Description of educational resources from open repositories using semantic technologies

María Auxilio Medina Nieto a,*, Jorge de la Calleja Mora b, Eduardo López Domínguez c, Ismael Everardo Bárcenas Patiño d, Delia Arrieta Díaz e and Araceli Ortiz Carranco f

a Graduate Departament, Universidad Politécnica de Puebla, Juan C. Bonilla, Puebla, México
E-mail: maria.medina@uppuebla.edu.mx

b Graduate Departament, Universidad Politécnica de Puebla, Juan C. Bonilla, Puebla, México
E-mail: jorge.delacalleja@uppuebla.edu.mx

c Department of Computer Science, Laboratorio Nacional de Informática Avanzada A. C., Xalapa, Veracruz México
E-mail: eduardo.lopez@lania.edu.mx

d Faculty of Engineering, Universidad Nacional Autónoma de México, Ciudad de México, México
E-mail: ebarcenas@unam.mx

e Faculty of Economics, Accounting and Management, Universidad Juárez del Estado de Durango, Durango, México
E-mail: dad@ujed.edu.mx

f Financial Engineering, Universidad Politécnica de Puebla, Juan C. Bonilla, Puebla, México
E-mail: araceli.ortiz@uppuebla.edu.mx

Abstract.

This paper presents a process to describe educational resources from open repositories, the goal is to provide the foundations for the deployment of semantic web services of shared interest to managers, developers and users within educational organizations. The stages of the process use semantic technologies, for example, ontologies that models users, resources and their properties found in Mexican institutional repositories according to three calls to construct repositories promoted by the National Council of Science and Technology in Mexico. These ontologies are written in English and Spanish languages, they enable the establishment of a common vocabulary that reduces ambiguity and allows users to attend their specific information needs and support knowledge acquisition from the repositories domain. These ontologies reuse the foaf, schema.org and skos vocabularies and their scope is limited by a set of competency questions. The paper suggests practical applications derived from the formalisms of ontologies such as working with incomplete information and describes the implementation of an ontological assessment method; subjects were organized in two groups: managers - developers and users. The first group followed a competency questions-based approach and determined by using SPARQL queries that the ontologies are able to represent questions and answers using its own terminology; whereas the second group participated in an exploratory and self-management survey, perceptions and opinions were gathered in questionnaires. The results showed positive attitude of users to promote ontologies. The process is of value as this fosters unique and formal definitions of concepts. A second type of impact is that this enables the construction of machine readable datasets that describe and enrich educational resources.

Keywords: educational resources, repositories, ontologies, machine readable datasets, semantic web

*Corresponding author. E-mail: maria.medina@uppuebla.edu.mx.

1. Introduction

For the National Council of Science and Technology (CONACYT), the main Mexican agency that funds re-
search, an institutional repository (IR) is a centralized technological framework that complies with international standards to store, manage, preserve and disseminate scientific, technological and innovative information that is derived from educational activities, technological developments and research processes [1], [2]. According to [3], IRs are technologies for knowledge sharing, they collect, index and share the intellectual capital of faculty and research staff, mainly scholarly publications and educational resources.

In a developing country like Mexico, endeavors to build IRs have been undergoing for several years; first, the Mexican Network of Institutional Repositories (REMERI, after the initials in the Spanish statement Red Mexicana de Repositorios Institucionales), integrated open access repositories [4] and the most recent, the National Repository (NR), another CONACYT initiative [5], has supported the construction of IRs through the 2015, 2016 and 2017 calls [6], [7] and [8]. The functionality and technical requirements for Mexican IRs were specified in [1] and [2]. At the time of this writing, the NR interoperates with 104 IRs and its home page indicates that this provides full-text access to 91,143 resources and that the number of queries is 8,925,985.

The NR has search mechanisms to retrieve educational resources that match keywords, authors, titles, subjects and dates. Technically, IRs implement the 2.0 version of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH protocol) [9], an international interoperability standard that uses unqualified Dublin Core (DC)\(^1\) as the default metadata format [10].

By exploring the descriptions of educational resources from four IRs from the central part of Mexico, (that belong to Universidad Politécnica de Puebla (UP-Pue) [11], Laboratorio Nacional de Informática Avanzada A. C. (LANIA) [12], Universidad de las Américas Puebla (UDLAP) [13], and Universidad Tecnológica de la Mixteca (UTM) [14]), differences in the use of DC elements were observed in spite of that these IRs interoperate with the NR and that the NR establishes predetermined general and technical requirements [1] and [2]. As a consequence, the retrieval of educational resources is difficult especially in scenarios where the integration of results from multiple IRs is expected.

As a way of illustration of the exposed problematic, consider that the DC contributor element sometimes stores the name of a thesis advisor, others the second author of an article or even the name of a main-funding organization\(^2\). Therefore, efforts are required along the lines of unifying uses and meanings in the descriptions of educational resources from IRs in order to attend specific information needs.

This paper presents a process to describe educational resources from IRs, the goal is to provide the foundations for the deployment of semantic web services of shared interest to managers, developers and users. The stages of the process use semantic technologies such as instances of an ontology called Onto4AIR\(^3\), a newer version of the Onto4AIR ontology originally described in [15].

This ontology models users, resources and their properties in English and Spanish languages, this has been designed to formally establish a common vocabulary that reduces ambiguity and enables knowledge acquisition from the repositories domain. The paper describes how foaf\(^4\), schema.org\(^5\) and skos\(^6\) vocabularies are integrated into Onto4AIR2 ontology, a set of competency questions expressed as SPARQL\(^7\) queries that determine its scope as well as its prototypical implementation for UPPUE and LANIA repositories.

Figure 1 shows the context of the proposed process. On one hand, educational organizations have one or more IRs that store educational resources, the descriptions of these resources are transformed into instances of Onto4AIR2 ontologies to construct machine-readable and semantically enriched datasets that support seamless data integration. Ontologies are knowledge representation models defined in [16] as formal specifications of shared conceptualizations. According to [17], ontologies define relevant concepts for a particular application.

The paper is organized as follows. Section 2 contains a detailed overview of related work. Section 3 presents the proposed process. Practical applications derived from the formalisms of Onto4AIR2 ontologies

---

\(^1\) Dublin Core is a standard set of basic metadata terms

\(^2\) More information about the usage of DC contributor element is available at: https://www.dublincore.org/specifications/dublin-core/usagewd/appendix_roles/

\(^3\) Onto4AIR2 is an acronym of Ontologies for All/Any Institutional Repositories, the number 2 denotes a second version as well as the use of two languages, English and Spanish.

\(^4\) http://xmlns.com/foaf/spec/

\(^5\) http://schema.org

\(^6\) https://www.w3.org/2004/02/skos/

\(^7\) SPARQL is an acronym for SPARQL protocol and RDF query language.
are explained in Section 4 while Section 5 describes the results of an ontological assessment method. Finally, we conclude in Section 6 with a summary of the present work along with further research perspectives.

2. Related work

Open repositories store educational resources from educational organizations such as universities and research centers around the world. A lot of IRs are supported by technological platforms such as Eprints [18], Digital commons [19], existDB [20], Greenstone [21], Fedora Commons [22] or DSpace [23]; these platforms have Graphical User Interfaces (GUIs) and information retrieval mechanisms. The impact of IRs is known and reported in the literature and databases such as OpenDOAR [24].

Frequently, IRs store descriptions of educational resources in relational databases like the repositories that interoperate with the NR, that is, descriptions are stored in tables according to specific metadata formats [25]. From our point of view, the disadvantages of this type of storage in scenarios where the integration of results from multiple IRs is expected are due to the visibility of the different elements in the basic or advanced GUIs for searching as well as the ambiguity in the result sets. This section presents related works that use semantic technologies to support knowledge acquisition, data integration and management of resource descriptions.

SpecINT is an hybrid framework for data integration and federation in semantic data query processing over RDF cheminformatics and bioinformatics repositories; this framework is designed for scenarios where integration is not possible because there are no mapping schemes between repositories [26]. SpecINT uses graph eigenvectors, vertices ranking and existing pattern SPARQL queries to identify relevant data sources.

ORDI is an open-source ontology middleware that enables data integration; this implements the RDF-like data model to integrate different structured data sources including Relational DataBase Management Systems (RDBMS). This Java middleware processes and storages metadata or context information regardless of specific ontologies [27].

Ontologies in [28] are used to integrate data resources and workflows to support on line science ac-
tivities, (semantic e-science); authors propose to focus on ontologies for representing scientific artifacts such as theories, models and on line tools to enable scientists to directly generate and test such representations as means of innovation.

In [29], semantic repositories refer to DataBase Management Systems (DBMSs) or other software components that store RDF data, support inference according to the semantics of these data and provide users of real-time query-answering mechanisms. OWLIM is a set of semantic repositories originally developed as part of the Semantic Knowledge Technologies (SEKT) and Triple Space Communication (TRIPCOM) European Research Projects. OWLIM has two variants: 1) SwiftOWLIM, the free-for-use version that uses an in memory RDF database, an inference-engine and a query-answering engine and 2) BigOWLIM, the commercial version, an enterprise-grade database management system that can handle huge volumes of RDF data; both variants are packaged as storage and inference layer for the Sesame open RDF framework. At present, OWLIM is used in the life sciences, telecoms and publishing sectors as a data-integration platform for massive amounts of heterogeneous data.

The use of ontologies in industrial scenarios is reported in [30]; the author describes ontologies as conceptual models and codification artifacts. Ontological design patterns are used during the creation, verification and validation of the ontologies which are then integrated into specialised tools.

An application of ontologies in the business domain is presented in [31], where an ontology serves as a reference model for an historical organizational memory (HOM). Such as IRs, a HOM stores digital resources. The methodology to construct this ontology implemented the next activities: a) experts were interviewed to gather main terms and b) main terms were used to represent document types, collections, subjects, the audience and the dissemination channel of the HOM. Search engines and information retrieval mechanisms use ontologies to index and classify the resources; equivalent relationships support multiple names for terms in different languages.

The use of ontologies in academic contexts, the management of descriptions and research documents using semantic technologies has been reported in the following works.

In the Tecnológico de Monterrey Institute (ITESM), an intelligent platform for knowledge acquisition is formed by an information system, a multi-agent system, a knowledge management system and a knowledge information interpreter that coordinates repositories, domain ontologies and databases; repositories store educational resources such as journal articles, research-based books, patents, technology licensing, trademarks and documents of technology-based start-up companies [32]. An analysis of ontology-based methodologies for integrating and reconciling information in IRs is presented in [33], due to ontologies deal with syntactic and semantic heterogeneity, authors proposed an agile method that minimizes the need of ontological expertise when semi-structured data are used (spread sheets); in this work, university ranking data are modeled as ontology instances.

The representation of students, teachers, monographs, theses and their relationships by using ontologies is also proposed in [34]; ontologies integrate a set of rules that formally model the two stages of the graduation process; main concepts for these ontologies were gathered by means of interviews to the staff of different universities.

The widely used Leigh University Benchmark [35] has an OWL ontology to describe the structure of a university with synthetically generated datasets. The description includes departments, individuals and relations between them. LUMB(8000) stores data from 8000 universities and contains about 1.1 Billion explicit statements, while LUMB(90000) contains over 12 Billion explicit statements and nearly 21 Billion after inference.

[36] describes the Bowlogna ontology, this was designed to support the gradual adoption of the overall description of the Bologna reform across European universities. This ontology models academic settings and includes the creation of new administrative procedures, this is also use to integrate information systems. End-users benefit from this ontology through a GUI that supports facetted search and browsing systems for course information.

The works listed below are more related with the process to integrate educational resources descriptions due to the origin of the data. On one hand, a solution for organizations that uses DSpace platform to support their IRs is described in [37], authors propose three methods for exporting data about scientific activity stored in the Current Research Information System (CRIS) at the University of Novi Sad; this system implements the Common European Research Information Format (CERIF) data model. One of these methods involves the use of the OAI-PMH protocol and the OAICat library. The exported data form datasets that enables the creation of graphs that show links.
between departments, faculties and researchers; these links are useful to discover research leaders and common research areas.

On the other hand, [38] describes a system that converts the IR database supported by the DSpace platform into an intermediate database with a normalized schema that is then transformed into an ontology; the aim is to share information from the IR with other information systems to discover common interests. The ontology obtained from the IR and ontologies from other systems are integrated by means of semantic correspondence between entities.

A solution that does not depend on DSpace platform is reported in [25], authors propose an ontology-facilitated sharing as an alternative to integrate IRs data; their method consists of transforming data from IRs into ontologies that are queried by users from a unique web page. An overview of how the ontology engineering field has evolved in the last decade that discusses some of the unsolved issues and opportunities for future research is described in [39].

3. Description of educational resources from IRs using semantic technologies

Figure 2 shows the process to describe educational resources from IRs. The purpose is in both the producer’s and consumer’s interest to make access to these resources as straightforward as possible. The main activities for each stage are explained in the following sections.

3.1. Stage 1: export data from IRs

IRs are supported by technological platforms that allow managers to export data into different formats such as CSV\(^8\). At this stage, it is necessary to choose the data that will be part of the machine readable datasets. As a way of illustration, by using DSpace 5.2, a set of theses from the UPPue IR produces a CSV file of 35 columns (of DC metadata) that are organized as follows:

**Group 1:** 11 columns are empty

**Group 2:** 7 columns store information related with the load and availability of the theses’ files

**Group 3:** 6 columns have the same and unique value for all the theses. (es and en are used for Spanish and English languages, respectively)

- `dc.description.statementofresponsibility[es]`
- `dc.publisher`
- `dc.right.uri[es]`
- `dc.rights[es]`
- `dc.language`
- `dc.type`

**Group 4:** 5 columns store information of authors:

- `dc.contributor.advisor[]`
- `dc.contributor.author`
- `dc.contributor.author[]`
- `dc.contributor[es]`
- `dc.creator[es]`

**Group 5:** 6 columns store content information:

- `dc.description.abstract[en]`
- `dc.description[es]`
- `dc.identifier.citation[es]`
- `dc.identifier.uri[]`
- `dc.subject.classification[es]`
- `dc.title[es]`

The chosen data belong to group 3, 4 and 5.

3.2. Stage 2: creation of ontologies

An overview of methodologies to create ontologies is presented in [40], some of them are based on IEEE standard 1074-1995. In the case of Onto4AIR2 ontology, the steps proposed in [41] were implemented, they are summarized as follows:

1. Determine domain and scope
2. Reuse existent vocabularies
3. Enumerate the main concepts
4. Define classes and construct its hierarchy
5. Assign properties for classes
6. Define properties between instances
7. Create instances

Those steps can be implemented in a different order or on a cyclical basis by using the Protégé ontology editor [42]. The implementation of these steps are explained in the next subsections.
3.2.1. Determine domain and scope

Onto4AIR2 ontology is designed to support knowledge acquisition from the repositories domain and to describe educational resources from IRs; this is written in English and Spanish languages. Table 1 shows groups roles of potential Onto4AIR2 users.

As it was mentioned in Section 1, the scope of Onto4AIR2 ontology is determined by a set of Competency Questions (CQs), (see Table 2), they represent key information requirements. More information about CQs can be found in [43] and [44].

3.2.2. Reuse of existent vocabularies

The use of these established vocabularies, mainly RDF and OWL for semantic web applications, minimizes terms ambiguity and improves reusability. Figure 3 shows the vocabularies integrated into Onto4AIR2 ontology.

3.2.3. Enumerate main concepts

By taking into account the CQs of Table 2, the main concepts of Onto4AIR2 ontology are the following:

- Institutional Repository
- Educational resources
- Person (producer’s and consumer’s of educational resources)
- Educational organization
- Subjects (knowledge areas and fields)

3.2.4. Define classes and construct its hierarchy

Once main concepts are enumerated, they are defined as classes, then remaining concepts are obtained by generalization, specialization and by implementing middle-out strategies. Figure 4 shows the class hierarchy of Onto4AIR2 ontology. It is worth to mention that classes for educational resources are disjoint.
Table 3

<table>
<thead>
<tr>
<th>Property</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>identifier</td>
<td>M</td>
</tr>
<tr>
<td>irName</td>
<td>M</td>
</tr>
<tr>
<td>languageOfDocuments</td>
<td>R</td>
</tr>
<tr>
<td>manager</td>
<td>M</td>
</tr>
<tr>
<td>numberOfDocuments</td>
<td>R</td>
</tr>
<tr>
<td>typeOfDocuments</td>
<td>R</td>
</tr>
<tr>
<td>updateDate</td>
<td>M</td>
</tr>
<tr>
<td>website</td>
<td>M</td>
</tr>
</tbody>
</table>

Fig. 5. Properties for an instance of the IR class

Definitions for some classes, that can be used as quick guidelines for ontology users, are included into rdfs:isDefinedby properties; clarity and completeness of definitions are supported by CONACYT documents [1] and [2]. Some fruitful comments and notes are also included in rdfs:comment and rdfs:seeAlso properties.

3.2.5. Assign properties for classes

Properties are assigned to classes, for example, Table 3 contains the properties for the InstitutionalRepository class. According to the NR specifications [2], the values can be mandatory (M) or recommended (R).

In general, properties for classes are modeled as data properties, for example, Figure 5 shows the properties for an instance of the InstitutionalRepository class.

3.2.6. Define properties between instances

A set of object properties is defined to relate persons, knowledge areas and fields with educational resources as is showed in Table 4. The notation is as follows: if a property has an asterisk *, this means that this accomplishes with that facet; the abbreviations are interpreted as functional (F), inverse functional (IF), asymmetric (A) and irreflexive (I).

Unlike properties of Table 4, note that 1) the facets from the knows property are reflexive, symmetric and transitive and 2) the contribution of authors is represented in the subproperties of isAuthorOf called isFirstAuthorOf and isCoauthorOf.

3.2.7. Create instances

Populating the ontologies with instances allow users to have knowledge bases. In the case on Onto4AIR ontology, most instances belongs to persons, educational resources, knowledge areas and fields, for example, Figure 6 shows information about an instance of the MasterThesis class.

3.3. Stage 3: description of educational resources

The original descriptions of educational resources come from the exported data of IRs, thus they are brief and pointed. According to [2], each resource has at least a title, a year, a subject and an author, these data are modeled as annotation properties (see Figure 6).

In summary, each educational resource is considered an information unit that is integrated into the ontology by means of instances and annotation, data and object properties. Although this is a time consuming task,
Table 4

<table>
<thead>
<tr>
<th>Properties between instances</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasAdvisorOfMasterThesis</td>
<td>Student</td>
</tr>
<tr>
<td>hasKnowledgeArea</td>
<td>CreativeWork, Document</td>
</tr>
<tr>
<td>hasKnowledgeField</td>
<td>CreativeWork, Document</td>
</tr>
<tr>
<td>isAuthorOf</td>
<td>Person</td>
</tr>
<tr>
<td>isCoauthorOf</td>
<td>Person</td>
</tr>
<tr>
<td>isFirstAuthorOf</td>
<td>Person</td>
</tr>
<tr>
<td>isManagedBy</td>
<td>InstitutionalRepository</td>
</tr>
<tr>
<td>knows</td>
<td>Person</td>
</tr>
<tr>
<td>manages</td>
<td>IR-manager</td>
</tr>
<tr>
<td>worksIn</td>
<td>Teacher</td>
</tr>
</tbody>
</table>

Reasoners use all the information to infer new relationships or to support logical consistency automatically. Therefore, the more information is added to instances, the more inferences can be done.

3.4. Stage 4: generation of datasets

By using the Protégé ontology editor [42], machine readable datasets are formed by exporting a populated ontology using established vocabularies such as RDF [45] or Ontology Web Language (OWL) [46]. Some datasets for UPPUE and LANIA repositories are available at: http://informatica.uppuebla.edu.mx/~mmedina/ontologias/. Therefore, in order to support seamless data integration, data from IRs can be stored in different instances of Onto4AIR2 ontology as desired to form semantically enriched datasets, that is, semantically lift data.

4. Practical applications derived from the formalisms of Onto4AIR2 ontology

The establishment of axioms, cardinality, domain and range restrictions support organizational policies such as the following:

- To take into account a maximum number of authors for each type of educational resource
- To manage relationships between educational resources, knowledge areas and fields
- To discover possible data inconsistencies

The implementation of organizational policies improves the value of the generated datasets. Some practical applications for managers and developers derived from the formalisms of Onto4AIR2 ontology are the following:

- **Modeling educational resources as instances.** Capacity to insert incomplete information. Examples of use:
  - A master thesis can be inserted although a student does not have an identifier
  - An educational resource can be stored even when there is no information about its subject
- **Modeling persons as instances.** Instances for user types can belong to any, one or more classes, (some classes are disjoint while others have common elements). Examples of use:
  - Instances can store information about university applicants
  - An undergraduate teacher and can also be a graduate student
  - Persons related with each other by means of educational resources
- **Working with disjoint classes.** The types of educational resources are disjoint classes, this implies that each resource has a unique type. Examples of use:
  - An educational resource is a thesis or an article but this can not be of both types simultaneously
  - Any educational resource that has more than an associated type will produce an inconsistency; this allows document managers to identify incorrect information
- **Management of cardinality restrictions.** Unlike relational database schema, cardinality restrictions are not limited to 1-1, 1-N or M-N relationships, they can have a minimum, maximum or exactly value. Examples of use:
A thesis has only a first author that belongs to the Student class.
Five is the maximum number of authors for an article.
All thesis has exactly one advisor.
An educational resource has exactly one title, although there can be alternative titles.

Data properties management. Properties between instances and data types have domain and range restrictions. Examples of use:
- All the educational resources have a title and a publication date.
- Authors are instances of the Student or Teacher classes.

Inference mechanisms. Universal and existential restrictions are used by reasoners to infer new knowledge. Examples of use:
- A person is an advisor if there is a thesis that has been addressed by him or her.
- All authors and co-authors are related with at least one educational resource.
- Any educational resource has a first author.
- New classes of educational resources can be formed by taking into account properties to which they relate.

Functional facets. Examples of use:
- A teacher only belongs to an educational organization.
- An educational resource is associated with a unique knowledge area.

Equivalence relationships between classes support English and Spanish languages, different names for the same instance can also be used.

5. Results and discussion

This section describes an ontological assessment method of Onto4AIR2 ontology, the purpose is to estimate its quality and relevance. Participants were organized in two groups: managers - developers and potential users (testers). Section 5.1 and 5.2 present the test scenarios and discuss the obtained results.

5.1. Assessment by managers - developers

CQs are a simple means to verify satisfiability of requirements by knowledge retrieval. Three managers and three developers experts in IRs domain and semantic technologies used the CQ-based approach described in [44] to validate the accomplishment of the CQs presented in Table 2. Figure 7 refers to CQ1, definitions for classes are included into skos:note and rdf:isDefinedBy properties. The types of educational resources (CQ2) are shown in Figure 8 while Section 3.3 summarizes how educational resources are described (CQ3). Figure 9 shows an example of the description for an article.

A Simple Knowledge Organization System (SKOS) was designed to classify educational resources according to the seven knowledge areas and 24 fields of CONACYT (CQ4). Figure 10 shows an SPARQL query that retrieves the labels for knowledge fields in Spanish and English languages.

Figure 11 illustrates some educational resources that are associated with an instance of the Teacher class.
class (CQ5); the relationships between persons and resources are retrieved by using SPARQL queries and object properties. For example, Figure 12 shows an excerpt of the isCoauthorOf property.

The SPARQL query associated with the affiliation of authors (CQ7) is the following:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX onto: <http://localhost:8080/mauxOntologies/onto4AIR2.owl#>
SELECT ?subject ?object ?name
WHERE { ?subject onto:worksIn ?object .
?subject foaf:firstName ?name .}
```

Fig. 11. A SPARQL query that uses the isCoauthorOf property.
The SPARQL query that retrieves the knowledge areas (CQ8) is as follows:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?area ?y
WHERE {
  ?area skos:inScheme ?y
} order by ?area
```

In summary, the implementation of the CQ-based approach described in [44] allows the group of managers and developers to determine that Onto4AIR2 ontology is able to represent CQs and their answers using its own terminology. This group was also in charge of logical consistency validation; all the inconsistencies were corrected before the ontology was sent to potential users group as is explained in the Section 5.2. The reasons used were Fact++ 1.6.5, Hermit 1.4.3.456 and Pellet.

### 5.2. Assessment by potential users

An exploratory and self-management survey was designed to gather the opinion and perception about the structure and the elements of Onto4AIR2 ontology from a group of 30 potential users; this group is formed by students and teachers randomly selected from the staff involved in projects to construct IRs from the next educational organizations: UPPue [11], UDLAP [13], LANIA [12] and UTM [14], (see Table 5). A scale of high, medium and low values is used to represent experience with IRs, ontologies and semantic technologies.

A questionnaire was designed to gather opinions and perceptions about correctness and language expressiveness, this was composed of five closed questions, three dichotomy questions, a multiple choice question and a section to write suggestions. Unlike the suggestion section, all the questions were considered mandatory. A Likert scale of five values (varying from 1 = total disagree to 5 = total agree) was used to answer the five closed questions about the following elements:

1. Class names
2. Class hierarchy (see Figure 4)
3. Definitions of classes
4. Description of an IR (data properties for instances of the InstitutionalRepository class, see Table 5)
5. Object properties (terms used in the labels of object properties, see Table 4)

True and false values were used to answer the three dichotomy questions, they were designed in such a way that users need to realize the following tasks using the version 5.2 of the Protégé editor:

- Complete the description of a thesis
- Find the types of educational resources
- Check if a person is a co-author of a specific educational resource

Finally, a multiple-choice question asks for ontologies usefulness. All users were asked to complete the questionnaires in a period of three days; questionnaires and a prototypical implementation on Onto4AIR2 ontology for the UPPUE repository were sent by email.

A manager was in charge of integrating the results of closed and dichotomy questions from the 30 questionnaires, the results join with the answers of the open questions were analysed by members of the managers - developers group during on line sessions; all the suggestions and fruitful comments were used to produce newer versions of the ontology until no changes were detected.

The managers and developers group made several tests to verify that definitions do not cause discrepancies with class axioms, object property axioms and data property axioms; additional synonyms and antonyms were also introduced. The latest version of the ontology was sent to users again and a new period of three days was assigned to complete the questionnaires for a second time.

Table 6 shows the averages per ontology elements. The overall average suggests that users have positive perceptions and opinions of these elements. Table 7 shows the results for dichotomy questions. The two incorrect answers of Table 6 row 1 represent that two users only complete the title as the unique descriptor for a thesis, whereas the four incorrect answers of row 43 indicate a high variability.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of teachers:</td>
<td>12 (6 female, 6 male)</td>
</tr>
<tr>
<td>Number of students:</td>
<td>18 (10 female, 8 male)</td>
</tr>
<tr>
<td>Average age of teachers:</td>
<td>47 years</td>
</tr>
<tr>
<td>Average age of students:</td>
<td>25 years</td>
</tr>
<tr>
<td>Average daily use of internet:</td>
<td>7.5 hours</td>
</tr>
<tr>
<td>IR frequency of use:</td>
<td>25 high, 5 medium</td>
</tr>
<tr>
<td>Experience with IRs:</td>
<td>26 high, 4 medium</td>
</tr>
<tr>
<td>Experience with ontologies:</td>
<td>28 high, 2 medium</td>
</tr>
</tbody>
</table>

Table 5

Table 6 shows the averages per ontology elements.
Table 6

<table>
<thead>
<tr>
<th>Ontologies element</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class names:</td>
<td>4.8</td>
</tr>
<tr>
<td>Class hierarchy:</td>
<td>3.8</td>
</tr>
<tr>
<td>Definitions of classes:</td>
<td>4.3</td>
</tr>
<tr>
<td>Description of an IR:</td>
<td>4.1</td>
</tr>
<tr>
<td>Object properties:</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>Overall average:</strong></td>
<td><strong>4.4</strong></td>
</tr>
</tbody>
</table>

Table 7

<table>
<thead>
<tr>
<th>Task for dichotomy question</th>
<th>Number of answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete the description of a thesis</td>
<td>28 correct, 2 incorrect</td>
</tr>
<tr>
<td>Find the types of educational resources</td>
<td>30 correct, 0 incorrect</td>
</tr>
<tr>
<td>Check if a person is a co-author of a specific educational resource</td>
<td>26 correct, 4 incorrect</td>
</tr>
</tbody>
</table>

**Percentage:** 93.33% correct, 6.67% incorrect

Table 8

<table>
<thead>
<tr>
<th>Use</th>
<th>Number of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop semantic web services</td>
<td>24</td>
</tr>
<tr>
<td>Construct machine readable datasets</td>
<td>26</td>
</tr>
<tr>
<td>Enrich data from IRs</td>
<td>23</td>
</tr>
<tr>
<td>Information retrieval mechanisms</td>
<td>23</td>
</tr>
<tr>
<td>Knowledge acquisition</td>
<td>28</td>
</tr>
<tr>
<td>Reuse knowledge</td>
<td>23</td>
</tr>
<tr>
<td>Share knowledge</td>
<td>25</td>
</tr>
</tbody>
</table>

3 refer to users did not find the difference between the concepts author and co-author. The results of dichotomy questions were positive with an overall percentage of correct answers of 93.33%.

The multiple choice question was used to gather data about uses of Onto4AIR ontology. Table 8 shows the possible choices and the number of users that identified each choice.

Additional uses mentioned by users were democratization and visualization of information. Furthermore, the following suggestions were gathered:

- Change the label of `updateDate` data property to `modificationDate` or `lastUpdate`
- Represent collaboration between persons of different educational organizations

Finally, the Net Promoter Score (NPS) was also introduced into the questionnaires that were sent for the second time to users, this is a standard question widely used to estimate users’ attitude about a tested object [47]. Users were asked for a numerical answer between 0 and 10, (minimum and maximum value); the answers between [0,6], [7,8] and [9,10] intervals are associated with a detractor, indifferent or promoter attitude, respectively. The average of the answers is considered as the final value for the tested object. The final value of the NPS for Onto4AIR ontology was 8.9; this value indicates a positive users’ attitude.

6. Conclusions

This paper presented a process to describe educational resources from IRs. The stages of this process use semantic technologies such as established vocabularies, ontologies and SPARQL queries. In particular, the Onto4AIR was described, this formally represents persons, educational resources and their properties in English and Spanish languages. Persons or IR users are modeled by using the foaf vocabulary, educational organizations and educational resources are represented by using schema.org while a SKOS was proposed to include knowledge areas and fields that were used to classify educational resources. The vocabulary of this ontology can be shared between users and computers.

The scope of Onto4AIR ontology was determined by a set of competency questions that were expressed as SPARQL queries, the answers to some of these questions were illustrated with figures that showed the use of the Protégé ontology editor.

Relevant related works have showed the successful use of ontologies for knowledge acquisition, integration and reconciling information in different contexts; however, information about assessment is limited. In this paper, managers, developers and testers participated in an ontological assessment method. On one hand, the managers and developers used reasoners for validation of logical consistency and followed a CQ-based approach, the experimental results allowed them to establish that the ontology was able to represent CQs and their answers using its own terminology. On the other hand, an exploratory and self-management sur-
vey was done to gather the opinions and perception of potential users; the results indicated a positive testers’ attitude to promote ontologies as useful software tools.

Onto4AIR ontology enables knowledge sharing and acquisition from the repositories domain and the construction and management of machine-readable datasets, this represents an alternative to increase quality, reliability and reusability of data from IRs. The ontology is of utility as this gives institutional visibility, this can be fostered by other educational organizations with little adjustments such as the incorporation of new terms, object properties or restrictions.

The paper is a preliminary guide that provides of foundations for systematic enrichment of data and knowledge-based applications. At present, we are working with the implementation of software modules to transform IR data into ontologies instances directly. We expect that our process foster the benefits of open access policies.

As future work, we plan to work in the implementation of new extensions that explore potential power of automated reasoning and support data integration for later analysis.

References


[48] CONACYT, Términos de referencia de la Convocatoria 2015 para desarrollar repositorios institucionales de acceso abierto a la información científica, tecnológica y de innovación, Consejo


