



# **Maestría en Ingeniería en Sistemas y Cómputo Inteligente**

**Adding semantic information to data of the  
UPPue institutional repository**

*Paulo Daniel Vázquez Mora*

*María Auxilio Medina Nieto*

*Jorge de la Calleja Mora*

*Antonio Benitez Ruiz*



*Symposium  
de Posgrado*



Paulo Daniel Vázquez, María Auxilio Medina, Jorge de la Calleja, Antonio Benitez  
 Maestría en Ingeniería en Sistemas y Cómputo Inteligente

{paulo.vazquez4303, maria.medina, jorge.delacalleja, antonio.benitez}@uppuebla.edu.mx  
 Tercer Carril del Ejido Serrano S/N, San Mateo Cuanalá, Juan C. Bonilla, Puebla, México

## 1. Introduction

The implementation of institutional repositories (IRs) satisfies preservation and publication needs. Search mechanisms in IRs are based on keywords or use descriptive data (or metadata).

This poster proposes the addition of semantic information to data of the theses collection that belongs to the Universidad Politécnica de Puebla repository (RI-UPPue) in an instance of the Onto4AIR ontology [1], the purpose is to show that semantic ontologies extend search mechanisms in IRs as well as to take advantage of the automatic checking of logical consistency. The access to these data is available through a web service called semantic retrieval-01, from now on, RS-01.

Before RS-01 design was made, REST service catalog of NR was verified by users with administrator profiles, some all datasets in NR can be accessed. Table 2 shows this information.

Table 2. Examples of web services supported by NR

Datasets in NR accessed by REST web services	
Knowledge areas	License
Knowledge fields	Location
Knowledge disciplines	County
Knowledge subdisciplines	Access level
Audience	Country
State	Person

## 3.2 RS-01 design

RS-01 service implementation is based on a design available in [2]. Figure 2 shows a high level design of this service.

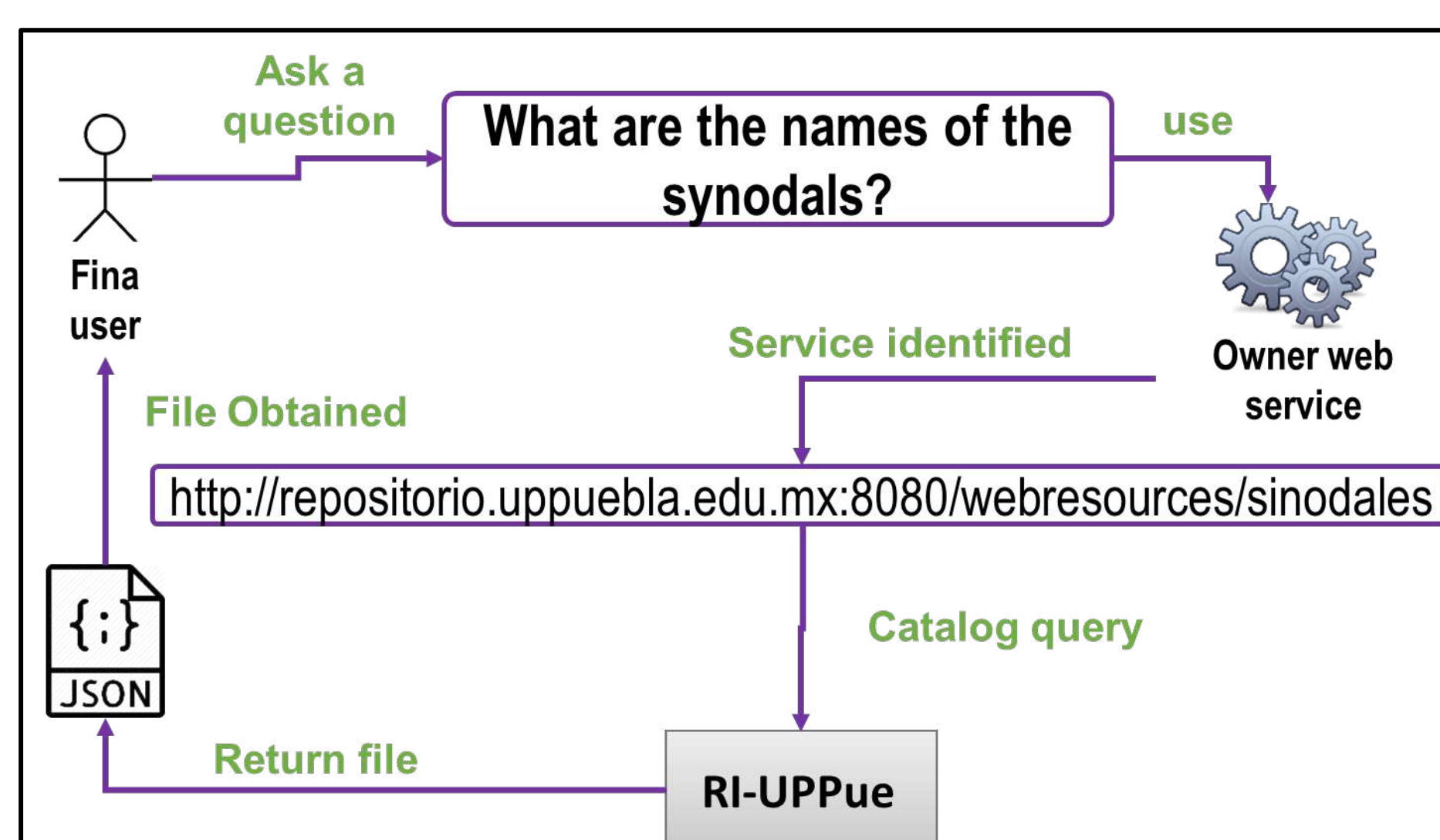


Figure 2. Service RS-01 architecture

The RS-01 service is invoked when a URL is submitted. Table 3 shows the RS-01 inputs and outputs; Figure 3 shows an example of results displayed, note that JSON format is used as output. Results can be processed by using the *requests* library of Python and storing them in any type of structure such as *list* or *tuple*. The POST, DELETE or PUT requests would be used to add, delete or update the information.

Table 3. RS-01 inputs / outputs.

Input	JSON output
<a href="http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales">http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales</a>	List of all synodals
<a href="http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales/&lt;cveSinodal&gt;">http://repositorio.uppuebla.edu.mx:8080/webresources/sinodales/&lt;cveSinodal&gt;</a>	Synodal data identified by "cveSinodal"



Figure 3. Example of JSON file result.

## 2. Goal

Manage semantic information related with synodals in a dataset of theses collection using Onto4AIR ontology and the RS-01 web service

## 3. Methodology

The methodology consists of using Onto4AIR ontology that is described in [1] and to implement the RS-01 web service; the general features of this services are presented in [2]. Figure 1 shows the main steps of this methodology.

1. Analysis of requirements
2. RS-01 design
3. Population of Onto4AIR ontology
4. Installation of DSpace RDF module at the RI-UPPue
5. RS-01 codification
6. Functionality and usability tests

Figure 1. Semantic information management at the RI-UPPue

## 3.1 Analysis of requirements

Actually the Mexican National Repository (NR) integrates data of 88 IRs [3], its interoperability is based on the implementation of the 2.0 version 2.0 of the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), this protocol uses the Dublin Core (DC) metadata standard. According with the technical specifications of RN, the DC element used to store the student name of a thesis is *dc:creator* while director or synodals names are stored in the *dc:contributor element*. As a result, there is no way to know which is the specific role of a teacher in a thesis. Table 1 shows the RS-01 requirements addressed to solve this problematic.

Table 1. Requirements of RS-01 service.

Querying of semantic information	High
Capacity to store RDF-tuples	High
Data exportation to non proprietary data as JSON or RDF	Medium
Data integration of other IRs	Medium

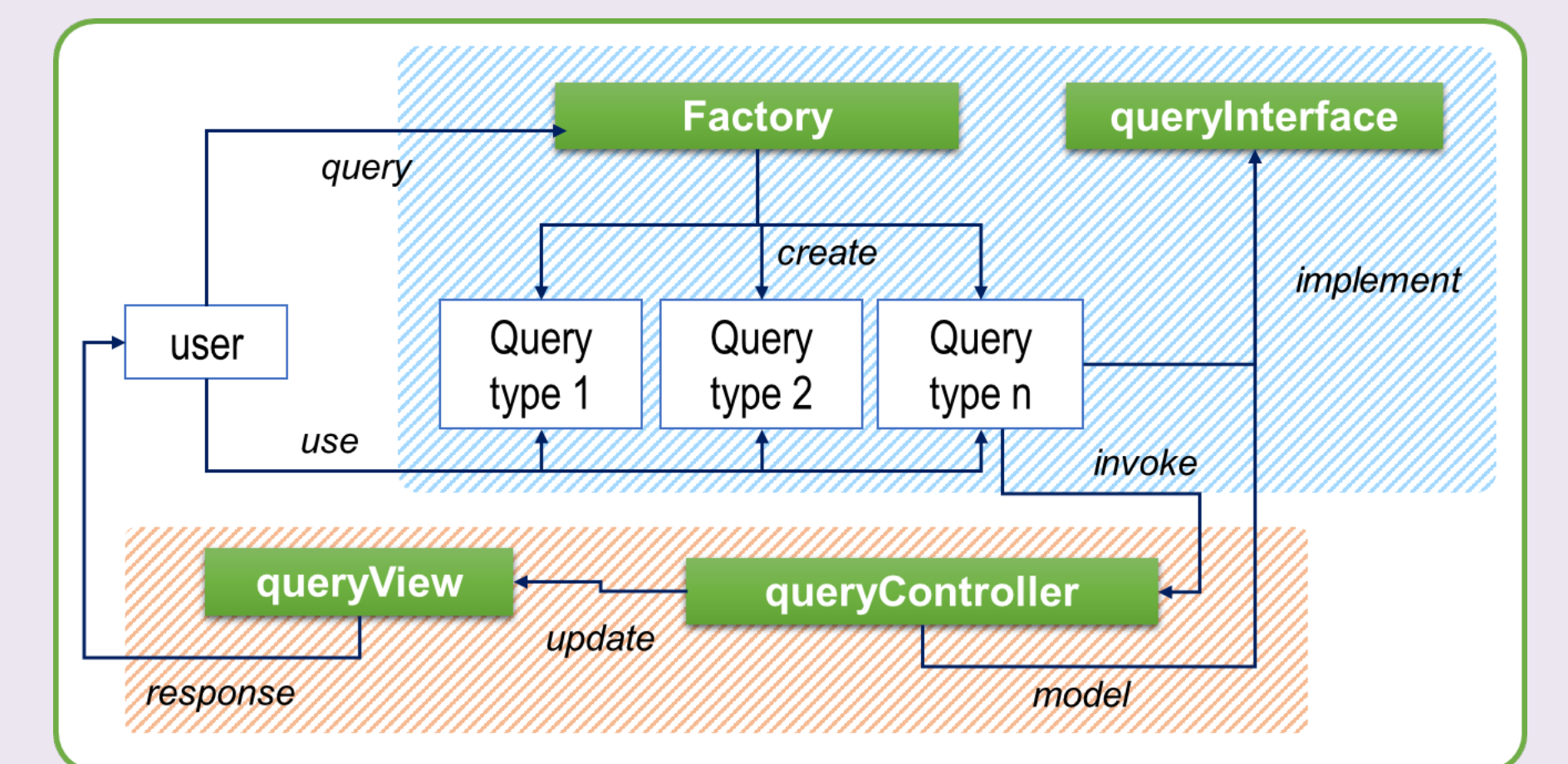


Figure 4. Classes diagram for RS-01 service.

Figure 4 shows classes and patterns design integrated into RS-01 service. The low level design is described in [2].

## 3.3 Population of Onto4AIR ontology

Some instances were added to make tests in Onto4AIR ontology, modeling relationships between them.

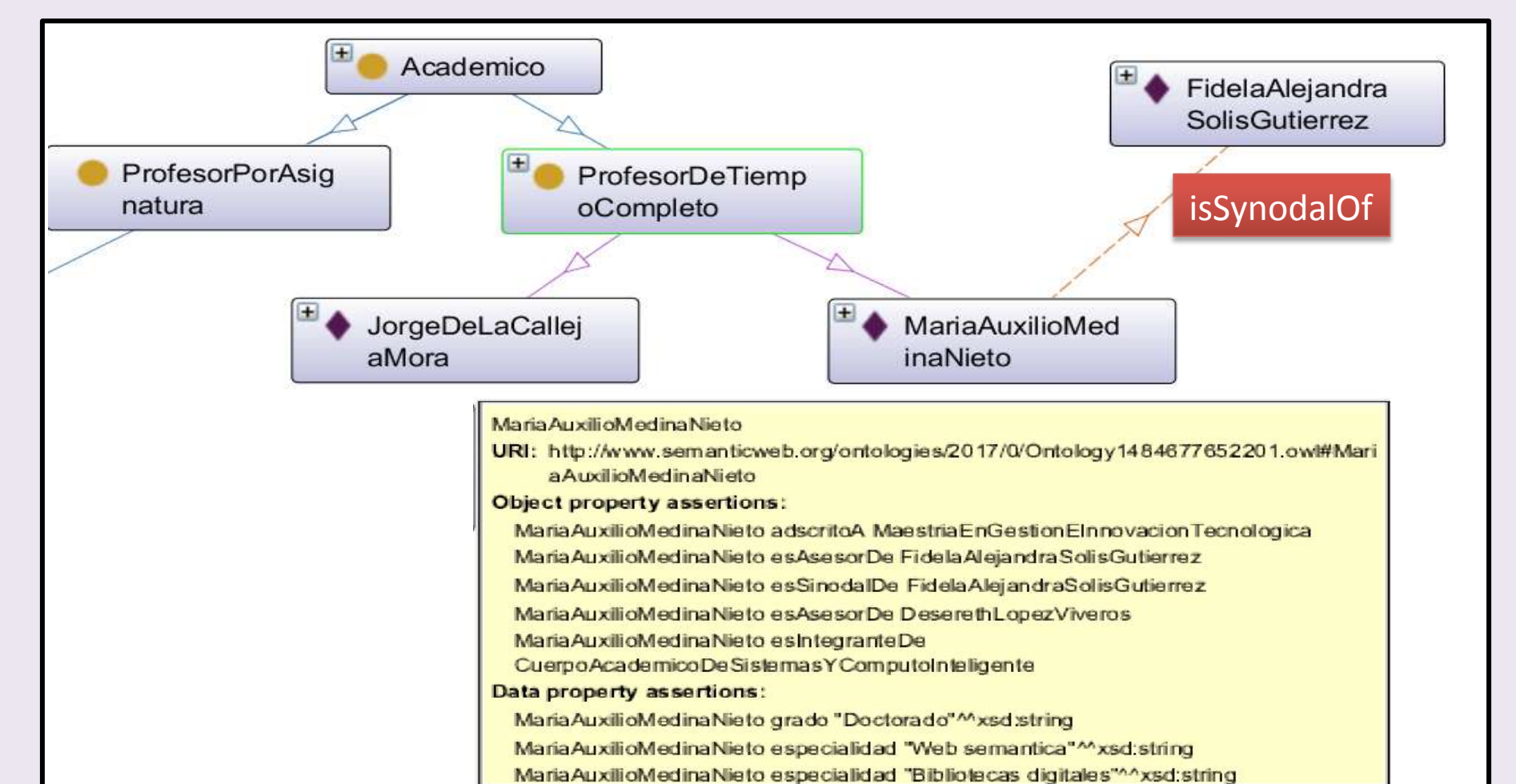


Figure 5. Instances relationships graph.

## 4. Results

Figure 6 shows a basic treatment of JSON string converted in a list structure in Python to make insertion of semantic information.

```
# Libraries
import requests
import json
# HTTP request to RS-01 type GET:
r = requests.get("http://repositorio.uppuebla.edu.mx:8080/webresources/autor")
# if HTTP state is 200 (OK):
if r.status_code == 200:
    # JSON string to list conversion
    data = json.loads(r.text)
    data.extend(lista)
    for element in data:
        # Data integration of synodals extract from RI-UPPue making comparisons between dc:Creator and "author" instance of Onto4AIR
```

Figure 6. Reduce code of web service RS-01.

## 5. Conclusions

Insertion of semantic information in two IRs increases the possibility of reuse, since it extends the current search services and allows to validate the logical consistency of the data automatically. In the short term, it is expected to integrate a catalog of semantic information retrieval services in the RI-UPPue, so that it distributes the data as linked open data.

## Acknowledgement

To CONACYT for being a scholarship holder of this academic program (scholarship number: 863914) and to all the people who support this project, including my family.

## References

- [1] Medina N. M. A., S. J. (2017). Semantic representation of operational and domain knowledge for institutional repositories. Copyright public record 03-2017-042511235500-01.
- [2] Vázquez M. P. D, Medina N. M. A., De la Calleja M. J., Benítez R. A., Vidal T. M., Alanís U. J. S. (2018). Web service design for extraction of semantic information from RI-UPPue. SOMI-UNAM, Instrumentation Congress, Torreon Coah. ISSN 2395-8499
- [3] CONACYT, (2019). National repository. Available in: <http://catalogs.repositorionacionalcti.mx/> [Acceded: March 30th 2019].
- [4] CONACYT, (2019). REST services catalog. Available in: <http://catalogs.repositorionacionalcti.mx/> [Acceded: March 30th 2019].





Este material se distribuye bajo los términos de la  
Licencia *Creative Commons* CC BY-NC-ND 2.5 MX

2019