

FONDO EUROPEO DE DESARROLLO
REGIONAL (FEDER)
Una manera de hacer Europa

Unión Europea

ABSTRACT

This document aims to summarise the current status of the Hercules project from the perspective of the development and integration of ontologies in the semantic network. A general description is first provided of the general Hércules project and its specific projects. Secondly, the OISA project and its network of ontologies are defined: what these ontologies are like, how they have been developed and where they can be found.

First of all, within all the applications of the Semantic Web is the use of ontologies, defined as a formal and explicit specification of a shared conceptualization, where by **conceptualization** we understand an abstract model of a world phenomenon, in which the relevant concepts identified for those phenomena. By **explicit**, we mean that the type of concepts used, and the restrictions on their use are explicitly defined. **Formal** refers to the fact that the ontology must be able to be interpreted by a machine. **Shared** reflects that the ontology must capture an accepted and consensual knowledge by a community.

HÉRCULES – “Semantics of University Research Data”

The project "Hercules - Semantics of University Research Data" is a CRUE initiative (Conference of Rectors of Spanish universities) with the aim of creating a Research Management System (RMS) based on open semantic data that offers a global vision of the data of research of the Spanish University System (SUS), to improve management, analysis and possible synergies between universities and the general public.

The project, in turn, is divided into 4 main projects with the following scope:

- **OISA Project:** Spanish University System Semantic Data Architecture and Information Ontological Infrastructure.

The Project aims to develop an efficient platform to store, manage and publish SGI data, based on the Ontological Infrastructure, with the ability to synchronize instances installed in different Universities. And the creation of a network of ontologies that can be used to describe the data of the research domain with fidelity and high granularity.

- **RMS Project:** Research Management System.

The project aims to develop an innovative prototype of a comprehensive research management system that can be integrated into any CRUE university. It will include all management activities related to research and must be integrated with the semantic architecture developed.

- **DEAM Project:** Data Enrichment from Internet sources and Development of Semantic Analysis Methods.

FONDO EUROPEO DE DESARROLLO
REGIONAL (FEDER)

Una manera de hacer Europa

The project aims to identify, extract, analyze and evaluate relevant data sets. Also the completion of data from external data sources.

- **RPA Project:** Robotic process automation.

The project aims to automate tasks and processes related to the field of university research management, using RPA technologies and artificial intelligence.

OISA Project

The project encompassed within the Hercules initiative in charge of developing an R&D service related to ontological infrastructure and the challenge of semantic architecture has been OISA. Given that the other projects that constitute HERCULES depend on both the Ontological Infrastructure and the Semantic Data Architecture, it is within the scope to interact with the other projects: development of an Innovative Prototype of Research Management for Universities and Enrichment of Data and Methods of Analysis.

The project has concluded, and it has been structured in two eliminatory phases:

- **Phase 1:** during this phase, completed during year 2019, the suitability of the proposed scientific-technological solutions, as well as the R&D plan to the challenge were raised, and the contractors verified the viability through a feasibility study. Two contractors were selected from this process to access Phase 2.
- **Phase 2:** It ended in May 2021 with the delivery of a prototype of the solution that allows the University of Murcia to benefit from an ontological infrastructure that describes the data that will be stored in the RMS. During this phase, R&D developments were carried out, the most relevant are listed below:
 - Ontology network, specification.
 - FAIR method analysis
 - RMS backend
 - Libraries (RDF validation, conversion, discovery and publishing)

Hercules Ontology Network

The Hercules Ontology Network (ROH) is divided into 2 main parts:

- **The generic ontology, core module** - contains the most important entities and properties to model information in the academic domain.
- **A set of vertical modules** which include, on one hand, specializations of some academic concepts for a given country domain and, on the other hand, controlled vocabularies, according to SKOS ontology, for different important areas in the academic domain.

Each of these ontological parts generates an automatic documentation, through the widoco tool, and that can be checked in the following links:

FONDO EUROPEO DE DESARROLLO
REGIONAL (FEDER)

Una manera de hacer Europa

- Widoco generated documentation for [ROH core module](#)
- Widoco generated documentation for [ROH vertical modules](#)

In Hercules, the ontology design process aims to model a Research Management System (RMS) for Spanish universities that is piloted and validated at the University of Murcia (UM). To achieve this, the following steps have been carried out:

1. Meet the requirements defined in "[Annex I. Ontology requirements analysis](#)" and "[Annex II. Ontologies and other resources to be used](#)". Delivered by GNOSS-DEUSTO as part of the feasibility study for “R&D service for the development of the ontological infrastructure and semantic architecture of the Research Management System (RMS) of the Hercules initiative”, file number: 2018/88/OT-AM
2. Selection and analysis of the main ontologies that model the academic environment. Including contrast with [CERIF - ERD](#) vocabulary, entity relationship diagram, since it is the standard information model for CRIS (Current Research Information System) systems.
3. Identification of the main entities and relationships to model the knowledge of the academic world. Fulfilling the requirements of the Semantic Architecture and Ontological Infrastructure (OISA) project.
4. Validation of the flexibility, completeness and integrity of the ROH ontology network through the following evaluations:
 - Review the [questions/competency queries](#) of the network listed by the UM and implement as a [suite of SPARQL queries](#) to validate their compliance. As a result of this validation, some new data and object properties were added.
 - Review datasets offered by the UM and check that their data can be modelled with the entities and properties defined within ROH. Extension and adaptation of ROH entities according to the determined needs.
 - [Mapping of FECYT CVN format data to the ROH ontology](#). Where there were unmodelled entities or relationships, they were included. Details of the mapping between CVN and ROH entities appear in the cvn/config folder.
5. Continuous refinement validated by a Continuous Integration (CI) process. A battery of regression tests regulate that new changes introduced continue to guarantee the quality of ROH, its flexibility and extensibility.

The ontology network (Hercules Ontology Network) is available in the Project's GITHUB repository <https://github.com/HerculesCRUE/ROH>

FONDO EUROPEO DE DESARROLLO
REGIONAL (FEDER)

Una manera de hacer Europa

CONCLUSION

The OISA Project has been successfully completed since the objectives have been achieved, specifically and in what is described here, an ontological infrastructure has been developed (ROH) that describes the data that the Research management system (RMS) will store.

The ontology network is available in the Project's GITHUB repository.

The rest of the projects within the HERCULES Project framework, Research Management System (RMS), Data Enrichment from Internet sources and Development of Semantic Analysis Methods (DEAM) and Robot Process Automation (RPA) are in the execution period, they will be available on github at the end of 2022.