main characteristics implemented in the ontology created for POHUA.

Table 4. Ontology metrics

Metric	Number of elements
Axioms created	1140
Logical axioms created	667
Class count	116
Data property count	98
Relationships count	16
Instances count	12
Annotations (description) count	431
Class assertions	16
Data property assertions	23
Relationship assertions	19

From the above table, it can be observed that there are multiple axioms created in the ontology which is an indicative of the variety of definitions, assertions and rules implemented. Additionally, the number of classes, properties and relationships added show the diversity of people, documents and interconnections needed to modelate accurately the functionality of an institutional repository.

Finally, the number of assertions obtained validate the correct construction of the ontology considering the consistency of the elements implemented and how well interact with the other definitions made.

4 Evaluation and Comparative Analysis

In this section, two main aspects associated to the accessibility and importance of the proposed ontology are presented and discussed. The first one is related to the creation of an evaluation tool to measure the ontology impact when it is tested by users that have a certain knowledge about digital repositories. The second one is associated to the analysis and comparison of two ontologies: the proposed ontology for POHUA and other used as inspiration or baseline, Onto4AIR.

4.1 Ontology Evaluation

One of the major challenges related to the analysis of ontologies is the technical evaluation of taxonomic main components (classes, properties, relationships and instances) for measuring the knowledge representation correctness and simplicity of specific domains [53]. For this reason, an evaluation tool was proposed in this paper for proving the acceptance and correct understanding of the semantic features created in the ontology over distinct potential users in the context of an institutional repository.

The users that tested the ontology using the evaluation tool, belong to two major groups: *expert users* that have experience in the creation of institutional repositories along with some knowledge associated to the creation of ontologies and *non-expert users* that only have some background related to the ontology terminology.

Considering the ontology main features (see Section 3.6), ten non-expert users and three expert users were selected to check different aspects related to the consistency of the ontology, taking into account six major categories: correct use of language and terminology, creation of relevant classes, use of representative properties, implementation of meaningful relationships, creation of ideal instances and discovery of new knowledge.

In this sense, Table 5 presents the overall results obtained for each group after analyzing the proposed ontology using the evaluation tool using a scale from one (worst evaluated) to five (best evaluated).

Table 5. Ontology evaluation results

Expert users					
Evaluation aspect	Overall evaluation				
	1	2	3	4	5
Software used (Protégé)			✓		
Ontology relevance				✓	
Ontology understanding					✓
Taxonomic structure					✓
Language used (vocabulary and definitions)				✓	
Classes created					✓
Data properties implemented				✓	
Relationships generated				✓	
Instances selected				✓	
Inference tools employed				✓	
Knowledge generated/inferred				✓	
Non-expert users					
Evaluation aspect	Overall evaluation				
	1	2	3	4	5
Software used (Protégé)		✓			
Ontology relevance					✓
Ontology understanding				✓	
Taxonomic structure				✓	
Language used (vocabulary and definitions)			✓		
Classes created				✓	
Data properties implemented			✓		
Relationships generated		✓			
Instances selected			✓		
Inference tools employed				✓	
Knowledge generated/inferred				✓	

From Table 5, it can be observed the following main aspects concerning the evaluation of the ontology:

- The experts and non-experts users surveyed about the importance of the ontology agree that the representation model helps to understand the overall importance of each element used on the construction of an institutional repository.
- For both types of users it can also be observed that the taxonomic structure of the ontology was easy to follow, which is an indicative of the organized nature of the information stored.
- In the case of expert users, they consider that classes, properties, relationships and instances are well organized and facilitate the understanding of the ontology. On

the other hand, the non-expert users have little difficulty understanding some of the terminology associated to these key elements in the ontology which suggests that some terminology must have concrete descriptions to improve readability.

- For the case of the inference information obtained from the ontology, both kinds of users consider that more information can be obtained if more rules or restrictions were included in the ontology.
- Finally, for both kinds of users the ontology tool (Protégé) was little difficult to follow but the structure of the ontology facilitates the overall understanding of the information, highlighting the proposed structure as a viable option to model an institutional repository in the context of a university.

4.2 Ontology Comparison

The following aspects highlight the main differences associated to the ontology created for POHUA and the ontology called Onto4AIR [39] which was used as baseline for implementing the main aspects of an institutional repository:

- The first major difference between the semantic models is that the POHUA ontology implements specific classes, data properties, relationships and instances that exemplifies the scientific and cultural production of the UDLAP community.
- The second difference concerns to the implementation of specific restrictions for the POHUA ontology (like the type of elements or their scope) according to the documents and users considered in the UDLAP repository.
- The third difference is associated to the annotation of each element created in the POHUA ontology to add semantic information to each element in the repository.

- The four difference deals with the addition of specific instances (types of documents and users) that show full functionality of the POHUA repository considering the specifics/requirements of UDLAP's academic and cultural production.
- The final major difference is the definition of a specific taxonomy for POHUA that can be adapted according to the university's usage of information compared to Onto4AIR, which establishes a number of flexible elements in the ontology to create a model depending on a university or institute specific requirements.

5 Conclusions and Future Work

In this paper, the steps performed to create an institutional repository ontology have been presented. For each step, the theoretical and practical implications have been discussed, pointing out hints and examples that illustrate the implementation of an ontology on a knowledge representation tool (Protégé). Considering the implications so far, the contributions as well as the proposed future work associated to the deployment of an ontology-domain are the following.

This work has the following contributions:

1. Review of the current state of the art trends related to the construction of digital repositories in Mexico and the rest of the world, considering the use of citation/integration engines.
2. Analysis of the university scientific and cultural production for determining the best way of implementing a digital repository ontology.
3. Extraction of actors and documents associated to the university's data sources to create classes that exemplifies relevant entities present in an institutional repository.
4. Implementation of data properties that helps to characterized and understand the nature of classes.

5. Creation of meaningful relationships among classes for emulating the interaction of main entities on an institutional repository.
6. Generation of suitable instances that illustrate the functionality of an institutional repository at UDLAP.
7. Usage of a practical tool for evaluating the applicability of an ontology on the deployment of a digital repository.
8. Analysis of the best practices for creating an open access repository associated to a university that can be applied to other institutions with a similar context.
9. Creation of a general template of the key elements and interactions related to an institutional repository that in turn can be used to understand better POHUA and therefore the best way for changing or updating specific elements.

We would like to mention as future work:

1. Creation of new classes, data properties and relationships that helps to improve the ontology created, considering new documents produced in the university.
2. Generation of new instances that covers all the participants involved in the functionality of an institutional repository.
3. Implementation of a query-based system on the semantic analysis of the ontology to discover or infer insightful knowledge related to the institutional repository.
4. Creation of distinct modules to interoperate Dspace software and the ontology proposed for adding semantic information related to the documents stored in the repository.

Acknowledgements

This work has been supported by the CONACYT grant with reference #284406 called "Repositorio institucional UDLAP: Un ambiente abierto para construir, compartir y visualizar conocimiento

universitario". The authors would also like to thank Antonio Felipe Razo Rodríguez, José Luis Velázquez García and Ivonne López Cuacuas for their invaluable help reviewing and updating this manuscript.

References

1. **Adame, S. I., Lloréns, L., & Schorr, M. (2013).** Retrospectiva de los repositorios de acceso abierto y tendencias en la socialización del conocimiento. *Revista Electrónica de Investigación Educativa*, Vol. 15, No. 2, pp. 148 – 162.
2. **Bankier J., G. & Perciali, I. (2008).** The institutional repository rediscovered: What can a university do for open access publishing? *Serials Review*, Vol. 34, No. 1, pp. 21–26.
3. **Beel, J. & Gipp, B. (2010).** Academic search engine spam and google scholar's resilience against it. *Journal of Electronic Publishing*, Vol. 13, No. 3, pp. 1–31.
4. **Bergmark, D., Phemphoonpanich, P., & Zhao, S. (2001).** Scraping the ACM digital library. *SIGIR Forum*, Vol. 35, No. 2, pp. 1–7.
5. **Borrego, A. (2017).** Institutional repositories versus researchgate: The depositing habits of Spanish researchers. *Learned Publishing*, Vol. 30, No. 3, pp. 185–192.
6. **British-columbia (2018).** Theses and dissertations. <http://guides.library.ubc.ca/theses/thesesubc>.
7. **Buehler, M. A. & Trauernicht, M. S. (2007).** From digital library to institutional repository: a brief look at one library's path. *OCLC Systems & Services: International digital library perspectives*, Vol. 23, No. 4, pp. 382–394.
8. **Burgess, L., Jefferies, N., Rumsey, S., Southall, J., Tomkins, D., & Wilson, J. (2016).** From compliance to curation: Ora-data at the University of Oxford. *Alexandria*, Vol. 26, No. 2, pp. 107–135.
9. **Caltech (2018).** Caltech library service. <https://authors.library.caltech.edu/>.
10. **Cambridge (2018).** Apollo, University of Cambridge Repository. <https://www.repository.cam.ac.uk/>.
11. **Cocciolo, A. (2010).** Can web 2.0 enhance community participation in an institutional repository? the case of pocket knowledge at teachers college, Columbia University. *The Journal of Academic Librarianship*, Vol. 36, No. 4, pp. 304 – 312.

12. **COLPOS (2018)**. Repositorio del Colegio de Postgraduados. [Http://www.biblio.colpos.mx:8080/jspui/](http://www.biblio.colpos.mx:8080/jspui/).
13. **Columbia (2018)**. Academic commons repository. <https://academiccommons.columbia.edu/>.
14. **CONACYT (2016)**. Convocatoria repositorios institucionales 2016. <https://www.conacyt.gob.mx/index.php/convocatorias-conacyt/convocatorias-conacyt/convocatorias-direccion-adjunta-de-planeacion-y-evaluacion/convocatoria-2016-repositorios-institucionales-aaicti/13336-convocatoria-ri-2016>.
15. **Conacyt (2018)**. Repositorio nacional. <https://www.repositorionacionalcti.mx/>.
16. **Cross, V. & Pal, A. (2008)**. An ontology analysis tool. *International Journal of General Systems*, Vol. 37, No. 1, pp. 17–44.
17. **CUDI (2018)**. Repositorio de corporación universitaria para el desarrollo de internet. <http://videoteca.cudi.edu.mx/>.
18. **Dialnet (2018)**. Repositorio institucional de la Universidad de la Rioja. <https://dialnet.unirioja.es/>.
19. **Dspace (2018)**. About dspace. <https://duraspace.org/dspace/about/>.
20. **Ebsco (2018)**. Who we are. <https://www.ebsco.com/about/who-we-are>.
21. **Galina, I. & Giménez, J. (2008)**. An overview of the development of open access journals and repositories in Mexico. *Proceedings ELPUB2008 Conference on Electronic Publishing*, pp. 280–287.
22. **Harvard (2018)**. Digital acces to scholarship at Harvard (dash). <https://dash.harvard.edu/>.
23. **Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., Reinhardt, W., & Stock, W. G. (2015)**. Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, Vol. 39, No. 2, pp. 255–264.
24. **IEEE (2018)**. About IEEE xplore digital library. <https://ieeexplore.ieee.org/xpl/aboutUs.jsp>.
25. **INSP (2018)**. Biblioteca digital del Instituto Nacional de Salud Pública. <http://www.inspvirtual.mx>.
26. **IPN (2018)**. Instituto Politécnico Nacional / Tesis Institucionales. <http://itzamna.bnct.ipn.mx/>.
27. **ITESM (2018)**. Repositorio Institucional del Tecnológico de Monterrey. <https://repositorio.itesm.mx/>.
28. **ITESO (2018)**. Repositorio institucional del Instituto Tecnológico de Estudios Superiores de Occidente. <https://rei.iteso.mx/>.
29. **Jones, R., Andrew, T., & MacColl, J. (2006)**. The institutional repository in the digital library. In *The Institutional Repository*, Chandos Information Professional Series. Chandos Publishing, pp. 1 – 30.
30. **Kurilovas, E. & Juskeviciene, A. (2015)**. Creation of web 2.0 tools ontology to improve learning. *Computers in Human Behavior*, Vol. 51, No. 1, pp. 1380–1386.
31. **Kutay, S. (2014)**. Advancing digital repository services for faculty primary research assets: An exploratory study. *The Journal of Academic Librarianship*, Vol. 40, No. 6, pp. 642 – 649.
32. **LA-Referencia (2018)**. Acerca de la referencia. <http://www.lareferencia.info/joomla/es/>.
33. **Lighton, P. & Hussein, S. (2013)**. Flexible design for simple digital library tools and services. *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pp. 160–169.
34. **Lovett, J. A., Rathemacher, A. J., Boukari, D., & Lang, C. (2017)**. Institutional repositories and academic social networks: Competition or complement? a study of open access policy compliance vs. researchgate participation. *Journal of Librarianship and Scholarly Communication*, Vol. 5, No. 1, pp. 1–35.
35. **MacKenzie, S. (2002)**. DSpace: An institutional repository from the MIT libraries and Hewlett Packard laboratories. *Research and Advanced Technology for Digital Libraries*, Springer Berlin Heidelberg, pp. 543–549.
36. **Manning, C. D., Raghavan, P., & Schütze, H. (2008)**. *Introduction to Information Retrieval*, chapter Computing scores in a complete search system. Cambridge University Press, pp. 135 – 149.
37. **Mariscal, J. L. & Girarte, J. L. (2017)**. Repositorios digitales para los procesos de formación e investigación en gestión cultural. *Revista de Investigación en Gestión Cultural*, Vol. 1, No. 3, pp. 1 – 22.
38. **Marx, W., Schier, H., & Wanitschek, M. (2001)**. Citation analysis using online databases: Feasibilities and shortcomings. *Scientometrics*, Vol. 52, No. 1, pp. 59–82.
39. **Medina, M. A., Sánchez, J. A., Cervantes, O., De la Calleja, J., & Benitez, A. (2017)**. Representación semántica de conocimiento operativo y de

- dominio para repositorios institucionales. Registro público del derecho de autor, México, Número: 03-2017-042511235500-01.
40. **MIT (2018).** Dspace@mit repository. <https://dspace.mit.edu/>.
 41. **Musen, M. A. (2015).** The Protégé project: A look back and a look forward. *AI Matters*, Vol. 1, No. 4, pp. 4–12.
 42. **Neuhaus, C. & Daniel, H. D. (2006).** Data sources for performing citation analysis: an overview. *Journal of Documentation*, Vol. 64, No. 2, pp. 193–210.
 43. **Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H. R. (2012).** Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing*, Vol. 25, No. 3, pp. 195–206.
 44. **Nisheva, M. M., Pavlov, P. I., & Stanchev, P. L. (2016).** Semantic search in heterogeneous digital repositories: Case studies. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, pp. 65–72.
 45. **Ortega, J. L. (2014).** *Academic Search Engines*, chapter AMiner: science networking as an information source. Chandos Publishing, pp. 47 – 70.
 46. **Ortega, J. L. (2014).** *Academic Search Engines*, chapter CiteSeerx: a scientific engine for scientists. Chandos Publishing, pp. 11 – 27.
 47. **Ortega, J. L. (2014).** *Academic Search Engines*, chapter Google Scholar: on the shoulders of a giant. Chandos Publishing, pp. 109 – 141.
 48. **Ortega, J. L. (2014).** *Academic Search Engines*, chapter Microsoft Academic Search: the multi-object engine. Chandos Publishing, pp. 71 – 107.
 49. **Oxford (2018).** Oxford university research archive. <https://ora.ox.ac.uk/>.
 50. **Paul, S. (2012).** Institutional repositories: Benefits and incentives. *The International Information & Library Review*, Vol. 44, No. 4, pp. 194 – 201.
 51. **PolyU (2018).** Hong kong polytechnic university research archive. <http://ira.lib.polyu.edu.hk/>.
 52. **Proquest (2018).** Who we are. <http://www.proquest.com/about/who-we-are.html>.
 53. **Raad, J. & Cruz, C. (2015).** A survey on ontology evaluation methods. *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 179–186.
 54. **RABCI (2018).** Repositório acadêmico de biblioteconomia e ciência da informação. <http://rabci.org/rabci/>.
 55. **Radovan, V. (2011).** Digital repositories and the future of preservation and use of scientific knowledge. *Informatologia*, Vol. 44, No. 1, pp. 55–62.
 56. **Redalyc (2018).** Acerca de redalyc. <http://www.redalyc.org:9081/home.oa>.
 57. **REMERI (2018).** Acerca de REMERI. www.remeri.org.mx/.
 58. **Rodrigues R., S., Taga, V., & Passos, M. F. (2016).** Research articles about open access indexed by Scopus: A content analysis. *Publications*, Vol. 4, No. 4, pp. 185–192.
 59. **Scielo (2018).** Biblioteca científica electrónica en línea. <http://www.scielo.org/php/index.php?lang=es>.
 60. **Simons, N. & Richardson, J. (2013).** *New Content in Digital Repositories*, chapter Introduction. Chandos Publishing, pp. 1 – 10.
 61. **Solomou, G. & Koutsomitropoulos, D. (2015).** Towards an evaluation of semantic searching in digital repositories: a dspace case-study. *Program*, Vol. 49, No. 1, pp. 63 – 90.
 62. **Springer (2018).** Springer driving academic publishing since 1842. <http://www.springer.com/la/about-springer/history>.
 63. **Swan, A. & Carr, L. (2008).** Institutions, their repositories and the web. *Serials Review*, Vol. 34, No. 1, pp. 31 – 35.
 64. **UACJ (2018).** Repositorio Institucional de la Universidad Autónoma de Ciudad Juárez. <http://ri.uacj.mx/vufind/>.
 65. **UAEH (2018).** Biblioteca Digital de la Universidad Autónoma del Estado de Hidalgo. <http://dgsa.uaeh.edu.mx:8080/bibliotecadigital/>.
 66. **UAM (2018).** Repositorio institucional de Universidad Autónoma Metropolitana. <http://zaloamati.azc.uam.mx/>.
 67. **UANL (2018).** Colección Digital de la Universidad Autónoma de Nuevo León. <https://cd.dgb.uanl.mx/>.
 68. **UASLP (2018).** Repositorio Institucional de Acceso Abierto de la Universidad Autónoma de San Luis Potosí. <http://ninive.uaslp.mx/jspui/>.
 69. **UDLAP (2017).** Índice global de impunidad. <http://www.udlap.mx/cesij/>.
 70. **UDLAP (2018).** Colección de arte. <http://www.udlap.mx/arteycultura/coleccion/>.

71. **UDLAP (2018)**. Colección de tesis digitales. [Http://catarina.udlap.mx/u_dl_a/tales/](http://catarina.udlap.mx/u_dl_a/tales/).
72. **UDLAP (2018)**. CVU único. [Http://inscripciones.udlap.mx/curriculum/](http://inscripciones.udlap.mx/curriculum/).
73. **UDLAP (2018)**. Editorial universitaria. [Http://www.udlap.mx/arteyculturaudlap/libros-y-publicaciones.aspx](http://www.udlap.mx/arteyculturaudlap/libros-y-publicaciones.aspx).
74. **UDLAP (2018)**. Índice global de impunidad. [Http://www.udlap.mx/igimex/](http://www.udlap.mx/igimex/).
75. **UDLAP (2018)**. Revista contexto. [Https://contexto.udlap.mx/](https://contexto.udlap.mx/).
76. **UDLAP (2018)**. Revista entorno. [Http://www.udlap.mx/entorno/](http://www.udlap.mx/entorno/).
77. **UGuadalajara (2018)**. Repositorio institucional de la Universidad de Guadalajara. [Http://www.crea.udg.mx/index.jsp](http://www.crea.udg.mx/index.jsp).
78. **UNAM (2018)**. Repositorio Institucional RAD/Universidad Nacional Autónoma de México. [Http://www.rad.unam.mx/](http://www.rad.unam.mx/).
79. **UNESP (2018)**. Repositório institucional unesp. [Https://repositorio.unesp.br/](https://repositorio.unesp.br/).
80. **UPM (2018)**. Archivo Digital de la Universidad Politecnica de Madrid. [Http://oa.upm.es/](http://oa.upm.es/).
81. **Xia, J. & Opperman, D. B. (2010)**. Current trends in institutional repositories for institutions offering master's and baccalaureate degrees. *Serials Review*, Vol. 36, No. 1, pp. 10 – 18.
82. **Yale (2018)**. About elischolar. [Https://elischolar.library.yale.edu/about.html](https://elischolar.library.yale.edu/about.html).
83. **Yin, Z. & Hsin L., C. (2015)**. Data management and curation practices: The case of using DSpace and implications. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, pp. 109–113.

*Article received on 09/08/2018; accepted on 25/01/2019.
Corresponding author is Esteban Castillo.*