



**ESCUELA SUPERIOR DE INGENIERÍA**  
**INGENIERÍA TÉCNICA INFORMÁTICA DE SISTEMAS**

MEMORIA ACADÉMICA UCA LINKED DATA

Lucía Batista Flores

12 de septiembre de 2014





ESCUELA SUPERIOR DE INGENIERÍA

INGENIERÍA TÉCNICA INFORMÁTICA DE SISTEMAS

MEMORIA ACADÉMICA UCA LINKED DATA

- Departamento: Ingeniería Informática
- Director del proyecto: Iván Ruiz Rube
- Autor del proyecto: Lucía Batista Flores

Cádiz, 12 de septiembre de 2014

Fdo: Lucía Batista Flores



Este documento se halla bajo la licencia FDL (Free Documentation License). Según estipula la licencia, se muestra aquí el aviso de copyright. Se ha usado la versión inglesa de la licencia, al ser la única reconocida oficialmente por la FSF (Free Software Foundation).

Copyright ©2014 Lucía Batista Flores.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

## ***Agradecimientos***

*Quisiera darle las gracias a Don Manuel Palomo y a Don Juan Manuel Dodero por darme esta idea tan original para mi proyecto de la Ingeniería Técnica Informática de Sistemas, ya que ellos fueron los primeros en hablar conmigo para ver cuál era mi opinión; a Don Iván Ruiz mi Director de Proyecto, y profesor que me ha ayudado en el desarrollo y corrección del mismo.*

*Gracias a los tres por tener paciencia y responderme a los sucesivos correos con mis dudas, además de permitirme participar en el congreso MSTR2012 donde aprendí más cosas que me van a ser muy útiles para este proyecto y me lleve una gratificante experiencia conociendo gente nueva. Por último quisiera darle las gracias a mis padres, a mi tía M<sup>a</sup> Angeles y a mis abuelos, además del resto de mi familia, amigos y compañeros de la ESI por el continuo apoyo y cariño que he recibido año tras año durante mi etapa de estudiante.*

## Resumen

“Memoria Académica UCA Linked Data” es un Proyecto de Fin de Carrera sobre las posibilidades que ofrece el método de publicación de datos entrelazados y el concepto de datos abiertos. El presente proyecto consta de las siguientes partes: en primer lugar, un análisis de la información actualmente presentada en las diferentes versiones de la memoria académica de la Universidad de Cádiz, para realizar una base de datos que contenga dicha información almacenada de forma estructurada.

Para este proyecto se tuvo acceso a un servidor de la UCA donde se creó una base de datos, a la que se aplica una ontología que ha sido previamente desarrollada y que permite realizar consultas en el *endpoint* o punto de consulta de información haciendo uso del lenguaje SPARQL. Finalmente, se crea una instancia de un portal que permite almacenar datos y ser reutilizados en el futuro para ir creando los distintos dataset que se necesiten y así fomentar el concepto de datos abiertos, además de una aplicación que hará uso del *endpoint* anterior permitiendo no sólo visualizar los datos de forma gráfica sino también algunos ejemplos que permiten comparar ficheros externos de otras universidades entre sí.

**Palabras clave:** Linked Data, Open data, D2RQ, CKAN, SPARQL, Base de datos, Endpoint, Ontologías, Conocimiento, RDF, OWL.





# Índice general

<b>I Prolegómeno</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
1.1. Motivación . . . . .	3
1.2. Descripción del sistema actual . . . . .	4
1.3. Objetivos y alcance del proyecto . . . . .	5
1.4. Organización del documento . . . . .	5
<b>2. Planificación</b>	<b>7</b>
2.1. Metodología de desarrollo . . . . .	7
2.2. Planificación del proyecto . . . . .	8
2.2.1. Iteraciones . . . . .	10
2.3. Organización . . . . .	11
2.4. Herramientas utilizadas . . . . .	11
2.4.1. Cacao . . . . .	11
2.4.2. ShareLatex . . . . .	11
2.4.3. Eclipse . . . . .	11
2.4.4. Subversion . . . . .	12
2.4.5. OpenRefine . . . . .	13
2.4.6. Pentaho Data Integration (PDI) . . . . .	13
2.5. Costes . . . . .	14
2.6. Gestión de riesgos . . . . .	15
2.6.1. Limitaciones de información . . . . .	15
<b>II Desarrollo</b>	<b>17</b>
<b>3. Estado del arte</b>	<b>19</b>
3.1. Conceptos básicos . . . . .	19
3.1.1. Open government . . . . .	19
3.1.2. Open data . . . . .	20
3.1.3. Estándares Linked Data . . . . .	20
3.1.4. Clasificación de fuentes Open Data . . . . .	21
3.1.5. Web Semántica . . . . .	22
3.1.6. Tripletas . . . . .	23
3.1.7. Linked Data . . . . .	24
3.1.8. SPARQL . . . . .	24
3.2. Otros proyectos ya desarrollados en Linked Data . . . . .	25

3.2.1.	Apertura de datos en instituciones universitarias . . . . .	28
<b>4.</b>	<b>Requisitos del sistema</b>	<b>31</b>
4.1.	Situación actual . . . . .	31
4.2.	Requisitos funcionales . . . . .	32
4.3.	Requisitos no funcionales . . . . .	32
4.4.	Requisitos de información . . . . .	33
<b>5.</b>	<b>Análisis</b>	<b>35</b>
5.1.	Casos de uso . . . . .	35
5.1.1.	Caso de uso: Añadir dataset. . . . .	35
5.1.2.	Caso de uso: Buscar un dataset. . . . .	35
5.1.3.	Caso de uso: Descargar un dataset. . . . .	36
5.1.4.	Caso de uso: Consultar de forma programática los datos. . . . .	36
5.1.5.	Caso de uso: Generar gráfica . . . . .	36
5.1.6.	Caso de uso: Comparar dos universidades . . . . .	36
5.1.7.	Diagrama de Secuencia . . . . .	37
5.2.	Modelo de interfaz de usuario . . . . .	38
5.2.1.	Boceto de la aplicación web . . . . .	38
<b>6.</b>	<b>Ingeniería Ontológica</b>	<b>41</b>
6.1.	¿Qué es una ontología y qué razones podemos tener para implementarla? . . . . .	41
6.2.	Diseño de vocabularios . . . . .	42
6.2.1.	DBpedia Ontology . . . . .	42
6.2.2.	The Friend of a Friend (FOAF) project . . . . .	43
6.2.3.	Dublin Core Metadata . . . . .	43
6.2.4.	Teaching Core Vocabulary . . . . .	43
6.2.5.	CERIF Semantic Vocabulary 1.3 . . . . .	43
6.2.6.	Academic Research Project Funding Ontology (ARPFO) . . . . .	44
6.2.7.	An Agent Certification Ontology . . . . .	44
6.2.8.	The Lingvoj Ontology . . . . .	44
6.2.9.	Neologism . . . . .	44
6.3.	Vocabularios desarrollados . . . . .	45
6.3.1.	Vocabulary about University Services . . . . .	46
6.3.2.	Vocabulary about the University Research . . . . .	49
6.3.3.	Vocabulary for University Infrastructure . . . . .	51
6.3.4.	Vocabulary for University Staff . . . . .	52
6.3.5.	Vocabulary for University Funding . . . . .	55
6.3.6.	Vocabulary for University Students and University Knowledge . . . . .	57
6.3.7.	Vocabulary for Energy Efficiency and Primary Resources . . . . .	61
<b>7.</b>	<b>Diseño del Sistema</b>	<b>63</b>
7.1.	Arquitectura . . . . .	63
7.1.1.	Arquitectura física . . . . .	63
7.1.2.	Arquitectura lógica . . . . .	64
7.1.3.	MySQL . . . . .	65
7.2.	Procesos ETL . . . . .	65
7.3.	Almacén de datos . . . . .	66

7.4. Publicación de datos RDF . . . . .	105
7.5. Aplicación de consulta de datos . . . . .	105
7.5.1. Diseño de la interfaz de usuario . . . . .	105
7.5.2. Arquitectura de diseño . . . . .	108
<b>8. Implementación del Sistema . . . . .</b>	<b>109</b>
8.1. Entorno tecnológico . . . . .	109
8.1.1. Interfaz con el usuario . . . . .	109
8.1.2. Lógica de servidor . . . . .	111
8.1.3. D2R-Server . . . . .	112
8.2. Código fuente . . . . .	113
8.2.1. Estructura de una aplicación Grails . . . . .	114
8.2.2. Scripts auxiliares de transformación de datos . . . . .	114
8.3. Procesos ETL . . . . .	115
<b>9. Pruebas del Sistema . . . . .</b>	<b>119</b>
9.1. Pruebas unitarias y de integración . . . . .	119
9.2. Pruebas de sistema . . . . .	120
9.2.1. Pruebas funcionales . . . . .	121
9.2.2. Pruebas no funcionales . . . . .	121
9.3. Pruebas de aceptación . . . . .	124
9.4. Caso de estudio . . . . .	125
9.4.1. Universidad Pablo de Olavide . . . . .	125
9.4.2. Comparación de la UPO con la UCA . . . . .	126
<b>III Epílogo . . . . .</b>	<b>129</b>
<b>10. Conclusiones . . . . .</b>	<b>131</b>
10.1. Objetivos cumplidos . . . . .	131
10.2. Lecciones aprendidas . . . . .	132
10.3. Trabajo futuro . . . . .	132
<b>IV Apéndice . . . . .</b>	<b>135</b>
<b>A. Manual de instalación y explotación . . . . .</b>	<b>137</b>
A.1. Introducción . . . . .	137
A.2. Requisitos previos . . . . .	137
A.3. Inventario de componentes . . . . .	137
A.4. Procedimientos de instalación . . . . .	138
A.4.1. D2RQ . . . . .	138
<b>B. Manual de usuario . . . . .</b>	<b>141</b>
B.1. Introducción . . . . .	141
B.2. Requisitos previos . . . . .	141
B.3. Utilización . . . . .	141
B.3.1. CKAN . . . . .	142
B.3.2. D2R Server . . . . .	143

B.3.3. Navegar el catálogo de datos . . . . .	144
B.3.4. Ejecutar una consulta SPARQL . . . . .	144
B.3.5. Aplicación Web . . . . .	145
<b>C. GNU Free Documentation License</b>	<b>149</b>
1. APPLICABILITY AND DEFINITIONS . . . . .	149
2. VERBATIM COPYING . . . . .	151
3. COPYING IN QUANTITY . . . . .	151
4. MODIFICATIONS . . . . .	151
5. COMBINING DOCUMENTS . . . . .	153
6. COLLECTIONS OF DOCUMENTS . . . . .	153
7. AGGREGATION WITH INDEPENDENT WORKS . . . . .	153
8. TRANSLATION . . . . .	154
9. TERMINATION . . . . .	154
10. FUTURE REVISIONS OF THIS LICENSE . . . . .	154
11. RELICENSING . . . . .	155
ADDENDUM: How to use this License for your documents . . . . .	155

# Índice de figuras

2.1. Modelo de ciclo de vida incremental . . . . .	7
2.2. Diagrama de estimación temporal . . . . .	9
2.3. Entorno de programación . . . . .	12
2.4. Interacción con el repositorio . . . . .	13
3.1. Capas de la Web Semántica . . . . .	23
3.2. Tripletas RDF . . . . .	24
3.3. LOD 2007 . . . . .	27
3.4. LOD 2011 . . . . .	27
5.1. Casos de uso . . . . .	37
5.2. Diagrama de Secuencia del caso de uso Generar Gráfica . . . . .	38
5.3. Diagrama de Secuencia del caso de uso Comparativa . . . . .	38
5.4. Boceto del Menu de la aplicación . . . . .	39
5.5. Boceto del Formulario de la aplicación . . . . .	40
6.1. Menu principal de neologism . . . . .	45
6.2. Vocabulario de Servicios Universitarios . . . . .	46
6.3. Vocabulario de Investigación Universitaria . . . . .	49
6.4. Vocabulario de Infraestructura Universitaria . . . . .	51
6.5. Vocabulario sobre el Personal Universitario . . . . .	52
6.6. Vocabulario de Financiación Universitaria . . . . .	55
6.7. Vocabulario sobre Alumnos y Conocimientos Universitarios . . . . .	57
6.8. Vocabulario de Eficiencia y Recursos primarios . . . . .	61
7.1. Vista general del sistema Linked Data . . . . .	63
7.2. Diagrama del Proceso Completo . . . . .	66
7.3. Prototipo de la aplicación . . . . .	106
7.4. Menu principal . . . . .	106
7.5. Formulario de Consulta . . . . .	107
7.6. Formulario de Comparación . . . . .	107
8.1. Estructura framework Grails . . . . .	112
8.2. Limpieza de datos con Open Refine . . . . .	115
8.3. Configuración de la base de datos . . . . .	116
8.4. Ejemplo de transformación . . . . .	117
8.5. Trabajo 1 . . . . .	117
8.6. Base de datos de infraestructura intermedia . . . . .	118

8.7. Trabajo 2 . . . . .	118
9.1. Datos de la Memoria Académica 2012-2013 . . . . .	119
9.2. Gráfico de la Aplicación . . . . .	120
9.3. Gráfico comparativo de tiempo de carga durante la prueba de estrés . . . . .	124
9.4. Conjunto de datos de la Universidad Pablo de Olavide . . . . .	125
9.5. Comparativa de Presupuestos de la UCA con la UPO en barras . . . . .	126
9.6. Comparativa de Presupuestos de la UCA con la UPO en líneas . . . . .	127
B.1. Creación de dataset . . . . .	143
B.2. Descarga de un recurso . . . . .	143
B.3. Página principal de D2R Server . . . . .	144
B.4. Interfaz Snorql de consulta . . . . .	145
B.5. Menu principal de la aplicación . . . . .	146
B.6. Formulario para generar la gráfica . . . . .	147

# Índice de cuadros

2.1. Tabla de Costes . . . . .	14
7.1. Acceso universitario . . . . .	67
7.2. Acciones Plan Social . . . . .	67
7.3. Actividades de comunicación por curso . . . . .	67
7.4. Actividades de idiomas . . . . .	68
7.5. Actuaciones del defensor universitario . . . . .	68
7.6. Alumnos extranjeros por curso . . . . .	68
7.7. Alumnos graduados . . . . .	69
7.8. Alumnos matriculados . . . . .	69
7.9. Alumnos de nuevo ingreso . . . . .	69
7.10. Alumnos por actividad . . . . .	70
7.11. Alumnos por campus . . . . .	70
7.12. Alumnos por centros . . . . .	70
7.13. Alumnos de tiempo completo . . . . .	71
7.14. Areas del PAIDI . . . . .	71
7.15. Aula de mayores matriculados . . . . .	71
7.16. Bau . . . . .	72
7.17. Becas de movilidad . . . . .	72
7.18. Beneficiarios de incentivos de jubilación . . . . .	72
7.19. Campus . . . . .	73
7.20. Aula de mayores matriculados . . . . .	73
7.21. Cátedras de Empresa . . . . .	73
7.22. Cau . . . . .	74
7.23. Clasificación por edad de alumnos . . . . .	74
7.24. Clasificación por edad de empleados . . . . .	74
7.25. Compromiso social por curso . . . . .	75
7.26. Contratos con empresas . . . . .	75
7.27. Convenios . . . . .	75
7.28. Curso académico . . . . .	76
7.29. Cursos de idiomas . . . . .	76
7.30. Cursos de prevención de riesgos laborales . . . . .	76
7.31. Difusión educativa . . . . .	77
7.32. Doctorados por curso . . . . .	77
7.33. Doctorados . . . . .	77
7.34. Eliminación de residuos . . . . .	78
7.35. Empresas vinculadas . . . . .	78

7.36. Enfermedades laborales . . . . .	78
7.37. Equipamientos informáticos . . . . .	79
7.38. Estructuras organizativas . . . . .	79
7.39. Evolución de ayudas a planes propios . . . . .	79
7.40. Evolución de la investigación . . . . .	80
7.41. Evolución de las mujeres . . . . .	80
7.42. Fondos de proyectos de convocatorias públicas . . . . .	80
7.43. Formación del PAS . . . . .	81
7.44. Formación del PDI . . . . .	81
7.45. fuentes primarias no renovables . . . . .	81
7.46. gastos e ingresos . . . . .	82
7.47. Grupos del PAS . . . . .	82
7.48. Horas de formación de plantilla del pas por sexo . . . . .	82
7.49. Horas de formación de la plantilla del pdi por sexo . . . . .	83
7.50. Igualdad entre sexos . . . . .	83
7.51. Indicadores de deporte . . . . .	83
7.52. Inserción laboral por curso . . . . .	84
7.53. Medidas de conciliación con la vida laboral . . . . .	84
7.54. Nuevas instalaciones . . . . .	84
7.55. Número de actividades . . . . .	85
7.56. Oferta de grupos de investigación . . . . .	85
7.57. Ofertas de titulación por cursos . . . . .	85
7.58. Origen de financiación . . . . .	86
7.59. residuos del papel . . . . .	86
7.60. Tabla del Pas . . . . .	86
7.61. Perfil de la Universidad . . . . .	87
7.62. Prácticas de empresa . . . . .	87
7.63. Premios y reconocimientos . . . . .	87
7.64. Presupuesto . . . . .	88
7.65. Profesorado . . . . .	88
7.66. Programas de salud laboral . . . . .	88
7.67. Programa Erasmus . . . . .	89
7.68. Protección de resultados de investigación . . . . .	89
7.69. Proyectos concedidos de convocatorias públicas . . . . .	89
7.70. Proyectos docentes . . . . .	90
7.71. Residuos eléctricos y electrónicos . . . . .	90
7.72. Ramas de conocimientos . . . . .	90
7.73. Recursos de investigación . . . . .	91
7.74. Rendimiento por curso . . . . .	91
7.75. Retribución mínima . . . . .	91
7.76. Satisfacción con servicios . . . . .	92
7.77. Satisfacción educativa por curso . . . . .	92
7.78. Servicios de biblioteca . . . . .	92
7.79. Servicios de Información . . . . .	93
7.80. Tesis Doctorales . . . . .	93
7.81. Tipos de accesos universitarios . . . . .	93
7.82. Tipos de acciones sociales . . . . .	94



7.83. Tipos de actividades de comunicación . . . . .	94
7.84. Tipos de alumnos extranjeros . . . . .	94
7.85. Tipos de ayudas a planes propios de investigación . . . . .	95
7.86. Tipos de bau . . . . .	95
7.87. Tipos de becas o contrato de investigación . . . . .	95
7.88. Tipos de becas de movilidad . . . . .	96
7.89. Tipos de compromiso social . . . . .	96
7.90. Tipos de convenios . . . . .	96
7.91. Tipos de difusión educativa . . . . .	97
7.92. Tipos de doctorados . . . . .	97
7.93. Tipos de enfermedades laborales . . . . .	97
7.94. Tipos de figura contratada . . . . .	98
7.95. Tipos de financiación . . . . .	98
7.96. Tipos de formación del PAS . . . . .	98
7.97. Tipos de formación del PDI . . . . .	99
7.98. Tipos de gastos e ingresos . . . . .	99
7.99. Tipos de idiomas . . . . .	99
7.100Tipos de igualdad entre sexos . . . . .	100
7.101Tipos de indicadores de deporte . . . . .	100
7.102Tipos de inserción laboral . . . . .	100
7.103Tipos de medidas de conciliación con vida laboral . . . . .	101
7.104Tipo de modalidades formativas . . . . .	101
7.105Tipos de protección de resultados de investigación . . . . .	101
7.106Tipos de proyectos de convocatorias públicas . . . . .	102
7.107Tipos de recursos de investigación . . . . .	102
7.108Tipos de rendimiento . . . . .	102
7.109Tipos de residuos eliminados . . . . .	103
7.110Tipos de satisfacción educativa . . . . .	103
7.111Tipos de servicios universitarios . . . . .	103
7.112Tipo de titulaciones . . . . .	104
7.113Titulaciones por rama y campus . . . . .	104



Parte I  
Prolegómeno



# Capítulo 1

## Introducción

A continuación se describe la motivación del proyecto, la descripción del sistema actual, los objetivos y la estructura del presente documento.

### 1.1. Motivación

La Memoria de la Universidad de Cádiz es el documento donde anualmente se recogen los datos de cada año académico y ordenan sistemáticamente. Hasta el curso académico 2009-2010, la Universidad publicaba dicha memoria en un documento PDF de más de 200 páginas con toda la información. En el curso académico 2010-2011, se decidió publicarla en un sitio web HTML con un dominio perteneciente a la Universidad. Este formato facilita el acceso a la información de una forma más rápida y con menor grado de dificultad para realizar búsquedas específicas en los datos.

Con el objetivo de mejorar la publicación de estos datos en la web que se encuentra ahora mismo, se desean aplicar las técnicas y tecnologías de Linked Data para acercarnos un poco más al concepto de Web Semántica, donde los datos están relacionados y se pueden realizar búsquedas por el significado.

La idea de este proyecto surge a raíz de la doctrina política Open Government, o también traducida como Gobierno Abierto, que sostiene el hecho de que tanto el gobierno como todas las administraciones públicas deben ser abiertos a todos los niveles posibles de transparencia, para así fomentar el debate público sobre la sociedad civil. Para esto debemos aplicar la filosofía del movimiento de software libre, que no es más que sustituir aquel software con licencia privativa por otros con licencia libre permitiendo al usuario mayor libertad.

En este proyecto se desea hacer uso de este nuevo enfoque para añadir a la Universidad de Cádiz en la lista de aquellas administraciones públicas que tenga sus datos publicados de tal forma que cumplan esta nueva doctrina política y así ser una de las primeras Universidades de España que lleva a cabo de este cometido, puesto que existen otras universidades del mundo en las que se han realizado proyectos de este tipo y están teniendo una gran influencia.

## 1.2. Descripción del sistema actual

Hasta el momento, el sistema parte de la información que tenemos en la página web [10]. En dicha página disponemos de la memoria académica del curso que se ha finalizado, en este caso 2012-2013, en formato web usando un conjunto de páginas estáticas HTML, sin ningún tipo de interactividad o dinamismo a excepción de un buscador, junto con las memorias de los años anteriores en formato PDF, aunque algunas de años anteriores también se conservan en formato HTML.

Desde el inicio de este proyecto, en el año 2012, la Universidad de Cádiz ha realizado otros esfuerzos adicionales de transparencia y apertura de datos, materializándose en iniciativas como la presentación pública de Presupuestos para el presente año 2014<sup>1</sup> o el Portal de Transparencia<sup>2</sup>, donde se publica o facilita la solicitud de diversos datos que pudieran resultar de interés a la ciudadanía en general, en el mismo formato online o PDF que la memoria académica.

Actualmente, a pesar de que los datos de dicha memoria anual están expuestos con licencia libre y es posible consultarlos, no es posible realizar con ellos ningún tipo de tratamiento automatizado tales como realizar comparaciones, consultas o intercambios entre instituciones usando un mismo esquema común. El inconveniente es que no se encuentra en ningún formato estructurado, tan sólo se ha llevado el texto de un formato (PDF) a otro (HTML), por lo que sigue siendo un conjunto de párrafos con información sólo legible para humanos (prosa), sin sentido alguno para una máquina, por lo tanto para el desarrollo del proyecto habrá que organizarlos.

En nuestros días la universidad cuenta con una web 2.0 que explicaremos a continuación. La universidad 2.0 es un término que surge cuando se hace uso de las tecnologías y herramientas web 2.0 (software social) en el seno de la universidad, creando un espacio de comunicación abierto entre los miembros de la comunidad universitaria (PDI, PAS, Alumnos y Gobierno) y los demás miembros de la sociedad para conseguir una universidad más social en la que cuyos integrantes puedan participar y colaborar de forma abierta.

Los objetivos esenciales de una universidad 2.0 son los siguientes:

- Crear un espacio de comunicación abierto en toda la comunidad universitaria y el conjunto de la sociedad.
- Fomentar los principios y actitudes de la filosofía web 2.0 en toda la comunidad universitaria y la sociedad.
- Establecer a la universidad como referente para la sociedad civil en la adopción y uso de las redes sociales.
- Promover la divulgación del conocimiento, los repositorios abiertos y alternativas de reputación científica, siguiendo el modelo Open Access.

El objetivo final de este proyecto es avanzar del estado actual de la publicación de la memoria actual en formato estático hacia la Web Semántica, aplicando la tecnología de Linked Data, a través de las herramientas que hemos usado para desarrollar este proyecto que nos permita dotar

---

<sup>1</sup><http://presupuestos.uca.es>

<sup>2</sup><http://transparencia.uca.es/>

de significado a la Web y permitir el tratamiento informatizado de datos, tanto por la propia Universidad como por cualquier otra persona o entidad.

### 1.3. Objetivos y alcance del proyecto

Los objetivos del desarrollo de la memoria académica en versión Linked Data son los siguientes:

- Publicar los datos de la memoria académica en la web bajo un formato libre. Un formato libre o abierto es aquel que puede ser usado por cualquier persona o entidad, puesto que está exento de restricciones legales en cuanto a su uso.
- Usar URIs y vincular los datos para obtener mejores búsquedas, cumpliendo los requisitos del modelo de 5 estrellas de Linked Data planteado por Berners-Lee [7].
- Datos reutilizables. Los datos académicos se pondrán a disposición de todos los usuarios, para su reutilización y generación de nuevos contenidos. Un ejemplo del uso estos datos puede ser crear gráficas con los datos que se desee comparar, estadísticas, etc.
- Desarrollar un portal de datos con la información de la memoria académica de la UCA.
- Incluir el dataset UCA al Linked Data Cloud.

### 1.4. Organización del documento

En esta sección, se tiene como objetivo describir los contenidos de la presente memoria, organizada en los siguientes capítulos:

- **Capítulo 1: Introducción.** Se realiza una introducción y motivación por la que se ha realizado el presente proyecto, además de especificar los objetivos y realizar un resumen de las distintas partes del documento.
- **Capítulo 2: Planificación.** Dentro de este apartado se especifica la metodología que se ha seguido para desarrollar el presente proyecto y la planificación a través del diagrama de Gantt.
- **Capítulo 3: Estado del arte.** Este apartado es el más importante, ya que recoge los conceptos previos que se han tenido que aprender para desarrollar el proyecto.
- **Capítulo 4: Requisitos del sistema.** Describe los requisitos que se necesitan para desarrollar el sistema.
- **Capítulo 5: Análisis.** Se incluyen los diagramas de caso de uso, diagramas de secuencia necesarios para el sistema total y se muestra un boceto de la aplicación
- **Capítulo 6: Ingeniería ontológica.** Se explica el concepto de ontología, las razones de su desarrollo, las buenas prácticas que se han seguido para desarrollar los vocabularios RDF, los términos reutilizados y los vocabularios que se desarrollaron con la descripción de clases y propiedades de cada uno de ellos.

- **Capítulo 7: Diseño del Sistema.** Se recoge la arquitectura general del sistema de información, el diseño de la interfaz de usuario, el diseño físico de datos y el diseño de componentes software.
- **Capítulo 8: Implementación del sistema.** Documentamos los detalles de implementación que son relevantes para el proyecto.
- **Capítulo 9: Pruebas.** Tras haber realizado la implementación y desarrollo del presente proyecto se documentan todas aquellas pruebas que se han realizado para verificar la calidad del producto.
- **Capítulo 10: Conclusiones.** Se refleja y valora los resultados que se han obtenido en el desarrollo del presente proyecto.
- **Apéndices.** Se divide en dos apartados:
  - El apéndice A es el **Manual de instalación y explotación del sistema.**
  - El apéndice B es el **Manual del usuario del sistema.**



## Capítulo 2

# Planificación

En esta sección se describen todos los aspectos relativos a la planificación del proyecto: metodología, organización, costes, planificación y gestión de riesgos.

### 2.1. Metodología de desarrollo

El desarrollo de este PFC se ha realizado de forma incremental. Como podemos ver en la figura 2.1, el modelo de planificación y desarrollo seguido se caracteriza por partir de una versión inicial del software con funcionalidad limitada, para evolucionarlo en sucesivas iteraciones, las cuales inician con el análisis de las especificaciones a implementar en dicha iteración, una segunda fase de diseño de los distintas partes del proyecto, una tercera fase de implementación y construcción y finalizan con una etapa de pruebas y la integración de las nuevas funcionalidades en el sistema.

Debido al carácter de investigación de este proyecto, y teniendo en cuenta las posibles aportaciones que se podrían realizar en un futuro en el sistema de la Universidad de Cádiz, la metodología usada ha permitido que se realizan correcciones y reaccionar de forma ágil ante las limitaciones que se han encontrado durante el proceso, como por ejemplo, el carácter sensible de algunos de los datos que se facilitan en este proyecto. Por lo tanto esta metodología es la más adecuada para analizar los requisitos que se desean, implementarlos y conseguir un resultado que posteriormente se prueba para mejorar hasta el final del desarrollo de este proyecto, pudiendo reaccionar a los cambios que ocurren durante el desarrollo e implantación con el menor retraso posible.

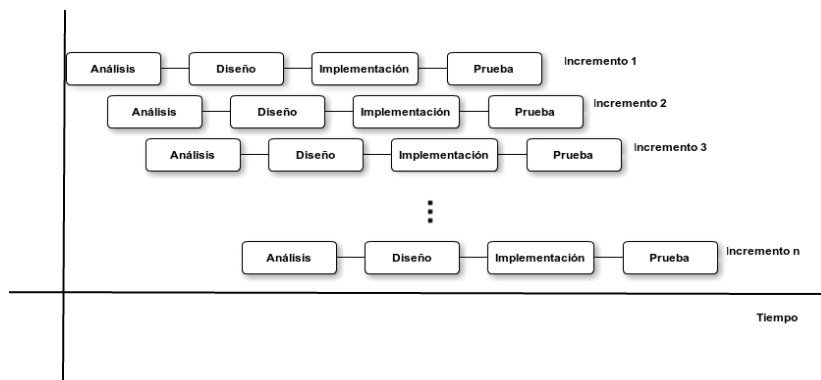


Figura 2.1: Modelo de ciclo de vida incremental

## 2.2. Planificación del proyecto

En la figura 2.2 se muestra la planificación del presente proyecto en forma de diagrama de Gantt, de forma que se represente de forma gráfica los diferentes hitos e iteraciones de cada fase, junto con su estimación temporal.

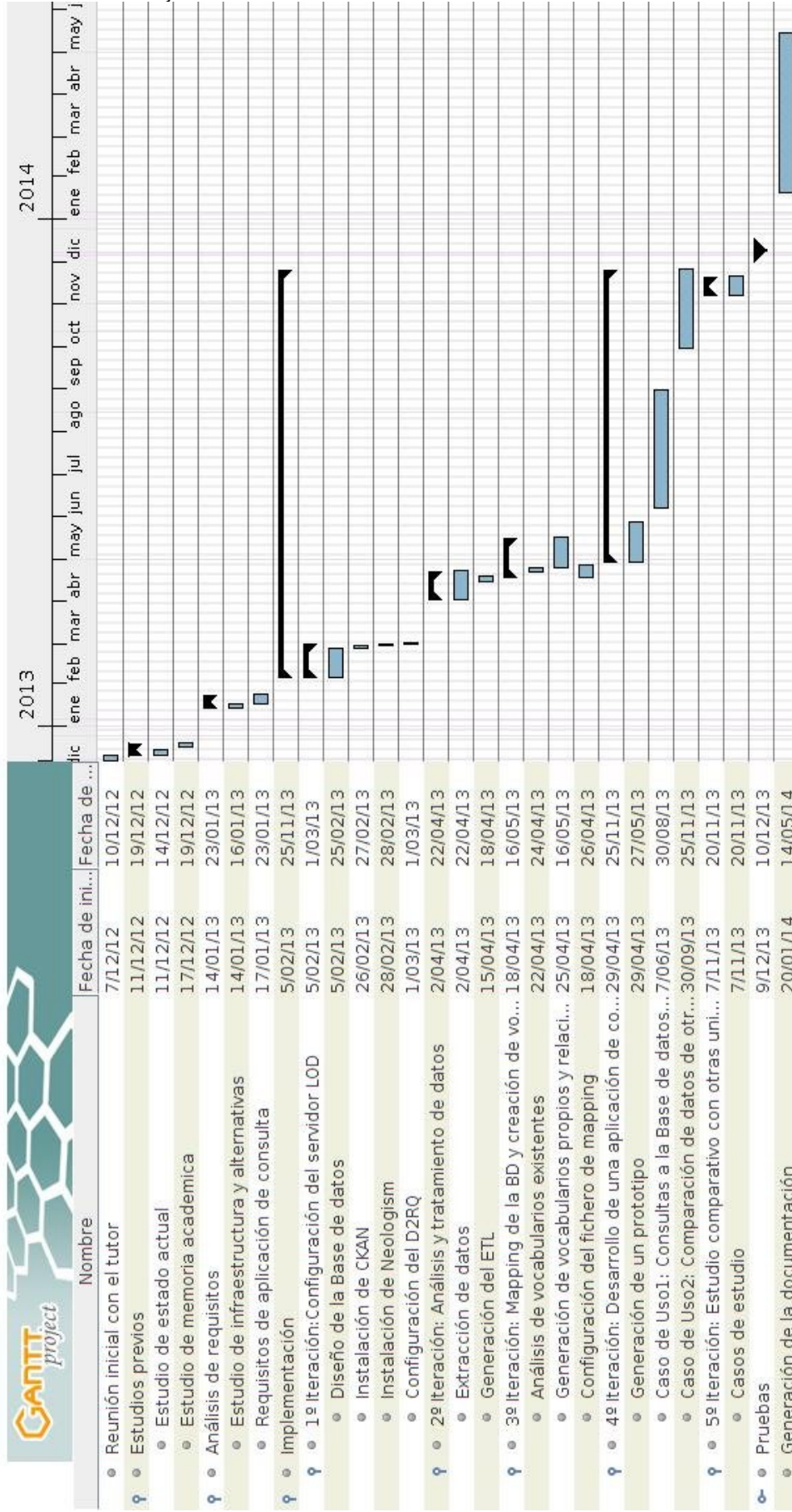


Figura 2.2: Diagrama de estimación temporal

### 2.2.1. Iteraciones

#### 1º Iteración: Diseño de infraestructura

En la primera etapa, se realizó una investigación inicial sobre los distintos sistemas Linked Data que ya hay disponibles y sobre como podemos adaptarlo a nuestra Universidad como entidad administrativa.

Además se hizo una búsqueda de herramientas y aplicaciones para aprender a publicar en Linked Data. Una vez decididas las herramientas que iban a utilizar, se configuró el servidor Linked Open Data (LOD). Se instalaron las aplicaciones que sirvieron como los cimientos principales para la arquitectura de información del proyecto: CKAN [1], como almacén de datos donde se publica el dataset de la memoria de la UCA y Neologism, como editor de ontologías necesario para crear los vocabularios RDF en la 3º iteración.

#### 2º Iteración: Análisis y tratamiento de datos

En esta iteración, con los almacenes creados, se realizó un análisis de los datos que hay disponibles en la memoria de la Universidad de Cádiz y se desarrolló de una base de datos con la idea de estructurar los datos disponibles. En esta fase se realizó la depuración de datos y corrección de los errores que tenían los ficheros CSV, de donde cargarían los datos de cada una de las tablas de la base de datos.

Además, se realizó un proceso ETL donde mostró cómo realizar una carga de datos directamente desde las bases de datos que recopilan toda la información de la memoria, la cual se encuentra repartida en diferentes lugares. Gracias a este proceso, se cargaron todos los datos directamente en la base de datos de la memoria de la UCA final, que es la que realmente está publicada en Linked Data.

#### 3º Iteración: Mapping de la BD y creación de vocabularios

Durante esta iteración se configuró el servidor D2RQ Server [8], cuya función principal es convertir datos estructurados SQL en RDF. Posteriormente, se diseñaron una serie de vocabularios RDF siguiendo unas directrices de buenas prácticas con el objetivo de que la Universidad de Cádiz o cualquier otra universidad puedan publicar sus datos siguiendo esquemas comunes.

Los distintos vocabularios se crearon haciendo uso del editor Neologism y posteriormente se realizaron las sustituciones pertinentes en el fichero de mapping de la aplicación del D2R Server, haciendo uso de un pequeño script en Python para evitar la tediosa labor manual.

#### 4º Iteración: Desarrollo de una aplicación de consumo

La penúltima fase de este proyecto consistió en realizar una aplicación que agilice la consultas de los datos de manera gráfica, cumpliendo así el objetivo más importante de Open Data y es que cualquier ciudadano pueda tener acceso a los datos de forma fácil y rápida.

#### 5º Iteración: Estudio comparativo con otra universidad

Esta última fase de proyecto se pretendió realizar un estudio de cómo están publicados los datos en otra universidad. Se realizó un caso de estudio comparando la universidad de Cádiz con la

Universidad de Pablo de Olavide analizando cómo publican los datos, el número de dataset que publican y el número de estrellas que tienen la publicación de estos datos siguiendo la clasificación de 5 estrellas de Tim Berners-Lee [7].

## 2.3. Organización

El conjunto de personas que ha participado o colaborado de alguna forma en este proyecto es el siguiente:

- Mi tutor académico que se ha encargado de supervisar, ayudar en la búsqueda de herramientas y valorar el trabajo realizado.
- Inmaculada Medina, que me ha permitido acceder algunos datos de carácter menos sensibles con la idea de realizar un proceso ETL, que explicaremos más adelante.
- La autora del presente proyecto, Lucía Batista Flores.

## 2.4. Herramientas utilizadas

### 2.4.1. Cacao

Esta herramienta online<sup>1</sup> se ha usado para la creación de todos los diagramas e ilustraciones de este documento. Se ha utilizado debido principalmente a su facilidad de uso para desarrollar varios diagramas a la vez.

### 2.4.2. ShareLatex

ShareLatex<sup>2</sup> se ha usado como el editor de  $\text{\LaTeX}$  para confeccionar y redactar los contenidos del presente documento, por tener un lugar centralizado desde el que poder editar el contenido desde cualquier equipo, disponiendo de los datos en la nube en cualquier momento, y teniendo en todo momento un visor de PDF integrado, lo cual facilita mucho la redacción al poder visualizar los cambios de forma instantánea.

### 2.4.3. Eclipse

Eclipse se ha usado como entorno de desarrollo (IDE) debido a existir una versión específicamente empaquetada con los componentes y configuración por defecto adecuados para el trabajo con el framework Grails (ver 8.1). Es, además, un editor que se adapta perfectamente al desarrollo con la plataforma Java.

---

<sup>1</sup><https://cacao.com>

<sup>2</sup><https://www.sharelatex.com>

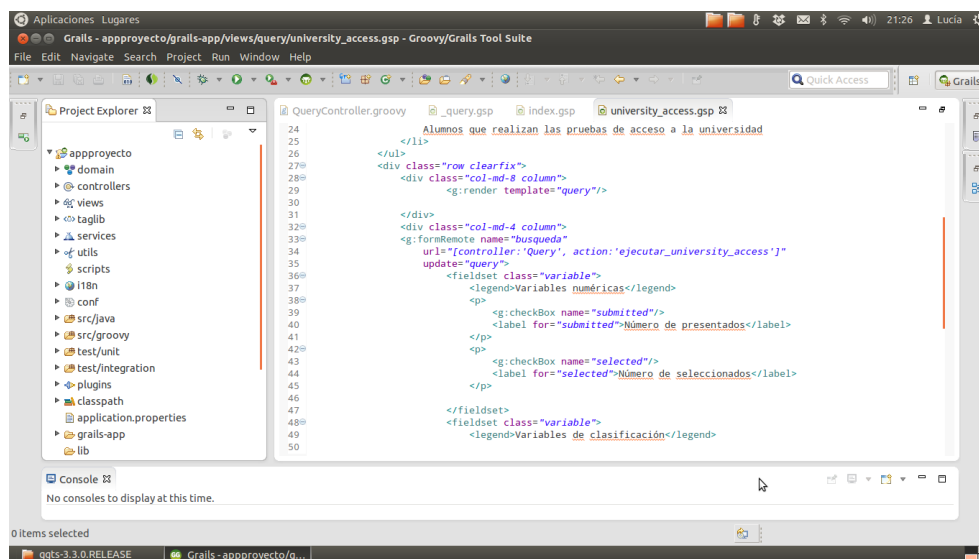


Figura 2.3: Entorno de programación

#### 2.4.4. Subversion

Esta herramienta permite el control de versiones de este proyecto, dicho sistema está centralizado ya que los cambios son enviados a un servidor utilizado por los desarrolladores del grupo.

Para interactuar con Subversion, se utilizan los siguientes comandos:

- Para descargar una copia del trabajo que se encuentra en el repositorio existente: `svn co`
- Para actualizar la copia de trabajo: `svn up`
- Para enviar los cambios hacemos commit: `svn ci -m "Cambios realizados"` La opción `-m` se usa para mandar una descripción de los cambios.

Sin embargo, para la realización de este proyecto hemos usado la herramienta RapidSVN<sup>3</sup> que se puede descargar de los repositorios de Ubuntu. De esta forma, hemos evitado tener abierta un terminal más para el empleo de Subversion.

<sup>3</sup><http://rapidsvn.tigris.org/>

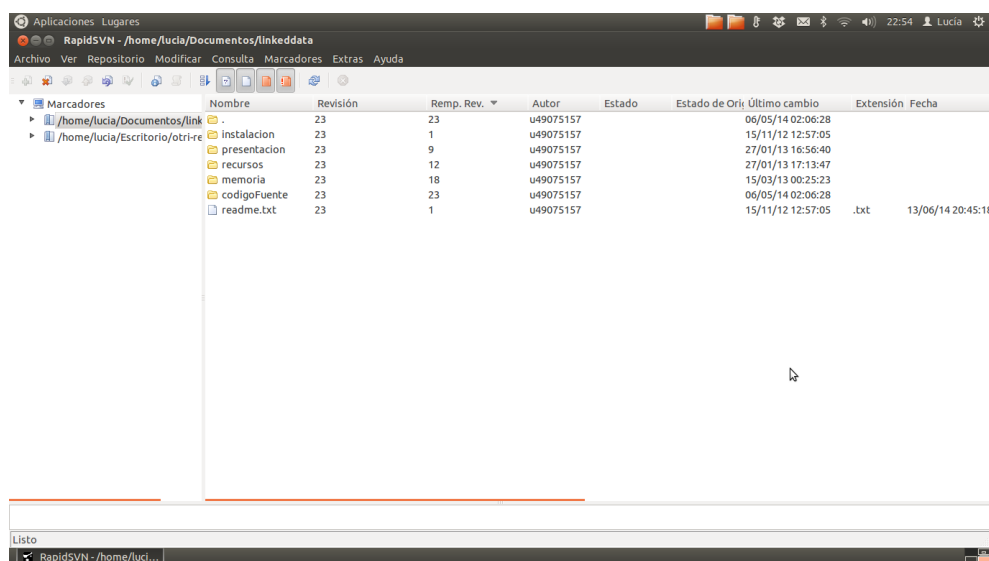


Figura 2.4: Interacción con el repositorio

Si desea obtener la última versión del proyecto puede obtenerla de su sitio web<sup>4</sup>.

### 2.4.5. OpenRefine

Esta herramienta se basa en la última versión de Google Refine8.3<sup>5</sup> que la empresa Google entregó a la comunidad. Sirve para manejar gran cantidad de datos, permite transformar de un formato a otro, hacer limpieza de una gran cantidad de datos, evitando así duplicidades en los valores u otras malformaciones que puedan tener los datos suministrados que puedan provocar problemas en las fases posteriores del proyecto.

### 2.4.6. Pentaho Data Integration (PDI)

Pentaho<sup>6</sup> es una herramienta Open Source, que también recibe el nombre de Kettle o PDI, nos permite realizar procesos ETL (extracción, transformación y carga, por sus siglas en inglés).

Para poder explicar como funciona PDI, tenemos que conocer previamente dos conceptos:

- **La transformación.** Es el conjunto de pasos enlazados entre sí por saltos. A través de los saltos fluye la información y están formados siempre por la salida de un paso y la entrada del siguiente.
- **El trabajo.** Es el conjuntos de tareas que sigue el programa para ejecutar cada una de las transformaciones que se desean.

El PDI integra un conjunto de programas, que cada uno cumple un propósito específico.

<sup>4</sup><https://proyectosii.uca.es/svn/linkedata/>

<sup>5</sup><https://github.com/OpenRefine>

<sup>6</sup><http://www.pentaho.com/>

- **Spoon:** es la herramienta gráfica que nos permite el diseño de las transformaciones y trabajos. Incluye opciones para previsualizar y testear los elementos desarrollados. Es la principal herramienta de trabajo de PDI y con la que construiremos y validaremos nuestros procesos ETL.
- **Pan:** es la herramienta que nos permite la ejecución de las transformaciones diseñadas en spoon (bien desde un fichero o desde el repositorio). Nos permite desde la línea de comandos preparar la ejecución mediante scripts.
- **Kitchen:** es similar a Pan, pero se encarga de ejecutar los trabajos.
- **Carte:** es un pequeño servidor web que permite la ejecución remota de transformaciones y los trabajos.

## 2.5. Costes

Para el desarrollo del presente trabajo los costes que ha supuesto este proyecto pueden dividirse en tres partes.

- Coste material. Para este proyecto se cuenta con el coste del servidor web de la Universidad de Cádiz donde se ha montado todas las herramientas necesarias para su desarrollo. El equipo informático donde se ha alojado el proyecto durante su fase de desarrollo y pruebas ha sido cedido por el Departamento de Ingeniería Informática, por lo que no ha generado ningún coste directo.
- Coste del programador. El sueldo del programador junior que ha realizado las tareas de programador (según aparece recogido en el Boletín Oficial del Estado del 4 de Abril de 2009 [9]), configuración del servidor y de investigación desarrolladas durante 200 días, con una media de 4 horas diarias.
- Coste de licencias de software. Al haber decidido utilizar un conjunto de herramientas disponibles como software libre, siendo coherentes con la filosofía de software libre y conocimiento abierto que propugna la Universidad de Cádiz, el coste en la adquisición de licencias de software ha sido nulo.

	Días	Número de horas	Coste por hora	Total
Coste del programador	308 días	1232 horas	6,57€/hora	8.094,24 €

Tabla 2.1: Tabla Salarial



## 2.6. Gestión de riesgos

En este apartado, vamos a valorar los riesgos de este proyecto y la posibilidad de que ocurra.

### 2.6.1. Limitaciones de información

Existe una limitación en la cantidad de información disponible en la web de la memoria actualmente publicada, por lo tanto habrá que realizar un análisis crítico acerca de que se desear guardar en la base de datos que va a contener la información y que es el principal foco del sistema.

Adicionalmente, durante el desarrollo del proyecto, a pesar de encontrarnos bajo un acuerdo de no revelación de secretos firmado al inicio del proyecto, nos encontramos con un problema con un conjunto de datos que no se deseaba que fueran públicos, principalmente por ser datos sensibles que entrarían directamente en conflicto con algunas de las leyes vigentes de protección de datos de carácter personal, por lo que siguiendo la recomendación de la Universidad, estos datos no han sido incluidos en el conjunto de datos que se han publicado.

Existe también un riesgo a medio/largo plazo para la continuidad del proyecto, consistente en la voluntad de los distintos cuerpos administrativos de la Universidad para facilitar y publicar los datos en este formato, por lo que será importante que todos los departamentos y escalas de la jerarquía universitaria cooperen y sean partícipes de la importancia de este proyecto para la comunidad, evitando que éste pueda caer en el desuso o en el olvido debido a trabas administrativas.



Parte II

Desarrollo



# Capítulo 3

## Estado del arte

En este capítulo se definen los conceptos básicos que necesitamos conocer para realizar el presente proyecto, la descripción de los vocabularios empleados para el desarrollo de los vocabularios RDF que se han creado en este proyecto y las herramientas que hemos empleado para construirlos.

### 3.1. Conceptos básicos

El movimiento open engloba desde una forma de entender la difusión de la información (tanto en forma documentos, como de software, como de conocimiento) en cuanto el servicio público, el trabajo por la comunidad y el bien común.

#### 3.1.1. Open government

Es una doctrina política cuyos antecedentes se han dado en Nueva Zelanda [12] y Estados Unidos [11], donde se propone que las administraciones y el gobierno deben de ser lo más transparentes posible, además deben de ir de la mano de la creación de espacios que permitan la participación ciudadana, permitiendo así una democracia participativa que sea el moto de una evolución de nuestro sistema democrático de convivencia y valores más allá del derecho a voto cada cuatro años.

Esta doctrina se fundamenta en tres pilares básicos: transparencia, colaboración y participación.

- La transparencia permite que las administraciones abran sus datos (open data) pasando así toda información pública a ser propiedad de los ciudadanos y eliminando las barreras de acceso a la información, ya que no existe otra forma posible para crear una ciudadanía madura, responsable y emprendedora. De esta manera se pueden acceder a los datos de forma fácil y gratuita por lo que facilita la rendición de cuentas de la administración pública.
- La colaboración, favorece a que los ciudadanos, las empresas, asociaciones y otras entidades se impliquen en el trabajo de la administración, dándole una parte de responsabilidad a la sociedad civil.
- La participación permite a la ciudadanía formar parte en la creación de las políticas públicas y así transmitir el conocimiento y experiencia de esta a la administración.

### 3.1.2. Open data

Es una filosofía y una práctica que requiere que ciertos datos sean de acceso público, sin limitaciones técnicas o legales. Para que los datos sean open data, se deben de cumplir los siguientes objetivos:

- **Públicos.** Los datos públicos no deben ser restringido de ninguna forma posible.
- **Detallados.** Los datos expuestos deben ser aquellos que proceden directamente de la fuente, como el mayor grado de simpleza posible.
- **Actualizados.** Se deben de hacer públicos con el mínimo tiempo posible.
- **Accesibles.** Deben de estar expuestos de la forma más clara posible para que cualquier ciudadano pueda acceder a ellos con facilidad.
- **Automatizados.** Deben ser publicados los datos en bruto en formato que puedan se encuentren estructurados y que a la vez puedan ser procesados por las máquinas.
- **Sin registro.** Los datos serán accesibles universalmente para todos, sin necesidad de previo registro.
- **Abiertos.** Los datos no se deben encontrar en formatos propietarios.
- **Libres.** Los datos tendrán licencia libre.

### 3.1.3. Estándares Linked Data

Los conjuntos de datos que se publican bajo estas características que hacen referencia a la categorización de los datos públicos en catálogo de datos, reciben el nombre de “dataset”, facilitando su indexación y localización, además deben de tener una descripción sobre la temática de los datos, una licencia para saber como se pueden distribuir con otros, la fecha que se produjo la última actualización y por último deben de estar en formatos abiertos como los siguientes ejemplos:

- **Texto plano.** Valores separados por coma, es un documento de texto plano donde se representan datos tabulados en tablas separados por comas o punto y coma, además las filas están separadas por saltos de líneas. Las extensiones son CSV o TXT.
- **XML.** Lenguaje de Etiquetado Extensible, es un formato ampliamente utilizado para el intercambio estructurado de datos, puesto que ofrece la posibilidad de mantener la estructura de los datos y la forma en que los archivos son construidos, permiten a los desarrolladores escribir parte de la documentación con datos sin interferir con la lectura de estos.
- **RDF.** Infraestructura para la Descripción de Recursos, es un modelo universal recomendado por el *World Wide Web Consortium* (W3C) [5] que hace uso de URLs como identificadores interconectando los datos abiertos entre sí. RDF permite intercambiar y enlazar a través de diferentes aplicaciones datos y recursos sin que pierdan su significado, facilitando la reutilización y el enriquecimiento de los recurso en la Web. Este modelo se basa en convertir las declaraciones de los recursos en expresiones con la forma sujeto-predicado-objeto, conocidas en términos RDF como tripletas (ver 3.1.6).

- **RDF Schema.** Es una extensión del estándar RDF [6], un lenguaje primitivo de ontologías que proporciona los elementos básicos para el modelado de vocabularios, con el objetivo de poder formar una estructura de los mismos, aportando una semántica y una interconexión entre todos los datos. Con esto se consigue que sea posible almacenar todo el conjunto de datos en una base de datos de tripletas (*triplestore*), de forma que sea posible ejecutar consultas SPARQL contra él, obteniendo la posibilidad de recuperar y filtrar los datos con un lenguaje de consultas con un nivel de abstracción mucho mayor, más cercano al lenguaje humano.
- **OWL.** El lenguaje Web de Ontologías es un lenguaje de marcado semántico para publicar y compartir ontologías en la *World Wide Web*. Este estándar [4] se ha desarrollado como una extensión del RDF y sus antecedentes son DAML + OIL, donde DAML (*DARPA Agent Markup Language*) es un lenguaje de marcado desarrollado por Agencia de Investigación Avanzada de Proyectos de EE.UU (DARPA) basándose en XML. Tiene mayor capacidad de establecer relaciones semánticas entre los objetos, creando así un mayor nivel de interoperabilidad entre los sitios web. Por el otro lado, OIL se puede considerar como la infraestructura de ontologías para la Web Semántica, basada en un lenguaje de marcado que es compatible con RDF.

La combinación de RDF con otras herramientas como RDF schema y OWL forman el conjunto de tecnologías básicas de la web semántica.

#### 3.1.4. Clasificación de fuentes Open Data

La propuesta de clasificación del grado de calidad de los datos “open data” por Tim Berners-Lee [7], considerado el “padre” de Internet y director del W3C, clasifica por estrellas los datos publicados, de tal modo que a mayor facilidad de uso y reutilización, mayor número de estrellas.

- **Una estrella (Datos publicados):** Se publican en formatos no estructurados (como, por ejemplo, ficheros PDF, ficheros de texto o imágenes). La ventaja de este nivel es que los datos, al no requerir ningún tipo de transformación previa, pueden ser publicados de una manera sencilla, pero tiene la desventaja de que estos formatos son muy difíciles de manipular y reutilizar.
- **Dos estrellas (Datos estructurados):** Se entregan los datos de manera estructurada, como en un archivo Excel o con extensión XLS. En este nivel se ofrece información y, aunque los datos se publican en formatos estructurados, esta publicación se hace en formatos propietarios (como, por ejemplo, ficheros Excel, XLS). La desventaja de este nivel, por lo tanto, es que para poder acceder a los datos publicados se requiere de herramientas no abiertas.
- **Tres estrellas (Formatos abiertos):** Entregar los datos en un formato que no sea propietario, como CSV en vez de excel, XML, RDF, etc. En este nivel se ofrece información y los datos se publican en formatos estructurados, abiertos y no propietarios (como, por ejemplo, en formato CSV, XML o RDF). La ventaja de este nivel es que el acceso y la manipulación de los datos pueden hacerse de manera sencilla al no existir limitaciones por las características del software. Desde el punto de vista de quien publica la información, la desventaja es que, en la mayoría de los casos, se requiere para la publicación el desarrollo de procesos de transformación previos a la misma.

- **Cuatro estrellas (URL para identificar un dato):** Usar URL para identificar cosas y propiedades, de manera que se pueda apuntar a los datos. Requiere usar un estándar: RDF. La principal ventaja de este nivel es que facilita la labor de reutilización de los datos y actualización de los mismos, al encontrarse siempre identificados por una misma URL. Para el publicador, la desventaja es que supone un mayor esfuerzo de estructuración, separación de los datos y asignación de las URLs.
- **Cinco estrellas:** En este nivel se trata de vincular los datos de un conjunto de datos con otros datos, en base al contexto de los diferentes conjuntos de datos. Para lograr llegar a enlazar conjuntos de datos de diversas fuentes es necesario el llevar a cabo un enriquecimiento semántico de los conjuntos de datos publicados y una estandarización del vocabulario utilizado en los diferentes contextos.

La clasificación de Berners-Lee es una recomendación para lograr el desarrollo de la web semántica, que vincula datos estructurados, frente a la web del hipertexto, que simplemente enlaza páginas o documentos en HTML y cuyos datos apenas se pueden manejar.

### 3.1.5. Web Semántica

La Web semántica es una nueva generación web, que se diferencia por dotar de mayor significado a los datos publicados en ella, lo que permite al usuario encontrar respuestas a sus preguntas de una forma más rápida y sencilla. Los principales componentes son XML, que provee una sintaxis elemental para las estructuras de contenidos dentro de los documentos, XML schema, RDF, RDF schema y OWL.

Las capas y funciones de estas dentro de la web semántica son las siguientes:

- **Unicode.** Es un estándar que proporciona la codificación de los textos en diferentes idiomas con un número de caracteres mucho mayor, gracias al uso de más bits por carácter, lo cual permite mostrar un mayor número de idiomas con alfabetos distintos a los que usan el alfabeto latino.
- **URI.** Es el acrónimo de *Uniform Resource Identifier* o identificador uniforme de recursos, que permite la localización de un recurso accediendo a través de Internet.
- **XML+NS+XMLSCHEMA.** Esta capa es la más técnica ya que se agrupan las tecnologías que hacen posible que los diferentes agente puedan entenderse entre ellos. XML es un estándar que ofrece un formato común para el intercambio de documentos, NS (los espacios de nombres o *namespaces*) nos proporcionan un método para cualificar elementos y atributos de nombres usados en documentos XML para asociarse con un espacio de nombre identificado por una URI, y por último, el XML Schema es el lenguaje que nos ofrece una plantilla para elaborar documentos estándar.
- **RDF+RDF Schema.** Esta capa, basada en la anterior, se define como el “lenguaje universal” con el cual podemos expresar diferentes ideas en la web semántica. RDF es lenguaje simple con el cual definimos sentencias en el formato de una 3-tupla o tripleta.
- **Lenguaje de ontologías.** Ofrece un criterio para catalogar y clasificar la información. El uso de ontologías permite describir objetos y sus relaciones con los demás. La ontología es la especificación formal de una conceptualización de un dominio concreto, a través de la creación de clases y propiedades que identifiquen los recursos.



- **Lógica.** Son las reglas de inferencia.  
Ej: Una ciudad tiene un código de estado.  
Una dirección es un código de la ciudad.  
Luego esa dirección tiene un código asociado, esto permite a los ordenadores deducir como un humano. Si no existiesen esas reglas de inferencia no sería posible realizar pruebas.
- **Pruebas.** A través de un lenguaje unificador que hace posible las inferencias lógicas hechas haciendo uso de las reglas de inferencia, se realizan intercambios de pruebas en este lenguaje.
- **Confianza.** Los agentes inteligentes realizan inferencias sobre la información que interactúa con el entorno sin necesidad de supervisión o control por el usuario, básicamente se encarga de recoger, filtrar y procesar información.
- **Firma digital.** Es un bloque cifrado de datos utilizados por los ordenadores y los agentes inteligentes para verificar que la información ha sido ofrecida por una fuente fiable.

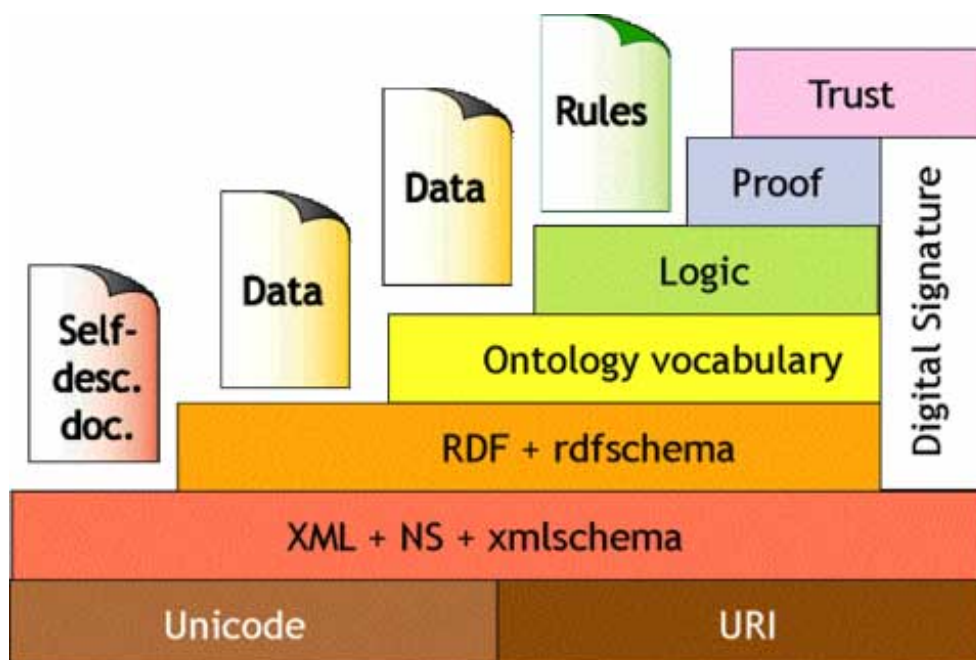


Figura 3.1: Capas de la Web Semántica

### 3.1.6. Tripletas

Una tripleta se define como un conjunto de 3 elementos asociados a una entidad particular. Los elementos dentro de una tripleta tienen la siguiente forma:

- **Sujeto.** Recurso al que nos referimos.
- **Predicado.** El recurso que nos indica que estamos definiendo.
- **Objeto.** Puede ser un recurso o literal que podríamos considerar el valor de lo que acabamos de definir.

El modelo RDF o Resource Description Framework es un modelo común, o mejor dicho, un framework que permite realizar descripciones sobre los recursos y permite que estos sean referenciables mediante URIs. Por su parte RDF Schema permite el modelo de datos con una semántica claramente definida, ofreciendo así en esta capa no sólo la descripción de los datos sino también cierta información semántica.

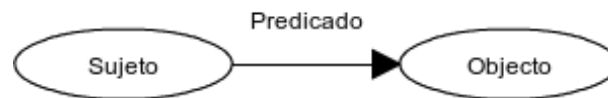


Figura 3.2: Tripleta RDF

### 3.1.7. Linked Data

El concepto de Open Data y el de Linked Data (datos vinculados), van de la mano con el desarrollo de la Web Semántica. Es un método de publicación de datos estructurados que se obtienen a partir del movimiento Open Data, para que estén interconectados entre sí y puedan ser consultados más fácilmente.

El objetivo de Linked Data es conseguir que la Web sea una gran base de datos interconectados y distribuidos, que sean perfectamente legibles por las máquinas, que serán las encargadas de hacer el trabajo de entender los requisitos que busca el usuario y buscar las respuestas automáticamente.

### 3.1.8. SPARQL

SPARQL es un acrónimo recursivo del inglés *SPARQL Protocol And RDF Query Language*, desarrollado por el W3C y surgió como recomendación para realizar un lenguaje de consultas en la Web Semántica que define un lenguaje para la recuperación de datos almacenados como RDF/RDFS.

Esta tecnología permite que los usuarios de las Web Semánticas no se tengan que preocupar por la estructuras de las base de datos o el formato que se utiliza para almacenar dichos datos sino que se centran únicamente en la información que se interesa consultar. Así, los desarrolladores y los usuarios finales son capaces de obtener los resultados obtenidos de las búsquedas en SPARQL a través de cualquier tipo de información que terminará por devolver un conjunto de sentencias en RDF.

- **SPARQL Query Language for RDF:** define la sintaxis y la semántica de las consultas, para la composición, concordancia y experimentación.
- **SPARQL Protocol for RDF:** define un protocolo para el acceso remoto entre una aplicación que emita consultas SPARQL y un servidor que le envíe los resultados.
- **SPARQL Query Results XML Format:** especifica un formato XML para los resultados de las consultas SPARQL.

Las sintaxis de SPARQL es muy parecida a la del lenguaje SQL usado para definir y consultar la información y estructura en los sistemas de gestión de bases de datos relacionales más extendidos, a continuación explicaremos las palabras reservadas más importantes:

- *PREFIX*. Define los prefijos para los espacios de nombres que se utilizarán en la consulta.
- *FROM*. Nombra los grafos que serán consultados.
- *WHERE*. Patrones para la consulta en forma de una lista de tripletas (dominio, relación, rango)
- *FILTER*. Establece restricciones sobre las tripletas de la consulta.
- *GROUP BY*. Agrupa los datos por valores.

A continuación se muestra un ejemplo de consulta SPARQL:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX type: <http://dbpedia.org/class/yago/>
PREFIX prop: <http://dbpedia.org/property/>
SELECT ?country_name ?population
WHERE {
    ?country a type:LandlockedCountries ;
            rdfs:label ?country_name ;
            prop:populationEstimate ?population .
    FILTER (?population > 15000000 && langMatches(lang(?country_name), "ES")) .
}
```

Esta consulta SPARQL realiza una petición de devolución de los registros coincidentes con los países (?country) rodeados exclusivamente de tierra (?country a type:LandlockedCountries) que posean una población (?population) mayor a 15000000, filtrando los nombres de los países al castellano (langMatches(lang(?country\_name), 'ES')), ya que por defecto nos devolverá resultados duplicados: un nombre de país por cada traducción del mismo. Esta consulta se puede ejecutar directamente contra el endpoint de la DBpedia a través de su interfaz de consultas web<sup>1</sup>.

### 3.2. Otros proyectos ya desarrollados en Linked Data

La Web de Datos o Linked Data tiene como objetivo el acceso, publicación y reutilización de los datos a través de la Web además del desarrollo de aplicaciones que hagan uso de esta tecnología. La idea principal es que se puedan reutilizar los mismos datos por aplicaciones distintas sin tener que hacer uso de documentos.

El auge de la publicación de los datos surge con la idea de Open Government Data que son iniciativas que pretenden difundir la idea de la publicación de los datos públicos de las administraciones públicas. Algunas de las iniciativas desarrolladas son las siguientes:

---

<sup>1</sup><http://dbpedia.org/sparql>

- **Dbpedia**<sup>2</sup> es un proyecto comunitario cuya misión es extraer información de Wikipedia y publicarla de forma estructurada en la web. Este proyecto está realizado por la Universidad de Leipzig, Universidad Libre de Berlín y la compañía OpenLink Software. En la base de datos se describen más 3.380.000 entidades, contiene 1.460.000 enlaces a imágenes, 5.543.000 enlaces a páginas externas, 4.878.000 enlaces a datasets externos y 415.000 categorías Wikipedia.
- **Data.gov**<sup>3</sup>: desarrollada por el gobierno de los Estados Unidos dentro de la Open Government Data Initiative. Su propósito es promover el acceso a datos públicos de gran calidad, accesibles a través de datasets tratables por máquinas generados por la Executive Branch del Gobierno Federal. Esta iniciativa contiene 3281 conjuntos de datos a fecha de Mayo de 2011.
- **Open Data Euskadi**<sup>4</sup>: iniciativa del gobierno del País Vasco, cuyo compromiso es exponer los datos públicos que obran en su poder de forma reutilizable, con el fin de que terceros puedan crear servicios derivados de los mismos<sup>5</sup>. Esta iniciativa contiene más de 1401 catálogos de datos diferentes a fecha de mayo de 2011.
- **EduBase**<sup>5</sup> es una base de datos de entidades educativas registrada por el Departamento de Educación del Gobierno de Reino Unido. Dicha base de datos es accesible como linked open data a través de un portal especializado para su consulta.

El portal de datos se basa en Elda<sup>6</sup>, una plataforma para ofrecer una API de acceso a datos. Esta elección de plataforma permite ofrecer una interfaz pública basada en servicios web REST (*Representational state transfer*) que permite que otros programas realicen consultas usando el protocolo HTTP de forma simple para obtener los datos necesarios. Soporta múltiples formatos de salida, como JSON, XML, RDF, CSV, etc.

El desarrollo de la Web de Datos, se mostrará a continuación con las imágenes de la evolución de los datasets recogido por LinkedData.org [15], donde se muestra la evolución desde el año 2007 hasta el 2011. Cada burbuja representa los diferentes conjuntos de datos y las flechas la relaciones que se establecen entre los distintos dataset, quedando así todo el conocimiento relacionado y mostrando la increíble expansión que ha tenido el movimiento.

---

<sup>2</sup><http://wiki.dbpedia.org/About>

<sup>3</sup><http://www.data.gov/>

<sup>4</sup><http://opendata.euskadi.net>

<sup>5</sup><http://education.data.gov.uk/>

<sup>6</sup><https://github.com/epimorphics/elda>

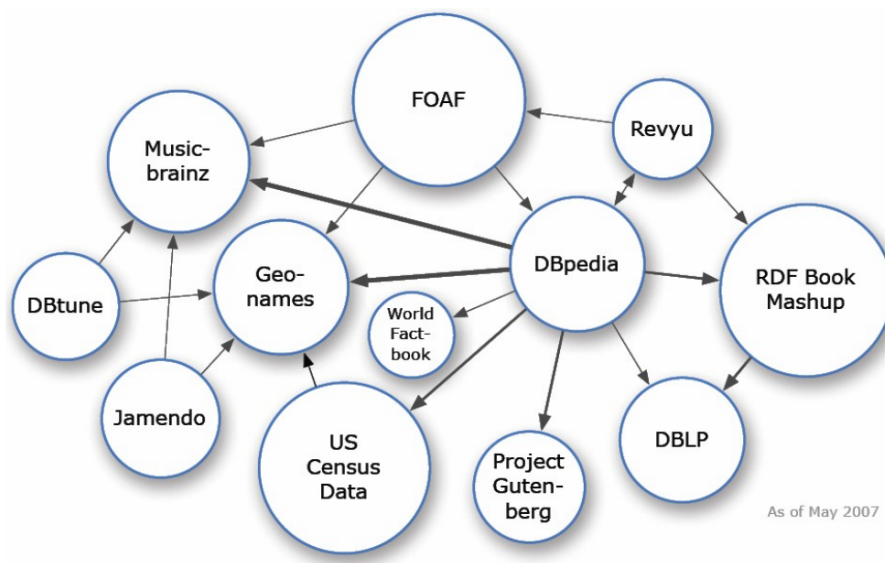


Figura 3.3: LOD 2007

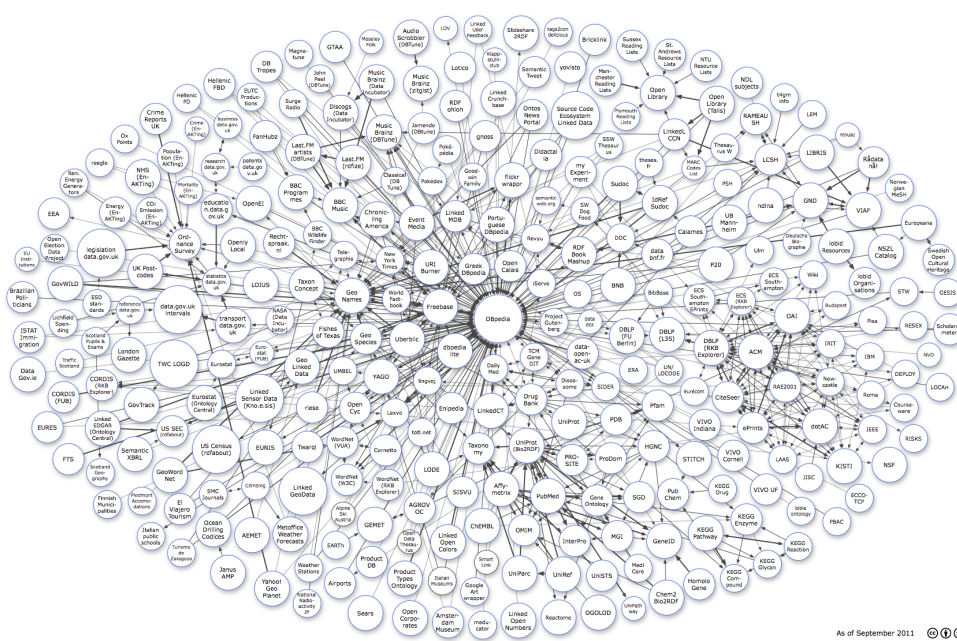


Figura 3.4: LOD 2011

Gracias a esta nueva tecnología, en el siguiente apartado nos enfocamos en como únicamente las Universidades como entidad administrativa, han realizado la apertura de sus datos.

### 3.2.1. Apertura de datos en instituciones universitarias

En estos momentos existe una alianza de las universidades europeas llamada Linked Universities<sup>7</sup> que se han unido para exponer sus datos empleando tecnología linked data, permitiendo que los datos en crudo dispongan de una referencia y estén entrelazados entre sí, dispongan de un fácil acceso, y sea un conjunto de información libre, conectada y reutilizable. La idea de que diferentes instituciones y organizaciones contribuyan en crear un Web de datos.

Actualmente, ya existen unas cuantas universidades que actualmente están exponiendo sus datos públicos como linked data, usando tecnologías como RDF y SPARQL que les permite acceder directamente a la información, como publicaciones, cursos, material educativo, etc.

Estas iniciativas están actualmente desconectadas las unas de las otras, porque cuando se desarrolla una plataforma de de datos públicos que además poseen sus datos vinculados requiere bastante esfuerzos, de desplegar un portal y formarse sobre cómo se deben publicar estos datos y crear ontologías o usar herramientas basadas en linked data. A la larga todo este trabajo no sólo va a beneficiar individualmente a cada institución publicando sus datos sobre educación e investigación, sino que también va a ayudar a agregar más información y comparar datos entre sí para sacar conclusiones. Es decir, que estaremos ayudando a un objetivo común.

Los principales objetivos de Linked Universities son:

- Identificar, dar soporte y desarrollar los vocabularios más comunes de linked data universitarias, para que puedan ser usados por cada universidad que use linked data para sus propios propósitos.
- Compartir el conocimiento sobre herramientas o prácticas para la exposición de los datos de las universidades en linked data.
- Apoyar, mediante el intercambio de experiencias y la reutilización de vocabularios, a las iniciativas para exponer los datos de las universidades como linked data.

Linked Universities mantiene un listado<sup>8</sup> con el conjunto de universidades de todo el mundo adheridas a su iniciativa, así como el conjunto de vocabularios<sup>9</sup> disponibles para su uso generados por la comunidad de participantes. No obstante, esta lista no es ni mucho menos exhaustiva.

A modo de ejemplo de los vocabularios que se pueden encontrar en el sitio de Linked Universities, comentaremos brevemente uno de los vocabularios publicados en dicho sitio web. Concretamente, se trata del vocabulario TEACH<sup>10</sup>, que permite relacionar entidades con cursos o asignaturas académicas impartidas por una universidad.

```
1 @prefix tisc: <http://observedchange.com/tisc/ns#>.
2 @prefix ical: <http://www.w3.org/2002/12/cal/icaltzd#>.
```

<sup>7</sup><http://linkeduniversities.org/>

<sup>8</sup><http://linkeduniversities.org/lu/index.php/members/index.html>

<sup>9</sup><http://linkeduniversities.org/lu/index.php/vocabularies/index.html>

<sup>10</sup><http://linkedscience.org/teach/ns/>

```

3 @prefix GE: <http://www.w3.org/2002/12/cal/tzd/Germany/Berlin#>.
4 @prefix teach: <http://linkedscience.org/teach/ns#>.
5 @prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
6 @prefix aiiso: <http://purl.org/vocab/aiiso/schema#>.
7 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
8 @prefix owl: <http://www.w3.org/2002/07/owl#>.
9 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
10
11 <http://data.uni-muenster.de/context/csa/course/197>
12   rdf:type          aiiso:Course;
13   teach:academicTerm "WS 2011/2012";
14   teach:bookingNumber "142256"^^xsd:int;
15   teach:courseDescription "BSc Studierenden koennen diesen Kurs im
16     Modul [...]."@de;
17   teach:courseTitle "Study Project: Open Floor Map"@en;
18   teach:ects "5"^^xsd:int;
19   teach:studentGroup <http://data.uni-muenster.de/context/csa/
20     group/197>;
21   teach:teacher <http://data.uni-muenster.de/context/csa/teacher/
22     ChristianKray>;
23   teach:weeklyHours "2"^^xsd:int;
24   owl:sameAs <http://data.uni-muenster.de/context/csa/course/197>;
25   foaf:homepage <http://data.uni-muenster.de/context/csa/course
26     /197>;
27   teach:module <http://data.uni-muenster.de/context/csa/module
28     /123>;
29   teach:studyProgram <http://data.uni-muenster.de/context/csa/
30     studyprogram/1234>;
31   teach:deadlineDraftReport "2012-06-10T23:59:59"^^xsd:dateTime;
32   teach:deadlineFinalReport "2012-06-30T23:59:59"^^xsd:dateTime;
33   teach:deadlineReviewReport "2012-06-20T23:59:59"^^xsd:dateTime;
34   teach:notArrangedAt "2012-06-19"^^xsd:date;
35   teach:grading "Grading of the seminar is as follows: 1/3 for the
36     seminar work, 1/3 for the opponent work and presentation of
37     the own work, 1/3 for the active participation."@en;
38   teach:nextReading <http://wiki.openstreetmap.org/wiki/
39     Indoor_Mapping>;
40   teach:arrangedAt [
41     rdf:type          ical:Vevent;
42     ical:byday "TH"^^xsd:string;
43     ical:freq "WEEKLY"^^xsd:string;
44     ical:interval "1"^^xsd:int;
45     ical:summary "Study Project: Open Floor Map (weekly
46       meeting) "@en;
47     teach:room <http://data.uni-muenster.de/context/page/
48       cris/rooms/ifgi/seminarroom>;
49     teach:building <http://data.uni-muenster.de/context/page
50       /cris/addresses/48151/WeselerStr253>;
51     ical:dtend "2012-04-05T16:00:00"^^GE:tz;
52     ical:dtstart "2012-06-12T14:00:00"^^GE:tz;
53     tisc:locatedAt "51.941269,7.610532" ];
54     ical:uid "2012-06-12T16: 95-69-1-7@lod";

```

El ejemplo anterior describe una asignatura impartida en la Universidad de Münster de Alemania. Como se puede observar a simple vista, encontramos propiedades reconocibles de una asignatura cualquiera, como pueden ser el trimestre académico en el que se imparte (`teach:academicTerm`), el título y la descripción de la asignatura (`teach:courseTitle`, `teach:courseDescription`), número de créditos ECTS (`teach:ects`), etc. Encontramos otras propiedades más específicas en cuanto a entregas, métodos de evaluación y próximas clases (haciendo uso del vocabulario `ical` de referencia de eventos).



## Capítulo 4

# Requisitos del sistema

En esta sección se describen los objetivos y el catálogo de requisitos de este Proyecto de Fin de Carrera.

### 4.1. Situación actual

Este proyecto se ha realizado en colaboración con el grupo de investigación de Mejora de Procesos Software y Métodos formales (SPI-FM, *Software Process Improvement and Formal Methods*), creado en el año 2004 dentro del Plan Andaluz de Investigación, que le asigna la referencia TIC-195. Está integrado por un grupo de profesores del Departamento de Ingeniería Informática y tiene su sede en la Universidad de Cádiz. Entre sus principales líneas de investigación se encuentra la Web Semántica, dentro del área de Ingeniería Web, a la que se adscribe el presente proyecto como labor de investigación. El objetivo de la misma es desarrollar tecnologías para publicar datos legibles por aplicaciones informáticas de forma automática, dotándolas de un contexto que les proporciona información reconocible por máquinas, a diferencia de el enfoque clásico en el que los datos carecían de entidad, siendo sólo usados por aquellos que supieran su significado y forma.

La labor de investigación de este grupo y su afán por la mejora continua al servicio de la sociedad planteó una iniciativa cuyo objetivo final es el de publicar multitud de datos de muy diversa índole acerca de la Universidad, tales como cifras relativas al alumnado, titulaciones, personal, infraestructuras, etc., disponibles de forma gratuita y libre a través de Internet de forma que cualquier ciudadano o entidad pueda en todo momento acceder a dicha información para su interés particular de forma transparente, siguiendo la filosofía que difunde el movimiento Open Data, a la vez que se dota a dicha información de la posibilidad de ser tratada automáticamente reteniendo su sentido intrínseco, colaborando con la popularización de los estándares libres de intercambio de información y conocimiento y ofreciendo, de forma fácil al ciudadano, una herramienta para consultar y comparar dichos datos sin necesidad de conocimientos específicos o herramientas para ello a través de una herramienta web.

Así todo lo expuesto anteriormente se materializa en el presente Proyecto de Fin de Carrera, que pretende asentar las bases de la iniciativa Linked Open Data dentro de la Universidad de Cádiz.

## 4.2. Requisitos funcionales

Tras las reuniones mantenidas a nivel de departamento, se decidió establecer los siguientes requisitos globales:

- Analizar el estado actual de la memoria académica de la Universidad de Cádiz y extraer datos estructurados a partir de ella.
- Diseñar y estructurar un proceso de extracción, transformación y carga (ETL) para unificar las distintas bases de datos existentes en la universidad.
- Analizar y crear vocabularios RDF específicos para describir datos publicados en el dominio de la universidad, reutilizando en la medida de lo posible los ya existentes.
- Crear una plataforma Linked Open Data que permita publicar esos datos.
- Diseñar e implementar una aplicación que explote los datos que ya han sido publicados permitiendo visualizar y comparar dichos datos de forma gráfica y accesible al usuario.

## 4.3. Requisitos no funcionales

En adición a los anteriores requisitos, claves para llevar a buen puerto la investigación y sus resultados, se establecieron otros requisitos que el producto de nuestro trabajo debe cumplir, como resultado de una aproximación científico-técnica, aplicando siempre un enfoque que asegure la total confianza en los resultados y, a fin de cuentas, garantizando la calidad de los mismos. Concretamente, nos referimos a:

- **Rendimiento:** Aplicando los conocimientos de las diferentes líneas de investigación del departamento y aquellos obtenidos durante el transcurso de la titulación, permitiendo desarrollar un producto que pueda trabajar de forma eficiente en cuanto a tiempo y recursos.
- **Interoperabilidad:**

"La interoperabilidad es la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada." [17]

Dado el objetivo de apertura de datos en el que se sustenta este proyecto es obligatorio usar un conjunto de estándares universales y libres que permitan a cualquier individuo interactuar con nuestra plataforma, ejemplos de esto son el estándar de codificación de caracteres Unicode<sup>1</sup>, XML<sup>2</sup> como base para la representación de datos estructurados y RDF<sup>3</sup> como un conjunto de estándares para aportar semántica a los datos.

Gracias al conjunto de todos estos estándares la Web se convierte en una gran base de datos, permitiendo así la reutilización de conocimiento, la navegación conceptual y la fusión de sistemas de organización del conocimiento a través de múltiples dominios.

- **Mantenibilidad:** Todo el sistema se encuentra completamente documentado en este documento.

---

<sup>1</sup><http://www.unicode.org/standard/standard.html>

<sup>2</sup><http://www.w3.org/XML/>

<sup>3</sup>[http://www.w3.org/standards/techs/rdf#w3c\\_all](http://www.w3.org/standards/techs/rdf#w3c_all)

## 4.4. Requisitos de información

Para el presente proyecto ha sido necesario contar con la siguiente información disponible públicamente en la memoria académica 2012-2013, la más reciente a fecha de hoy, y anteriores ediciones. La información que se ha extraído pertenece a distintas categorías, las cuales se enumeran a continuación:

- Alumnos (matrículas, presentados a pruebas de acceso, titulación/ciclo, etc.)
- Personal (PDI, PAS, jubilaciones, salarios, etc.)
- Financiación (contratos, becas, financiación de proyectos, etc.)
- Servicios universitarios (campus virtual, compromiso social, niveles de satisfacción, etc.)
- Investigación (proyectos de innovación docente, proyectos de investigación, etc.)
- Infraestructuras
- Gestión de energía y residuos (consumo de agua, reciclado de materiales, etc.)



# Capítulo 5

## Análisis

Se presentan en este capítulo los distintos casos de uso que se han tenido en cuenta a la hora de diseñar la aplicación de consulta de datos (*Linked Data Viewer*), así como las diferentes alternativas y flujos de trabajo que un usuario puede encontrarse durante su uso. Adicionalmente, se muestra los diagramas de comportamiento de la aplicación, dando una vista en conjunto de la interacción entre los distintos componentes del sistema durante la ejecución del mismo.

### 5.1. Casos de uso

A continuación vamos a detallar los casos de uso que tiene nuestro sistema completo.

#### 5.1.1. Caso de uso: Añadir dataset.

1. Loguearse en el CKAN.
  - a) Registrarse como usuario de CKAN si no dispone de usuario.
2. Seleccionar en la barra superior añadir dataset.
3. Escribir los datos necesarios para añadir el dataset: título, URL, licencia, descripción y grupo al que pertenece.
4. Añadir los datos mediante un link, una API o un fichero.
5. Añadir una descripción de los cambios que se han realizado en los datos.
6. Pulsar el botón de añadir dataset.

#### 5.1.2. Caso de uso: Buscar un dataset.

1. Loguearse en el CKAN.
  - a) Registrarse como usuario de CKAN si no dispone de usuario.
2. Seleccionar en la barra superior "buscar".
3. Escribir en la barra de búsqueda el dataset deseado.
4. Seleccionar el dataset que estaba buscando.

### 5.1.3. Caso de uso: Descargar un dataset.

1. Loguearse en el CKAN.
  - a) Registrarse como usuario de CKAN si no dispone de usuario.
2. Seleccionar en la barra superior "buscar".
3. Escribir en la barra de búsqueda el dataset deseado.
4. Seleccionar el dataset que se desea consultar.
5. Seleccionar la opción de recursos de la barra, elija el recurso que desea descargar y pulse el botón de descarga.

### 5.1.4. Caso de uso: Consultar de forma programática los datos.

1. Vaya a la página del endpoint de SPARQL del D2RQ Server<sup>1</sup> instalado en el servidor LOD.
2. Escriba la consulta que desee.
  - a) La consulta está mal y devuelve un error, vuelva a escribirla.
3. Devuelve el resultado de la consulta con éxito.

### 5.1.5. Caso de uso: Generar gráfica

**Actores:** Usuario

**Escenario Principal:**

1. El usuario visualiza el menú principal y elige un submenú dependiendo la temática que desee consultar.
2. Dentro del submenú, el usuario vuelve a elegir un tópico de datos que desee consultar gráficamente.
3. El usuario accede al formulario de consulta, donde debe de elegir una variable de tipo numérico (si existe más de una) y una variable de clasificación.
  - a) Si el usuario no elige alguna de las dos variables, la aplicación mostrará un error diciendo que debe seleccionar una variable de cada tipo.
4. Pulsa sobre el botón mostrar y aparece la gráfica seleccionada.

### 5.1.6. Caso de uso: Comparar dos universidades

**Actores:** Usuario

**Escenario Principal:**

1. El usuario visualiza el menu principal y elige pulsar sobre la opción de la barra superior de *Comparativas con otras universidades*.

---

<sup>1</sup><http://lod.uca.es/d2rq/snorql/>

2. El usuario accede al formulario de consulta, donde debe subir un fichero con formato el cual es una forma abreviada de serialización no-XML de modelos en RDF, diseñado pensando en la legibilidad por parte de humanos.
  - a) Si el usuario no sube un fichero, la aplicación le mostrará un error de que es un campo obligatorio.
  - b) Si el fichero subido por el usuario no cumple el formato exigido, la aplicación devolverá un error diciendo que la sintaxis del RDF no es correcta.
3. El usuario elige una variable de tipo numérico y una de tipo de clasificación para poder visualizar la gráfica.
  - a) Si el usuario no elige alguna de las dos variables, la aplicación mostrará un error diciendo que debe seleccionar una variable de cada tipo.
4. Por último, el usuario selecciona el botón Mostrar para visualizar la gráfica comparativa.

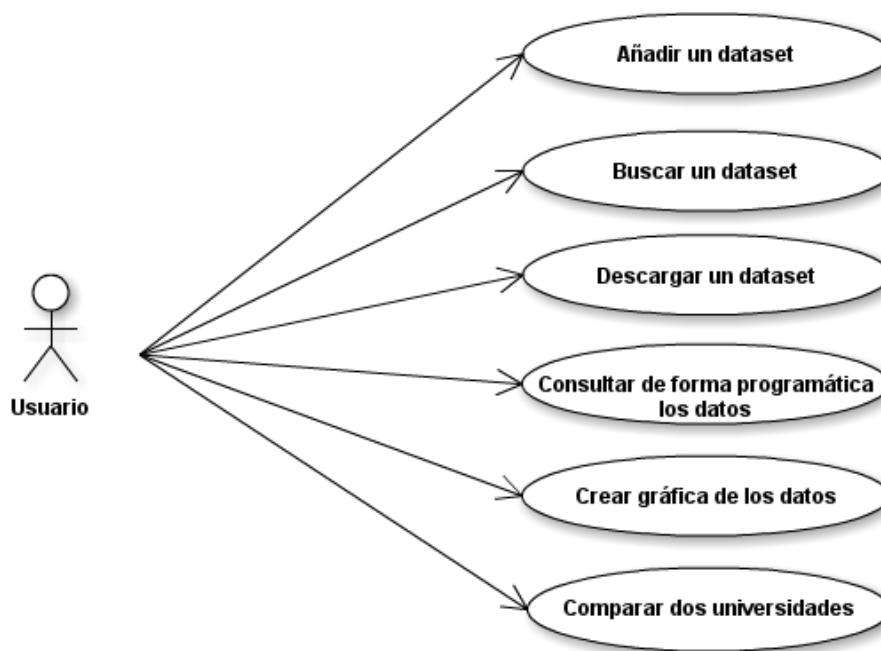


Figura 5.1: Casos de uso

### 5.1.7. Diagrama de Secuencia

A continuación, se muestra los diagramas de secuencia únicamente de la aplicación ya que el diagrama de secuencia del caso de uso de consultar de forma programática es exactamente igual que el paso 4º del diagrama del Caso de uso de Generar Gráfico y del Caso de Uso de Comparativa.

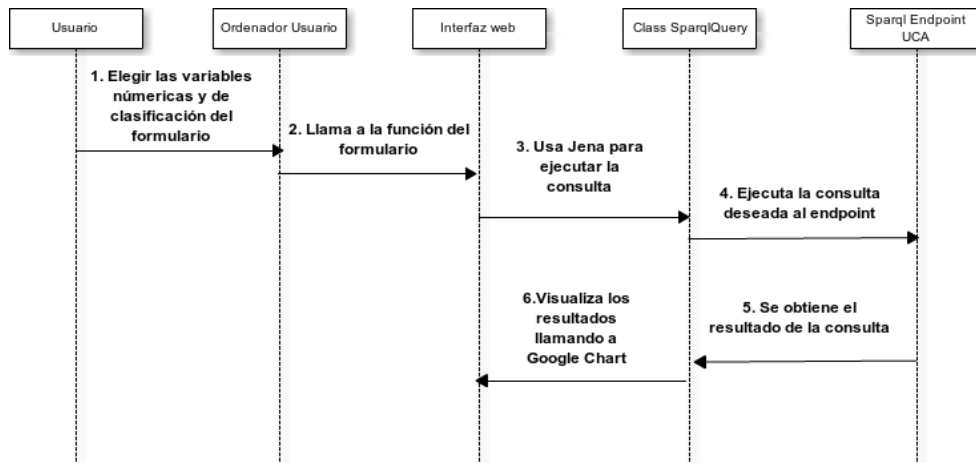


Figura 5.2: Diagrama de Secuencia del caso de uso Generar Gráfica

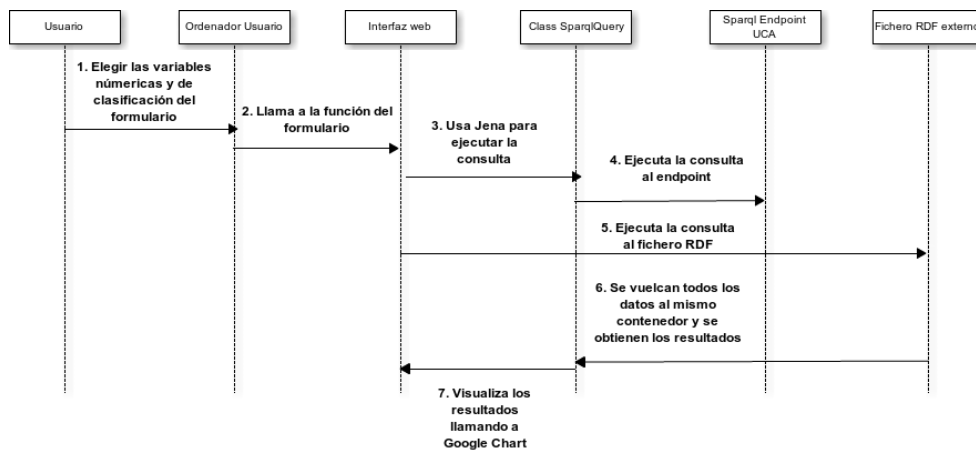


Figura 5.3: Diagrama de Secuencia del caso de uso Comparativa

## 5.2. Modelo de interfaz de usuario

Durante el diseño de la interfaz de usuario de la aplicación de consulta de datos, se realizaron diferentes bocetos con distintas aproximaciones para presentar la información y la navegación a través de la aplicación. Tras la evaluación pertinentes se decidió dar como definitivos los siguientes bocetos que se presentan.

### 5.2.1. Boceto de la aplicación web

El boceto 5.4 presenta la interfaz de la página principal de la aplicación, que sirve a modo de portada y menú de navegación principal del sitio.



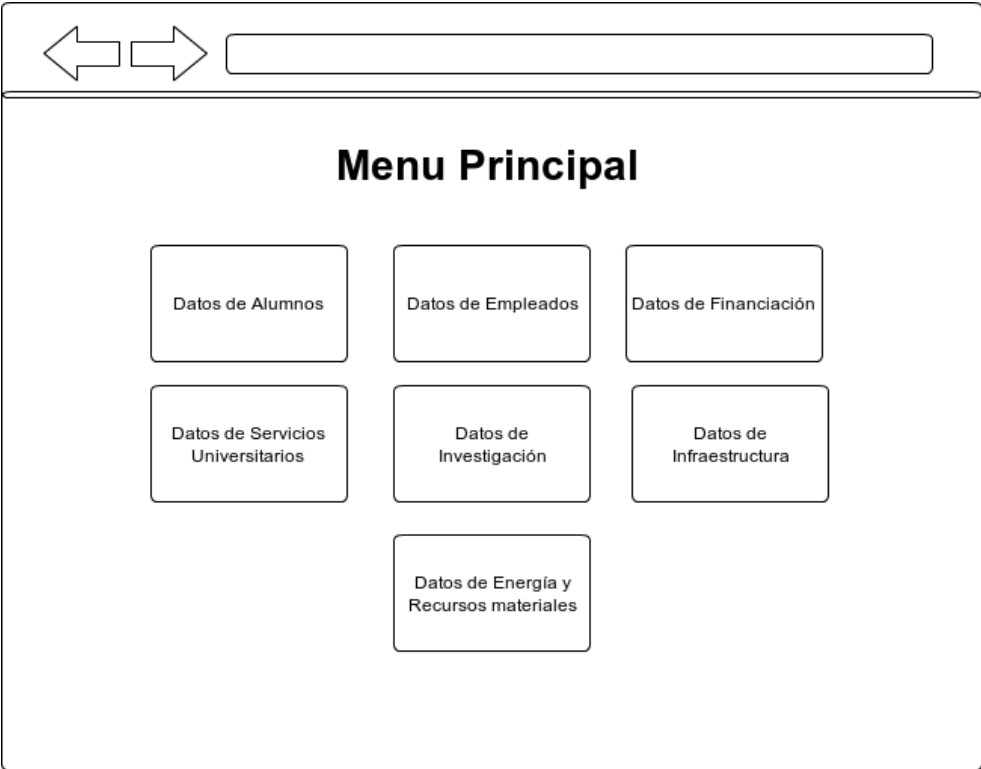


Figura 5.4: Boceto del Menu de la aplicación

El boceto 5.5 presenta la pantalla de consulta y muestra de resultados al usuario en el formato seleccionado.



Figura 5.5: Boceto del Formulario de la aplicación

## Capítulo 6

# Ingeniería Ontológica

La Ingeniería Ontológica se refiere al conjunto de actividades de desarrollo de una ontología, el ciclo de vida de una ontología, así como a las metodologías, herramientas y lenguajes necesarios para la construcción de ontologías.

### 6.1. ¿Qué es una ontología y qué razones podemos tener para implementarla?

Una ontología define el vocabulario común para investigadores que necesitan compartir información en un dominio. En ella se definen los conceptos y las relaciones que se establecen entre los conceptos para que puedan ser interpretadas por una máquina. Algunas de las razones por las que deseáramos desarrollar una ontología son las siguientes:

- Compartir el conocimiento común sobre la estructura de la información entre personas y máquinas. Supongamos que existen varios sistemas Web sobre información médica o provean servicios de e-commerce médico. Si los sitios Web comparten y publican siguiendo una ontología subyacente de los términos que se usan, entonces los agentes software podrían extraer y agregar información, además de utilizar los datos de entrada para realizar aplicaciones que exploten estos datos. Aplicados al caso de este proyecto, sería una manera de unificar los sistemas de información de las distintas universidades del país, ya que todas tienen que publicar los datos de la universidad de cada año. Sería una buena práctica conseguir que el sistema información de donde se sacan los datos, contase con una ontología subyacente que permitiese el intercambio de información y comparación de datos entre distintas universidades que permita un análisis más rápido de los datos.
- Permitir la reutilización del conocimiento sobre un determinado dominio. Para construir ontologías nuevas podemos reutilizar las existentes que describen porciones del dominio más grande o también podemos basarnos en un ontología general y extenderla para describir el dominio que nos interesa.
- Explicitar suposiciones de un dominio. Las especificaciones son útiles para nuevos usuarios que deben aprender el significado de cada término del dominio.
- Separar el conocimiento del dominio del conocimiento operacional. Es separar el conocimiento del dominio del conocimiento sobre cómo se fabrica o la forma en la que se opera. Un ejemplo puede ser describir una tarea de configuración de un producto a partir de sus

componentes de acuerdo a especificaciones requeridas e implementar un programa que haga independiente esta configuración de los productos y componentes en sí.

- Analizar el conocimiento de un dominio. Para desarrollar una ontología es muy importante analizar el conjunto de datos para extraer estructuras conceptuales de estos, permitiendo así la reutilización de las ontologías existentes y también poder extenderlas.

Las ontologías están formadas por conceptos de un dominio de discurso llamadas clases o conceptos, propiedades o slots de cada concepto describiendo las características y atributos del concepto y las restricciones o facetas de las propiedades.

### Ventajas de las ontologías

- Desde el punto de vista de la recuperación de datos, facilitan el empleo simultáneo de distintas bases de datos o almacenes de información que pueden ser consultados simultáneamente de una forma homogénea para recuperar la información que deseamos obtener.
- La posibilidad de utilizar la información preexistente en dicha base de datos a través de la ontología sin tener que renunciar a la base de datos de partida, conviviendo base de datos y ontología simultáneamente.

## 6.2. Diseño de vocabularios

Dado que una de las principales premisas en el empleo de Linked Open Data es la reutilización de vocabularios existentes, con el fin de evitar duplicidades y una descripción homogénea de las entidades, cualesquiera que sea su fuente, antes de diseñar un vocabulario específico es preciso realizar un análisis de los vocabularios ya existentes.

Para este proyecto, nos hemos basado en los siguientes vocabularios ya existentes:

### 6.2.1. DBpedia Ontology

El DBpedia<sup>1</sup> Ontología es una ontología que ha sido creado de forma manual basado en los tópicos más utilizados de Wikipedia. La ontología actualmente cubre 685 clases que forman una jerarquía y se describen por 2.795 propiedades diferentes.

#### Términos usados

- dbpedia:year.
- dbpedia:statisticValue.
- dbpedia:headquarter.
- dbpedia:president.

---

<sup>1</sup><http://wiki.dbpedia.org/Ontology>

### 6.2.2. The Friend of a Friend (FOAF) project

Es una ontología<sup>2</sup> legible para las máquinas que describe a las personas, sus actividades y sus relaciones con otras personas y objetos.

#### Términos usados

- foaf:age.
- foaf:gender.
- foaf:name.
- foaf:workplaceHomepage.

### 6.2.3. Dublin Core Metadata

Es un vocabulario<sup>3</sup> que permite describir tanto recursos físicos como recursos web.

#### Términos usados

- dcterms:description.

### 6.2.4. Teaching Core Vocabulary

Es un vocabulario<sup>4</sup> ligero que ofrece condiciones para que los profesores puedan relacionar términos con los cursos. Este vocabulario se basa en los términos para describir los seminarios y descripciones de cursos como Linked Data.

#### Términos usados

- ns:academicTerm.

### 6.2.5. CERIF Semantic Vocabulary 1.3

Es un vocabulario<sup>5</sup> que representa la relación de términos para describir el ámbito de la investigación. Muchos de estos términos se han tomado desde el editor del grupo SPI-FM de la Universidad.

#### Términos usados

- cerif:PriceAwards.
- semcerif:awarded.
- semcerif:Funding.
- semcerif2:DoctoralThesis.
- semcerif2:Doctorate.

---

<sup>2</sup><http://www.foaf-project.org/>

<sup>3</sup><http://dublincore.org/documents/dcmi-terms/>

<sup>4</sup><http://linkedscience.org/teach/ns/>

<sup>5</sup><http://www.eurocris.org/ontology/semcerif>

### 6.2.6. Academic Research Project Funding Ontology (ARPFO)

Este vocabulario<sup>6</sup> ofrece clases y propiedades para describir la estructura de la financiación de proyectos de investigación académica, y también ofrece clases y propiedades para codificar las relaciones entre las ofertas de proyectos, proyectos, grupos de investigación y las salidas resultantes.

#### Términos usados

- arpfo:Funding.

### 6.2.7. An Agent Certification Ontology

Este vocabulario<sup>7</sup> permite expresar en Linked Data la existencia de avales o certificaciones de agentes oficiales.

#### Términos usados

- acrt:qualification.

### 6.2.8. The Lingvoj Ontology

Un vocabulario<sup>8</sup> para describir el uso de las lenguas por las personas y las organizaciones, su ámbito geográfico y el estado, así como su uso en documentos o durante eventos.

#### Términos usados

- lingvo:Language.

Tras este paso usaremos la siguiente herramienta para desarrollar los vocabularios.

### 6.2.9. Neologism

Es una plataforma basada en Drupal donde se publica el vocabulario para usarlo en la Web de datos, ágil y fácil de utilizar además de compatibilizar con los principios de Linked Data. Neologism es gratuito y de código abierto<sup>9</sup>.

Con esta herramienta desarrollaremos los distintos vocabularios RDF para informar sobre los resultados de la Universidad y desarrollando las propiedades y clases respectivas de cada uno de ellos.

---

<sup>6</sup><http://vocab.ox.ac.uk/projectfunding>

<sup>7</sup><http://privatealpha.com/ontology/certification/1#>

<sup>8</sup><http://www.lingvoj.org/ontology#>

<sup>9</sup><http://neologism.deri.ie/>

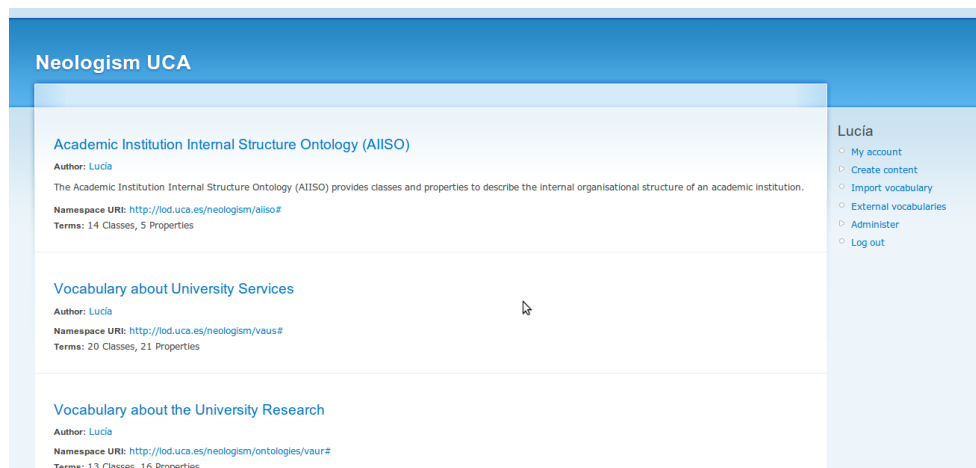


Figura 6.1: Menu principal de neologism

Para ver más detalles sobre la implementación de cada uno de los vocabularios RDF, se debe acceder a la instancia de Neologism<sup>10</sup> montada en el servidor LOD.

### 6.3. Vocabularios desarrollados

Tras realizar un estudio de los vocabularios que podemos desarrollar, necesitamos ofrecer más información de la que permiten los vocabularios anteriores, por lo que entonces se ha tenido que diseñar una serie de vocabulario nuevos. Para el diseño de los vocabularios se ha empleado una serie de guías y recomendaciones de buenas prácticas que se resumen en los siguientes pasos:

- **Determinar el dominio y alcance de la ontología.** Comprender cuál es el campo de la aplicación de vocabulario RDF que se ha generado, su uso y fomentar la generalización de los términos con el objetivo de que puedan ser reutilizados.
- **Considerar la reutilización de ontologías existentes.** Tal y como se comenta en la sección 6.2.
- **Enumerar términos importantes para la ontología.** En nuestro caso, los términos importantes serán cada una de las tablas de la base de datos y sus respectivas columnas puesto que la aplicación del D2R Server mapeará la base de datos como un grafo y necesitaremos un término para cada dato y columna de cada tabla.
- **Definir las clases y la jerarquía de clases.** Cada ontología puede tener tres tipos de proceso de desarrollo: *top-down*, *bottom-up* y *combinado*. En el desarrollo de nuestros vocabularios, se ha seguido un proceso de desarrollo combinado donde en algunos casos se ha desarrollado de los conceptos más específicos a los más generales como en el caso del vocabulario sobre energía y recursos y sin embargo para el vocabulario RDF sobre alumnos y conocimientos universitarios se ha realizado el proceso contrario de lo más general a lo más específico debido a que el número de términos necesarios era mayor y por lo tanto era más fácil establecer las relaciones entre las clases.

<sup>10</sup><http://lod.uca.es/neologism/>

- **Definir las propiedades de las clases: slots.** Tras haber definido las clases se han creado las propiedades, buscando siempre poder reutilizar el mayor número de propiedades ya creadas cumpliendo así con el principio de reutilización de vocabularios que fomenta la tecnología Linked Data.
- **Definir las facetas de una propiedad.** Se tuvo que realizar un análisis que tipo de dato es cada una de las propiedades que se han creado, es decir, si el valor de la propiedad es un entero, un valor estadístico, etc.
- **Crear instancias.** Este paso se ha realizado cuando ya existen recursos que estén definidos por las clases y propiedades que se han reutilizado o creado previamente.

Además hemos realizado el vocabulario pensando en su generalización para que pueda ser utilizados por terceros.

Los vocabularios desarrollados y aplicados a la Base de datos anterior, se pueden clasificar en 5 dominios distintos:

### 6.3.1. Vocabulary about University Services

Este vocabulario especifica todos aquellos términos que hacen referencia a cualquier tipo de servicio que ofrece la universidad.

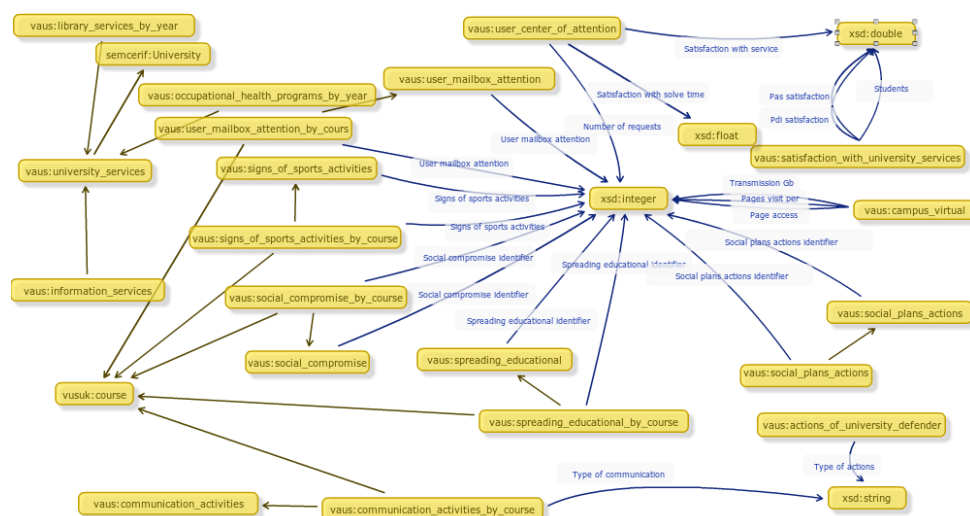


Figura 6.2: Vocabulario de Servicios Universitarios

A continuación se describen las clases y propiedades del vocabulario RDF:

#### Clases

- `actions_of_university_defender`. Número de cada tipo de actuación del defensor universitario.



- `campus_virtual`. Entorno web dedicado a la docencia facilitando a los alumnos el desarrollo de sus estudios.
- `communication_activities`. Descripción de la temática de los mensajes del tablón virtual de anuncios.
- `communication_activities_by_course`. Porcentaje de cada una de las temáticas de los mensajes en el tablón de anuncios en cada curso académico.
- `information_services`. Número de servicios que ofrece la universidad con el objetivo de informar a la comunidad universitaria.
- `library_services_by_year`. Los servicios de las diferentes bibliotecas de la universidad a la comunidad universitaria.
- `occupational_health_programs_by_year`. Número de programas de salud en la universidad.
- `satisfaction_with_university_services`. Valoración de satisfacción de los servicios universitarios de los alumnos, pas y pdi en un curso académico.
- `signs_of_sports_activities`. Descripción de los indicadores de actividad y satisfacción con las actividades deportivas.
- `signs_of_sports_activities_by_course`. Valor de cada uno de los indicadores de actividad o satisfacción con las actividades deportivas cada curso académico.
- `social_compromise`. Descripción de cada indicador de las actividades de acción social y solidaria desarrolladas en la universidad.
- `social_compromise_by_course`. Descripción de las acciones del plan de acción social.
- `social_plans_actions_by_year`. Número de indicadores de actividades de acción social y solidaria en cada curso académico.
- `spreading_educational`. Descripción de las diferentes eventos en los que se promueven su oferta educativa a los estudiantes que aún no pertenecen a la comunidad universitaria.
- `spreading_educational_by_course`. Número de actividades que tratan de difundir la educación en la universidad.
- `university_services`. Servicios a la comunidad universitaria.
- `user_center_of_attention`. Número de solicitudes en cada año, valoración del tiempo de respuesta con el servicio y el tiempo de respuesta.
- `user_mailbox_attention`. Descripción de los distintos aspectos de evaluación del buzón de atención.
- `user_mailbox_attention_by_course`. Número de mensajes al buzón de atención al usuario en cada aspecto de evaluación y el tiempo promedio de respuesta.

## Propiedades

- `activity_identifier`. El código que permite identificar un único tipo de actividad desarrollada en la universidad.
- `number_of_days`. Número de días que dura una determinada actividad.
- `number_of_requests`. Número de solicitudes enviadas al centro de la atención del usuario con el fin de que sean respondidas.
- `number_of_subjects`. Número de asignaturas que se enseñan en la universidad.
- `pages_access`. Número de accesos a la página web del campus virtual.
- `pages_visit_per_day`. Número de páginas web visitadas por días.
- `pas_satisfaction`. Valoración de satisfacción del pas con los servicios universitarios.
- `pdi_satisfaction`. Valoración de satisfacción del pdi con los servicios universitarios.
- `satisfaction_with_answer_time`. Valoración del grado de satisfacción con el servicio de respuesta del centro de atención al usuario.
- `satisfaction_with_service`. Valoración del grado de satisfacción con el servicio del centro de atención al usuario.
- `satisfaction_with_solve_time`. Valoración de la satisfacción con el tiempo de respuesta del centro de atención al usuario.
- `signs_of_sports_activities_identifier`. El código que le permite identificar un único tipo de la actividad deportiva en un curso académico.
- `social_compromise_identifier`. El código que permite la identificación de un único tipo de compromiso social de la universidad.
- `social_plans_actions_identifier`. El código que permite identificar un único tipo de plan de acción social.
- `spreading_educational_identifier`. El código que permite identificar a un único tipo de difusión educativa.
- `students_satisfaction`. Valoración de satisfacción de los estudiantes con los servicios universitarios.
- `transmission_Gb_by_day`. Gigabyte de información que se transfiere en un día.
- `type_communication`. Los diferentes tipos de comunicación entre la comunidad universitaria.
- `type_of_actions`. Tipo de actuación que realiza el defensor universitario.
- `university_services_identifier`. El código que permite identificar un único tipo de servicio universitario.
- `user_mailbox_attention_identifier`. El código que permite identificar un único aspecto de la evaluación del servicio de buzón de usuario.

### 6.3.2. Vocabulary about the University Research

Este vocabulario especifica todos aquellos términos que hacen referencia a la investigación en la universidad.

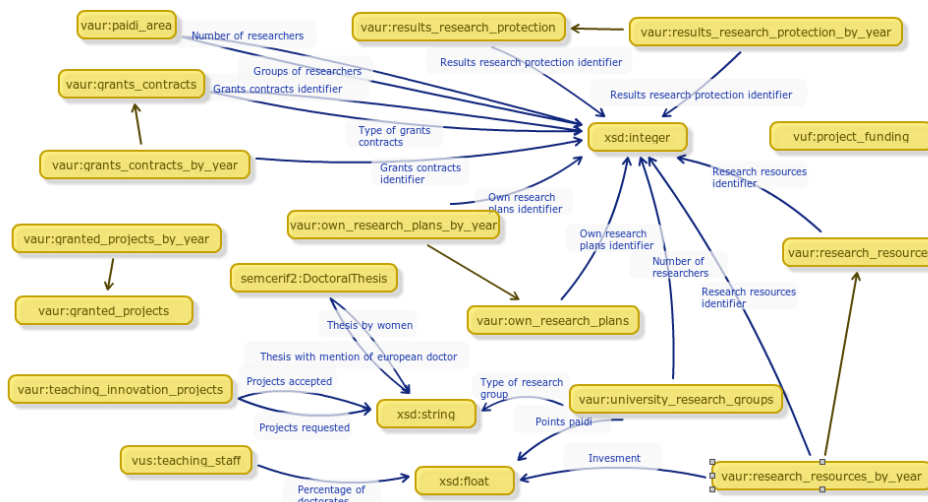


Figura 6.3: Vocabulario de Investigación Universitaria

A continuación se describen las clases y propiedades del vocabulario RDF:

#### Clases

- `granted_projects`. Descripción de los distintos proyectos.
- `granted_projects_by_year`. Aquellos proyectos que han sido subvencionados para que se pueda proseguir con su investigación.
- `grants_contracts`. Descripción del tipo de inversiones en investigación en becas y contratos para la universidad.
- `grants_contracts_by_year`. Número de inversiones (subvenciones o contratos) en la investigación universitaria en un año.
- `own_research_plans`. Descripción de las ayudas para los planes propios de investigación.
- `own_research_plans_by_year`. Números de cada plan propio de investigación de la universidad en un año.
- `paid_area`. Información sobre los recursos de cada área de estudio de los grupos de investigación que pertenecen al área andaluza de investigación.
- `research_resources`. Descripción de los recursos de los que dispone la universidad para ayudar a la investigación.
- `research_resources_by_year`. Información sobre los recursos de los que dispone la universidad para ayudar a la investigación en un año.

- `results_research_protection`. Descripción de los tipos de patentes y licencias de software que se desarrollan en la universidad con el fin de proteger los resultados de la investigación.
- `results_research_protection_by_year`. Cantidad de patentes y licencias de software que se desarrollan en la universidad en un año.
- `teaching_innovation_projects`. Este tipo de objetivo principal del proyecto es involucrar a los maestros en la adopción de los cambios metodológicos en la enseñanza que conduzcan al proceso de enseñanza-aprendizaje más efectivo.
- `university_research_groups`. Información de cada tipo de grupo de investigación de la universidad.

### Propiedades

- `grants_contracts_identifier`. El código que permite identificar a un solo tipo de subvención o contrato para la investigación en la universidad.
- `groups_of_researchers`. Número de grupos de investigadores que pertenecen al PAIDI por cada área de investigación.
- `investment`. Importe total recibido en total en cada tipo de ayuda para el plan propio de la investigación de la universidad.
- `mention_of_european_doctor`. Número de tesis presentadas que tenga la mención de doctor europeo.
- `number_of_researchers`. Número de personas que forman parte de la comunidad universitaria que se dedican a la investigación.
- `own_research_plans_identifier`. El código que permite un único tipo de plan propio de investigación.
- `percentage_doctorates`. Porcentaje de doctorados en cada año o curso académico.
- `points_paidi`. Número de puntos que recibe un grupo de investigadores de la universidad pertenecientes al PAIDI.
- `projects_identifier`. El código que permite identificar a un único proyecto de investigación perteneciente a la universidad.
- `project_accepted`. Número de proyectos que se han aprobado cumplen los requisitos publicados de mejora docente.
- `project_requested`. Número de proyectos que se han tramitado con el fin de que sean aceptados, para cumplir los objetivos publicados para de mejora docente en el curso académico.
- `research_resources_identifier`. El código que permite la identificación de un único tipo de recurso de investigación para la investigación en la universidad.
- `results_research_protection_identifier`. El código que permite la identificación de un único tipo de protección a los resultados de investigación de la universidad.

- `thesis_by_women`. Número de tesis que se han presentado por mujeres en la universidad durante un curso académico.
- `type_of_grants_contracts`. Propiedad para clasificar los tipos de inversiones entre contratos o becas.
- `type_of_research_group`. Clasificación de los grupos que pertenecen al PAIDI entre grupos competitivos o consolidados.

### 6.3.3. Vocabulary for University Infrastructure

Este vocabulario especifica todos aquellos términos que hacen referencia a la investigación en la universidad.

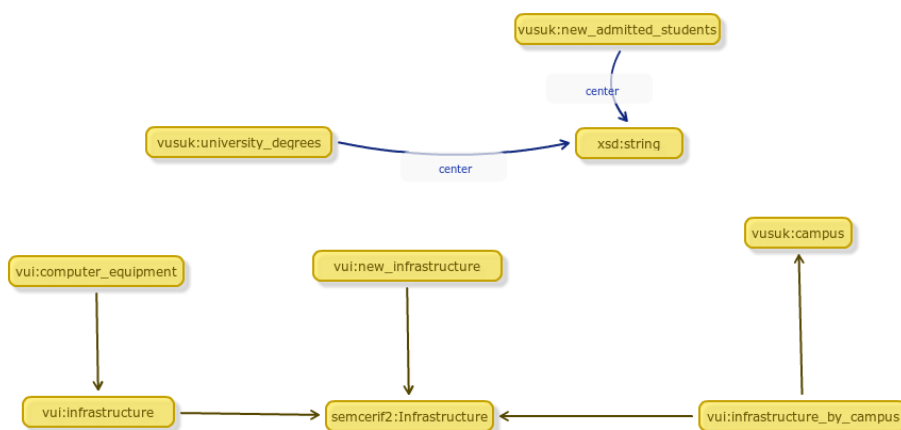


Figura 6.4: Vocabulario de Infraestructura Universitaria

A continuación se describen las clases y propiedades del vocabulario RDF:

#### Clases

- `computer_equipment`. Descripción de los distintos recursos informáticos usados para la docencia en la universidad.
- `infrastructure`. Todos los diferentes edificios que forman parte de la universidad.
- `infrastructure_by_campus`. Todos los diferentes edificios que forman parte de la universidad divididos por campus en el que se encuentra.
- `new_infrastructure`. Las nuevas construcciones y mejoras de las instalaciones que se desarrollan en un curso académico.



A continuación se describen las clases y propiedades del vocabulario RDF:

### Clases

- `conciliation_measures_working_life`. Descripción de las medidas ofrecidas por la universidad para la conciliación de la vida personal, familiar y profesional de sus trabajadores.
- `conciliation_measures_working_life_by_course`. Número de beneficiarios de cada tipo de medida de conciliación con la vida laboral.
- `hours_pas_training`. Número de horas que el personal de la universidad tienen que tomar clases de capacitación.
- `hours_pdi_training`. Número de horas que los investigadores de la universidad tienen que tomar clases de capacitación.
- `illnesses`. Descripción de los distintos tipos de enfermedades por las que se pueden dar de baja los trabajadores.
- `illnesses_by_year`. Número de empleados por sexo que han sufrido este tipo de enfermedad.
- `laboral_risk`. Los accidentes de trabajo, las enfermedades profesionales y las víctimas mortales que se pueden dar en el trabajo.
- `minimum_salary_by_year`. Salario mínimo interprofesional, según cada categoría de empleado cada año.
- `occupational_risk_prevention_courses`. Actividades formativas en salud laboral destinada a una determinada parte de la comunidad universitaria y el número de ediciones que se han impartido de los mismos.
- `pas`. Personal de administración y servicios.
- `pas_groups`. Grupos en los que se clasifica el personal de administración y servicios.
- `pas_training`. Descripción de los distintos tipos de formaciones del PAS.
- `pas_training_by_year`. Número de horas de formación recibida por el personal de administración y servicios en función del género y el grupo o escala administrativa.
- `pdi_training`. Descripción de los distintos tipos de formaciones del PDI.
- `pdi_training_by_year`. Número de horas de formación recibida por el personal docente e investigador en función del género y de la categoría.
- `retirement_incentive_recipients`. Número de empleados que son beneficiarios de los incentivos de jubilación en un año.
- `staff`. Plantilla de empleados de la universidad.
- `staff_classified_by_age`. Número de empleados clasificados por tipo, por edad y por sexo.
- `teaching_staff`. Plantilla de empleados que se dedican a enseñar en la universidad.
- `university_structures`. Las distintas estructuras de las que se compone la universidad.

### Propiedades

- `conciliation_measures_working_life_identifier`. El código que permite la identificación de un único tipo de medida de conciliación de la vida laboral.
- `illness_identifier`. El código que permite la identificación de un único tipo de enfermedad.
- `minimum_interprofesional_wage`. Coeficiente del salario mínimo interprofesional.
- `number_of_beneficiaries`. Número de personas que se benefician de una determinada actividad.
- `number_of_employees`. Número de personas que pertenece a los empleados de la universidad.
- `number_of_hours`. Número de horas de una determinada actividad.
- `number_of_pas`. Número de personas del PAS.
- `number_of_pdi`. Número de personas del PDI.
- `number_of_sick`. Número de enfermos.
- `number_of_teachers`. Número de personas que se dedican a la docencia en la universidad.
- `pas_groups_identifier`. El código que permite la identificación de un único tipo de grupo de PAS.
- `pas_training_identifier`. El código que permite identificar un único tipo de formación del pas.
- `pdi_training_identifier`. El código que permite identificar un único tipo de formación del pdi.
- `staff_identifier`. El código que permite un único tipo de perfil de personal de la universidad.
- `type_university_structure`. Tipo de estructura universitaria.
- `type_work_time`. Especificación de el tipo de jornada laboral de los trabajadores si es completa, a tiempo parcial o de sustitución.
- `university_structures_identifier`. El código que permite identificar un único tipo de estructura universitaria.



### 6.3.5. Vocabulary for University Funding

Este vocabulario especifica todos aquellos términos que hacen referencia a la financiación de la universidad.

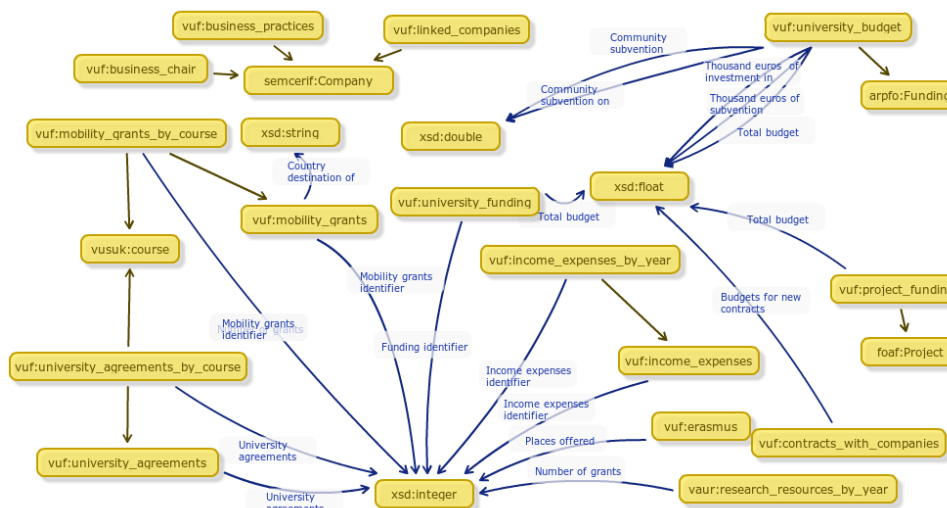


Figura 6.6: Vocabulario de Financiación Universitaria

A continuación se describen las clases y propiedades del vocabulario RDF:

#### Clases

- `business_chair`. La manera de establecer una relación entre empresas, fundaciones y otras entidades que están vinculadas a la universidad.
- `business_practices`. El trabajo a desarrollar en las empresas por los estudiantes que pertenecen a cualquier tipo de grado.
- `contracts_with_companies`. Relación de las empresas con la universidad.
- `erasmus`. Programa de acción de la Comunidad Europea para la Movilidad de Estudiantes Universitarios.
- `income_expenses`. Descripción de los gastos e ingresos de la universidad.
- `income_expenses_by_year`. Valor en millones de euros de gastos e ingresos que se realizan en un determinado año del presupuesto de la universidad.
- `linked_companies`. Aquellas empresas a través de las cuales se apoyan a jóvenes empresas innovadoras, que se dedican a los resultados de investigación de la universidad.
- `mobility_grants`. Descripción de cada tipo de beca de movilidad.
- `mobility_grants_by_course`. Número de cada tipo de beca de movilidad por curso académico.

- `project_funding`. Financiación recibida para cada tipo de proyecto en un año.
- `university_agreements`. Descripción de cada tipo de convenio firmado por la universidad.
- `university_agreements_by_course`. Número de cada tipo de convenio firmado por la universidad en un curso académico.
- `university_budget`. El presupuesto total de la universidad, las donaciones y las inversiones en la investigación por parte de la comunidad autónoma.
- `university_funding`. El presupuesto recibido de diversas fuentes de financiación universitaria cada año.

### Propiedades

- `budgets_for_new_contracts`. Presupuesto asignado para los nuevos contratos formalizados entre empresas y la universidad en un año.
- `community_subvention`. Presupuesto que recibe la universidad procedente de la comunidad autónoma. En este caso, de la Junta de Andalucía.
- `community_subvention_on_research`. Presupuesto que recibe la universidad procedente de la comunidad autónoma destinado a la investigación.
- `country_destination_of_grants`. País destinado el tipo de beca de movilidad.
- `euros`. Número de euros.
- `funding_identifier`. El código que permite identificar a un único tipo de financiación.
- `income_expenses_identifier`. El código que permite la identificación de un único tipo de ingresos o gastos de la universidad.
- `million_euros`. Millón de euros.
- `mobility_grants_identifier`. El código que permite identificar a un único tipo de beca de movilidad.
- `number_of_grants`. Número de becas o subvenciones destinadas a un propósito específico.
- `places_offered`. Numero de plazas ofertadas para un determinado objetivo.
- `thousand_euros`. Miles de euros.
- `thousand_euros_research_investment`. Valor de cada inversión en investigación de la universidad en miles de euros.
- `thousand_euros_subvention`. Valor de cada subvención de la universidad en miles de euros.
- `total_budget`. Presupuesto total para un determinado fin.
- `university_agreements_identifier`. El código que permite identificar a un único convenio entre empresa y la universidad.



- `doctorate_by_branch_of_knowledge`. Los doctorados divididos por rama de conocimiento a la que pertenece.
- `doctorate_by_course`. Doctorado por curso.
- `educational_satisfaction`. Descripción de los distintos aspectos consultados para medir la satisfacción de los alumnos con la docencia.
- `educational_satisfaction_by_course`. Valoración de cada aspecto consultado para medir la satisfacción de los alumnos con la docencia en cada curso académico.
- `employment`. Descripción de los aspectos analizado sobre la inserción laboral de sus antiguos alumnos según la fecha de comienzo y finalización de sus estudios.
- `employment_by_course`. Porcentaje de cada aspecto analizado sobre la inserción laboral de sus antiguos alumnos según la fecha de comienzo y finalización de sus estudios.
- `foreign_students`. Estudiantes extranjeros.
- `foreign_students_by_course`. Número de estudiantes en cada clase de alumnos extranjeros por cada curso.
- `full_time_students`. Número de estudiantes que están estudiando una titulación en la universidad en un curso académico.
- `gender_relations`. Descripción de las incidencias producidas por las relaciones entre sexos en la comunidad universitaria.
- `gender_relations_by_course`. Número de cada una de las incidencias producidas por las relaciones entre sexos en la comunidad universitaria por cada curso.
- `graduate_students`. Número de estudiantes que han terminado sus estudios en un curso académico por cada rama de conocimiento.
- `languages_certifications`. Número de cursos, idiomas y certificaciones por cada año.
- `new_admitted_students`. Número de estudiantes por rama de conocimiento en cada curso, según el tipo de centro donde van a cursar los estudios.
- `non-formal_education`. La educación no formal es aquella con la que se obtiene un grado que no es reconocido por un organismo oficial. Es lo que incluye todas las enseñanzas, lecciones, cursos o seminarios sobre diversos temas que se pueden hacer para comenzar y especializarse en un tema o simplemente como un hobby.
- `older_classroom_enrolled`. Número de personas que han comenzado el aprendizaje de alguna titulación o actividad universitaria que supera la un cierto periodo de edad, por cada campus en cada curso académico.
- `older_classroom_graduate`. Número de personas que han terminado el aprendizaje de alguna titulación o actividad universitaria que supera un cierto periodo de edad, por cada curso en un curso académico.
- `students_activities`. Número de estudiantes que desarrollan una actividad de educación no formal y el tipo de presencia de los estudiantes en la actividad en cada año.

- `students_by_campus`. Número de estudiantes por cada campus.
- `students_by_school`. Número de alumnos de centros propios, alumnos de centros adscritos, número de universidades de la comunidad y universidades del país.
- `students_classified_by_age`. Número de estudiantes y porcentaje que representa sobre el total de alumnos que se clasifican según la titulación y el periodo de edad al que pertenecen.
- `students_enrolled`. Número de estudiantes que han proseguido sus estudios en cada curso académico clasificados por rama de conocimiento y tipo de titulación.
- `university_access`. Los diferentes caminos desde los que se puede acceder a la universidad.
- `university_access_by_year`. Número y porcentaje de alumnos que aprueban las pruebas de acceso a la universidad, y número de alumnos seleccionados en cada campus, año y tipo de acceso a la universidad.
- `university_activities`. Número de actividades que no forman parte de la educación reglada impartida por la universidad y el tipo de presencia que se requiere en dicha actividad.
- `university_courses`. Número de cursos universitarios dedicados a cada titulación separados por ramas de conocimiento.
- `university_degrees`. La descripción de las titulaciones universitarias que se enseñan en la Universidad. Esta descripción contiene el nombre del título, la rama de conocimiento de la misma, el campus y el centro universitario donde se enseña.
- `university_profile`. Perfil de la universidad.
- `women_progress`. Porcentaje de alumnas o mujeres graduadas en la universidad.

### Propiedades

- `academic_performance_identifier`. El código que permite un único tipo de rendimiento académico.
- `approved_percentage`. Porcentaje de alumnos que superen las pruebas de acceso a la universidad.
- `attached_college`. Número de estudiantes universitarios que reciben un determinado tipo de educación en un centro privado.
- `branch_of_knowledge_identifier`. El código que permite identificar un único tipo de conocimiento que se enseña en la universidad en la universidad.
- `campus_identifier`. El código que permite identificar un único tipo de campus universitario.
- `country_universities`. Número de estudiantes universitarios que pertenecen al país donde se encuentra la universidad.
- `degree_identifier`. El código que permite identificar a una única titulación.
- `doctorate_identifier`. El código que permite identificar un único doctorado que se imparte en la universidad.

- `educational_satisfaction_identifier`. El código que permite identificar un simple aspecto para la evaluación de la satisfacción educativa.
- `employment_identifier`. El código que le permite identificar un único aspecto analizado de la inserción laboral de sus estudiantes graduados.
- `foreign_students_identifier`. El código que permite identificar un único tipo de alumnos extranjeros.
- `gender_relations_identifier`. El código que permite identificar un único tipo de relaciones entre sexos.
- `incoming_students`. Número de estudiantes extranjeros que vienen al país para estudiar o desarrollar alguna actividad en la universidad.
- `language_identifier`. El código que permite identificar un solo idioma del que se puede obtener una certificación.
- `non-formal_education_identifier`. El código que permite identificar un único tipo de educación no reglada.
- `number_of_certifications`. Número de certificaciones de idiomas.
- `number_of_first_cycle_students`. Número de estudiantes de primer ciclo.
- `number_of_members`. Número de miembros o entidades de una clase.
- `number_of_second_cycle_students`. Número de estudiantes de segundo ciclo.
- `number_of_students`. Número de estudiantes.
- `outgoing_students`. Número de estudiantes de la universidad que se van al extranjero para estudiar o desarrollar alguna actividad en una universidad extranjera.
- `own_college`. Número de estudiantes universitarios que recibe cierta educación en un centro universitario.
- `pass`. Número de estudiantes que aprueban las pruebas de acceso a la universidad.
- `percentage_women`. Porcentaje de cada aspecto de la evaluación del sector femenino, que es parte de la universidad en cada curso académico.
- `percentage_women_graduate`. Porcentaje de mujeres graduadas en la universidad en cada curso.
- `percentage_women_student`. Porcentaje de mujeres estudiantes en la universidad en cada curso.
- `presence`. Tipo de presencia de un curso o asignatura en la que se imparte en la universidad.
- `promotion`. La fecha en que los estudiantes han terminado sus estudios.
- `selected`. Número de personas que se presentaron a las pruebas de acceso y han sido seleccionadas para acceder a la universidad.

- submitted. Número de personas que se presentan a las pruebas de acceso a la universidad.
- universities\_community. Número de estudiantes universitarios que pertenecen a la comunidad donde se encuentra la universidad.
- university\_access\_identifier. El código que permite identificar un único tipo de acceso universitario de los alumnos.
- university\_community. Todas las personas o grupo de ellas que participan en las actividades universitarias.

### 6.3.7. Vocabulary for Energy Efficiency and Primary Resources

Este vocabulario especifica todos aquellos términos que hacen referencia al consumo de energía y recursos que realizan la universidad, así como la eliminación de residuos.

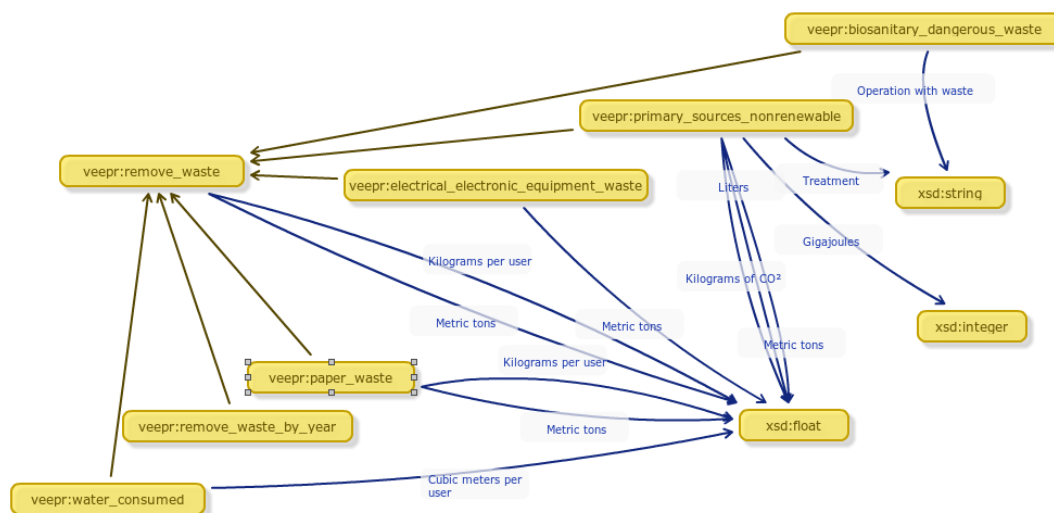


Figura 6.8: Vocabulario de Eficiencia y Recursos primarios

A continuación se describen las clases y propiedades del vocabulario RDF:

#### Clases

- biosanitary\_dangerous\_waste. Describe el tratamiento y el tipo de operación que una organización le da a sus residuos biosanitarios.
- electrical\_electronic\_equipment\_waste. Los aparatos eléctricos y electrónicos que son recolectados para ser reciclados por una organización.
- paper\_waste. Los residuos de papel que se reciclan por una organización.
- primary\_sources\_nonrenewable. Consumo indirecto de energía por fuentes primarias no renovables en un año, que describe el tipo de energía consumida por la organización, las unidades necesarias para expresar su consumo y su impacto en el ambiente.

- `remove_waste`. Describe los datos de como eliminar cualquier residuo en una organización.
- `remove_waste_by_year`. Cantidad de residuos que se eliminan al año en la universidad.
- `water_consumed`. Describe el número de metros cúbicos por usuario consumido en un campus universitario.

### Propiedades

- `cubic_meter_per_user`. Unidad del sistema de medida que le permite conocer el volumen de los líquidos que se consume por usuario en una organización.
- `cubic_meters`. Unidad del sistema de medida que le permite conocer el volumen de los líquidos.
- `gigaJoules`. Unidad de medida que pertenece al Sistema Internacional de Unidades, equivale a un millón de julios.
- `kilograms_of_CO2`. Unidad de medida para medir el impacto y la contaminación que se produce en la atmósfera.
- `kilograms_per_user`. Unidad de medida que permite conocer la masa del los residuos recogidos por usuario.
- `liters`. Unidad de medida que permite conocer la capacidad de fuentes primarias no renovables.
- `metric_tons`. Unidad de medida que permite saber la capacidad de un recurso.
- `operation_with_waste`. Describe la operación del tratamiento del residuo.
- `treatment`. Describe la operación del tratamiento del residuo.
- `waste_identifier`. Código que deja identificar un único tipo de residuo.



## Capítulo 7

# Diseño del Sistema

En este capítulo se recoge la arquitectura de la plataforma e iremos describiendo cada uno de los componentes del sistema: el proceso ETL, el almacén de datos, la publicación de datos RDF y por último el diseño de la aplicación de consulta de datos.

### 7.1. Arquitectura

A continuación se muestra con mayor nivel de detalle los requisitos y las consideraciones de diseño que se han tenido en cuenta a la hora de estructurar el sistema de información. Por ello, partiendo de una visión a nivel global proporcionada en los capítulos anteriores, diferenciaremos los componentes que forman la capa física de la lógica.

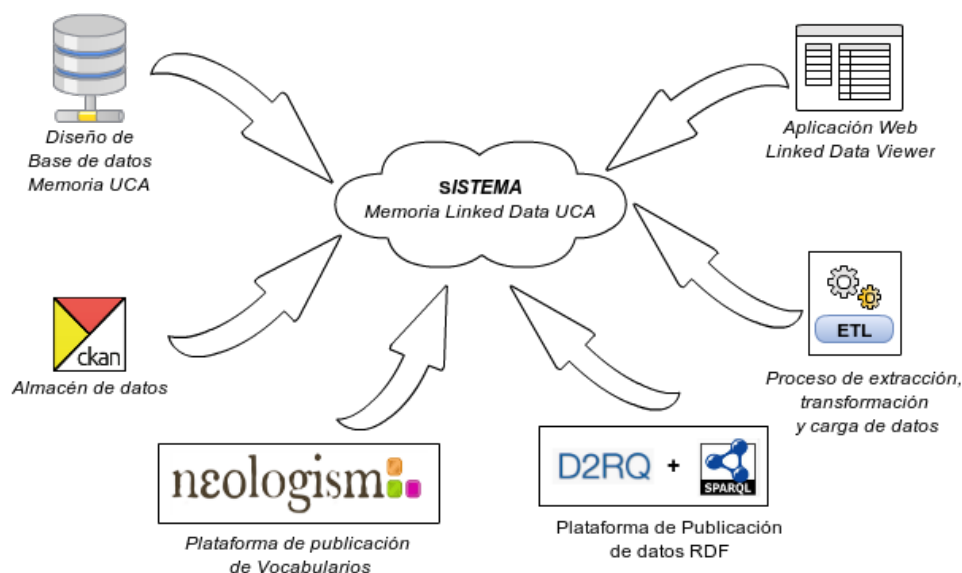


Figura 7.1: Vista general del sistema Linked Data

#### 7.1.1. Arquitectura física

El presente proyecto no requiere de ningún requisito especial en cuanto a requisitos físicos, fuera aparte de un sistema de cómputo tradicional capaz de ejecutar un sistema GNU/Linux compa-

tible con todas las aplicaciones y tecnologías que se mencionan en la presente memoria.

Es preciso indicar que el proyecto ha sido desarrollado usando un equipo propiedad del departamento de SPI-FM para este propósito, en el cual se han realizado todas las pruebas mencionadas. No obstante, se han hecho todos los esfuerzos posibles para evitar cualquier tipo de dependencia en cuanto a sistema y arquitectura, con el fin de posibilitar el despliegue del proyecto en la infraestructura de servidores propia de la Universidad de Cádiz, radicalmente diferente de la usada en la fase de desarrollo.

Debido a que este proyecto hace un uso relativamente intensivo de la plataforma Java, conocida por requerir una cantidad de recursos moderada para poder ejecutar aplicaciones con cierta soltura, se estima que la capacidad del servidor de producción deberá ser algo más elevada que la del usado durante el desarrollo. Hoy día, dado el reducido coste de la memoria y la potencia de las CPU actuales, esto no debería incurrir en costes significativamente más altos que los de un despliegue en un servidor más comedido en cuanto a recursos.

Según unas estimaciones iniciales, el servidor debería disponer como mínimo de 4 GiB de memoria de trabajo y varios procesadores a fin de poder atender las peticiones de forma más eficiente. A pesar de ello, debido a la alta modularidad del sistema, llegado el momento no debería presentar muchos inconvenientes separar los componentes del sistema en varias máquinas independientes según su función, pudiendo por ejemplo albergar la base de datos en un sistema cuya arquitectura y configuración del sistema estén optimizados para su modo de funcionamiento, con acceso a disco más rápido, pues se beneficiaría mucho más de esta mejora de lo que pudiera hacerlo el servidor web, cuyas lecturas a disco son más ocasionales, separando así la carga de una sola máquina y dividiendo el trabajo. De todas formas, esta decisión deberá ser estudiada detenidamente por el personal a cargo de la infraestructura informática del centro y no corresponde al ámbito de este proyecto.

En cualquier caso, como para cualquier aplicación o servicio, una monitorización constante es vital para poder identificar fallos de forma preventiva, sobre todo en los casos en los que el número de usuarios comience a crecer, llegando a los límites de la máquina.

### 7.1.2. Arquitectura lógica

La arquitectura lógica del sistema está formada por los elementos software (servicios, aplicaciones, librerías, frameworks, etc.) que componen el software base, más el software desarrollado para cumplir los requisitos de la aplicación.

#### D2R-Server

D2R-Server es una herramienta para publicar contenido de bases de datos relacionales en la Web Semántica. Este servidor hace uso de HTML y RDF para que los navegadores (como Google Chrome, Mozilla, Opera, etc.) les permita navegar por el contenido de la base de datos de la Web Semántica y también permite hacer consultas usando SPARQL.

### Ckan

Es una herramienta Open Source, que consiste en un poderoso manejador de contenido cuyo objetivo es permitirnos crear un portal de datos abiertos. Esto es posible gracias a que se proporcionan herramientas para agilizar la edición, la búsqueda, compartir los datos y usarlos. Está dirigido a los editores de datos como los gobiernos nacionales y regionales, empresas y organizaciones que quieran hacer sus datos abiertos.

### Apache Tomcat

Apache Tomcat<sup>1</sup> es un servidor web, aunque es más conocido por su capacidad de servir aplicaciones Java. Proporciona la capacidad de servir tanto D2RQ como la aplicación de visualización, desplegadas como un fichero WAR (Web application ARchive).

### Apache Jena

Apache Jena<sup>2</sup> es un framework de desarrollo para el lenguaje Java que proporciona las herramientas necesarias para trabajar de forma fácil con Linked data y web semántica.

### Neologism

Neologism<sup>3</sup> es una plataforma de edición y publicación de ontologías basado en el sistema de gestión de contenidos Drupal, que permite la creación de vocabularios, clases y propiedades RDF de una forma fácil e interactiva.

### 7.1.3. MySQL

MySQL<sup>4</sup> es un sistema de gestión de base de datos relacionales (RDBMS) propiedad de Oracle Corp. ofrecido como software libre en una de sus versiones, y uno de los sistemas más populares actualmente para desarrollo de aplicaciones web.

## 7.2. Procesos ETL

Se ha desarrollado un proceso ETL (Extract, Transform and Load) con el fin de unificar la información de la que dispone la Universidad de Cádiz que actualmente se encuentra dispersa en diferentes almacenes y bases de datos en distintas localizaciones. Esto nos permitirá centralizar y, por tango, agilizar y simplificar el acceso a los mismos a la hora de su consulta por parte de la plataforma Linked Open Data.

---

<sup>1</sup><http://tomcat.apache.org/>

<sup>2</sup><https://jena.apache.org/>

<sup>3</sup><http://neologism.deri.ie/>

<sup>4</sup><http://www.mysql.com/>

Con el fin de mostrar gráficamente el diseño del ETL, tenemos el siguiente diagrama 7.2:

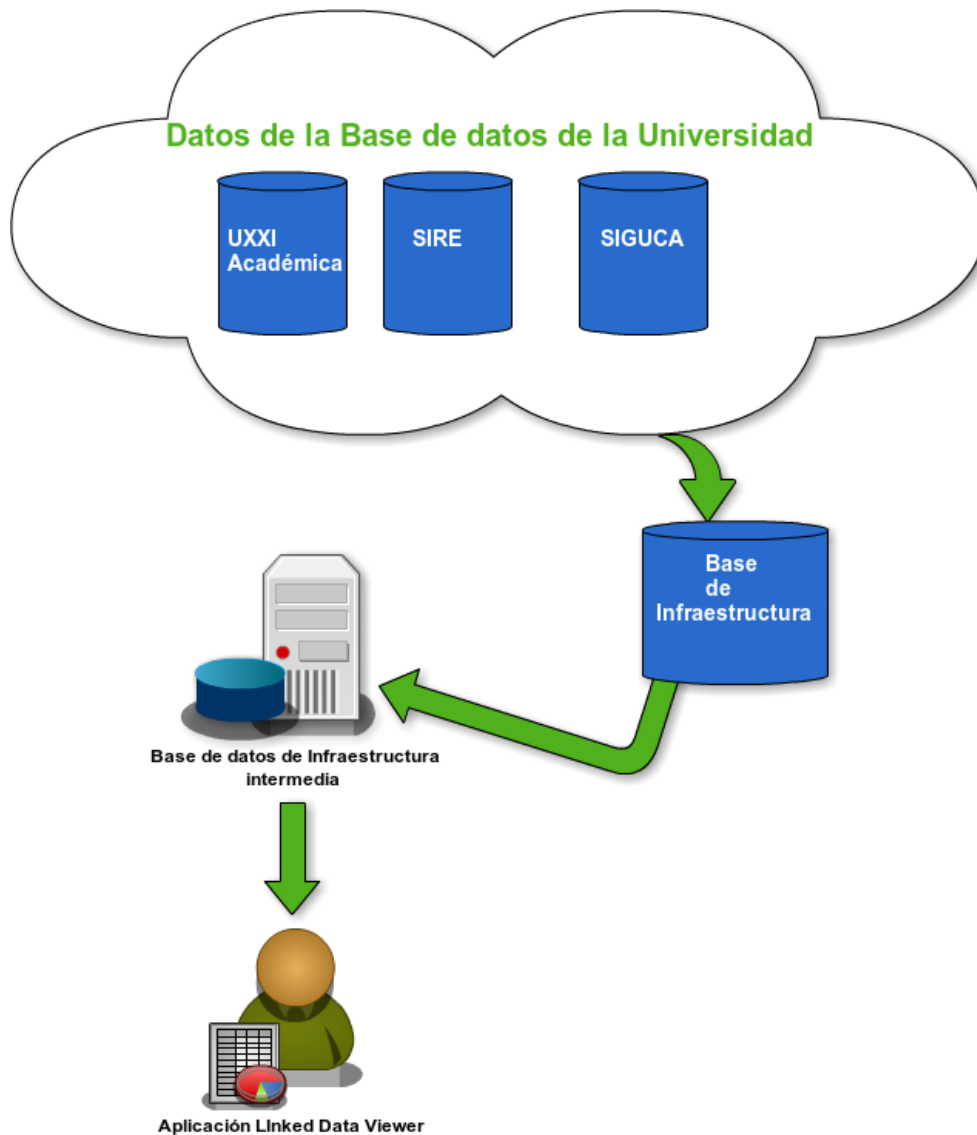


Figura 7.2: Diagrama del Proceso Completo

### 7.3. Almacén de datos

El D2RQ es la capa de abstracción de la Base de datos en MySQL. La estructura del sistema de gestión de la base de datos va a ser la misma que las clases y las propiedades generadas del D2R. Por lo tanto en las tablas mostradas en la sección 7.3, se muestra la especificación de los datos.

A continuación se muestra un listado de todas las tablas que componen la base de datos como resultado de la creación de los vocabularios:

**Acceso Universitario**

Acceso_universitario → vusuk:university_access		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_campus	Número	vusuk:campus_identifier
id_acceso	Número	vusuk:university_access_identifier
presentados	Número	vusuk:submitted
aprobados	Número	vusuk:selected
porcentaje_aprobados	Double	vusuk:approved_percentage

Tabla 7.1: Acceso universitario

**Acciones plan social**

Acciones_plan_social → vaus:social_plans_actions_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_accion_social	Número	vaus:social_plans_actions_identifier
num_acciones	Número	vusuk:number_of_members

Tabla 7.2: Acciones Plan Social

**Actividades comunicación por curso**

Actividades_comunicación_por_curso → vaus:communication_activities_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_activ_comunicacion	Número	vaus:activity_identifier
id_curso	Cadena	ns:academicTerm
tipo_comunicacion	Cadena	vaus:type_communication
uso_comunicacion	Número	vusuk:number_of_members
porcentaje	Double	dbpedia:statisticValue

Tabla 7.3: Actividades de comunicación por curso

**Actividades de idiomas**

Actividades_idiomas → vusuk:languages_activities		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
num_alumnos	Número	vusuk:number_of_students
num_certificaciones	Número	vusuk:number_of_members
num_cursos	Número	vusuk:number_of_certifications

Tabla 7.4: Actividades de idiomas

**Actuaciones de defensor universitario**

Actuaciones_defensor_universitario → vaus:actions_of_university_defender		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
tipo_actuacion	Cadena	vaus:type_of_actions
comunidad_defendida	Cadena	vusuk:university_community
num_actuaciones	Número	vusuk:number_of_members

Tabla 7.5: Actuaciones del defensor universitario

**Alumnos extranjeros por curso**

Alumnos_extranjeros_por_curso → veepr:water_consumed		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_alumnos_extranjeros	Número	vusuk:foreign_students_identifier
alumnos_extranjeros	Número	vusuk:number_of_members

Tabla 7.6: Alumnos extranjeros por curso

**Alumnos graduados**

Alumnos_graduados → vusuk:graduate_students		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_rama	Cadena	vusuk:branch_of_knowledge_identifier
graduados	Número	vusuk:number_of_members
porcentaje_mujeres_graduadas	Flotante	vusuk:percentage_women

Tabla 7.7: Alumnos graduados

**Alumnos matriculados**

Alumnos_matriculados → vusuk:students_enrolled		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_rama	Número	ns:academicTerm
id_titulacion	Número	vusuk:branch_of_knowledge_identifier
centro	Cadena	vui:center
matriculados	Número	vusuk:number_of_members
porcentaje_mujeres_matriculadas	Flotante	vusuk:percentage_women

Tabla 7.8: Alumnos matriculados

**Alumnos nuevo ingreso**

Alumnos_nuevoingreso → vusuk:new_admitted_students		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_rama	Número	vusuk:branch_of_knowledge_identifier
centro	Cadena	vui:center
nuevo_ingreso	Número	vusuk:number_of_members

Tabla 7.9: Alumnos de nuevo ingreso

**Alumnos por actividad**

Alumnos_por_actividad → vusuk:students_activities		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	vusuk:university_access_identifier
id_otroestudios	Número	vusuk:non-formal_education_identifier
tipo_presencia	Cadena	vusuk:presence
alumnos_por_actividad	Número	vusuk:number_of_members

Tabla 7.10: Alumnos por actividad

**Alumnos por campus**

Alumnos_por_campus → vusuk:students_by_campus		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_campus	Número	vusuk:campus_identifier
num_alumnos	Número	vusuk:number_of_members

Tabla 7.11: Alumnos por campus

**Alumnos por centros**

Alumnos_por_centros → vusuk:students_by_school		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
centros_propios	Número	vusuk:own_college
centros_ascritos	Número	vusuk:attached_college
univ_españolas	Número	vusuk:country_universities
univ_comunidad	Número	vusuk:universities_community

Tabla 7.12: Alumnos por centros



**Alumnos de tiempo completo**

Alumnos_tiempo_completo → vusuk:full_time_students		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_titulacion	Número	vusuk:degree_identifier
num_alumnosTC	Número	vusuk:number_of_members

Tabla 7.13: Alumnos de tiempo completo

**Area paidi**

Area_paidi → vaur:paidi_area		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
area	Cadena	dcterms:description
grupos	Número	vaur:groups_of_researchers
investigadores	Número	vaur:number_of_researchers

Tabla 7.14: Areas del PAIDI

**Aula mayores graduados**

Aula_mayores_matriculados → vusuk:older_classroom_enrolled		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_campus	Número	vusuk:campus_identifier
num_primer_ciclo	Número	vusuk:number_of_first_cycle_students
num_segundo_ciclo	Número	vusuk:number_of_second_cycle_students

Tabla 7.15: Aula de mayores matriculados

**Bau**

Bau → vaus:user_mailbox_attention_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_bau	Número	vaus:user_mailbox_attention_identifier
num_baus	Número	vusuk:number_of_members
num_dias_respuesta	Número	vaus:number_of_days

Tabla 7.16: Bau

**Becas de movilidad**

Becas_de_movilidad → vuf:mobility_grants_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_beca	Número	vuf:mobility_grants_identifier
numero_becas	Número	vuf:number_of_grants

Tabla 7.17: Becas de movilidad

**Beneficiarios incentivos jubilación**

Beneficiarios_incentivos_jubilacion → vus:retirement_incentive_recipients		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
num_beneficiarios	Número	vusuk:number_of_members
tipo_beneficiario	Cadena	vusuk:university_community

Tabla 7.18: Beneficiarios de incentivos de jubilación

**Campus**

Campus → vusuk:campus		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_campus	Número	vusuk:campus_identifier
nombre_campus	Cadena	dcterms:description

Tabla 7.19: Campus

**Campus virtual**

Campus_virtual → vaus:campus_virtual		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
asignaturas	Número	vaus:number_of_subjects
alumnos	Número	vusuk:number_of_students
profesores	Número	vus:number_of_teachers
accesos	Número	vaus:page_access
info_transferida_gb_dia	Flotante	vaus:transmission_Gb_by_day
pags_visitantes	Número	vaus:pages_visit_per_day

Tabla 7.20: Tabla Aula de mayores matriculados

**Cátedras empresa**

Catedras_empresa → vuf:business_chair		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
catedras_empresa	Cadena	dcterms:description

Tabla 7.21: Cátedras de Empresa

## Cau

Cau → vaus:user_center_of_attention		
	Tipo	Término en el vocabulario
año	Número	dbpedia:year
num_solicitudes	Número	vaus:number_of_requests
satisfacion_con_tiempo_respuesta	Double	vaus:satisfaction_with_answer_time
satisfacion_con_servicio	Double	vaus:satisfaction_with_service
satisfacion_con_tiempo_solucion	Double	vaus:satisfaction_with_solve_time

Tabla 7.22: Cau

## Clasificación por edad alumnos

Clasificacion_por_edad_alumnos → vusuk:students_classified_by_age		
	Tipo	Término en el vocabulario
id_titulacion	Número	vusuk:degree_identifier
edad	Número	foaf:age
num_alumnos	Flotante	vusuk:number_of_students
porcentaje_alumnos	Flotante	dbpedia:statisticValue

Tabla 7.23: Clasificación por edad de alumnos

## Clasificación por edad empleados

Clasificacion_por_edad_empleados → vus:staff_classified_by_age		
	Tipo	Término en el vocabulario
tipo_empleado	Cadena	vusuk:university_community
edad	Cadena	foaf:age
sexo	Cadena	foaf:gender
num_empleados	Número	vus:number_of_employees

Tabla 7.24: Clasificación por edad de empleados

**Compromiso social por curso**

Compromiso_social_por_curso → vaus:social_compromise_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_compromiso	Número	vaus:social_compromise_identifier
num_indicadores	Número	vusuk:number_of_members

Tabla 7.25: Compromiso social por curso

**Contratos con empresas**

Contratos_con_empresas → vuf:contracts_with_companies		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
num_contratos	Número	vusuk:number_of_members
num_profesores	Número	vus:number_of_teachers
euros_consignados_por_contratos	Flotante	vuf:euros

Tabla 7.26: Contratos con empresas

**Convenios**

Convenios → vusuk:university_agreements_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_convenio	Número	vuf:university_agreements_identifier
num_convenios	Número	vusuk:number_of_members

Tabla 7.27: Convenios

**Curso**

Curso → vusuk:course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm

Tabla 7.28: Curso académico

**Cursos de idiomas**

Cursos_de_idiomas → vusuk:courses_languages		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_idioma	Número	vusuk:language_identifier
tipo_certificacion	Cadena	acrt:qualification
num_cursos	Número	vusuk:number_of_members

Tabla 7.29: Cursos de idiomas

**Cursos de prevención de riesgos laborales**

Cursos_prevenicion_riesgos_laborales → vus:occupational_risk_prevention_courses		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
descripcion_riesgo_laboral	Cadena	dcterms:description
num_cursos	Número	vusuk:number_of_members
destinatarios	Cadena	vusuk:university_community

Tabla 7.30: Cursos de prevención de riesgos laborales

**Difusión educativa**

Difusion_educativa → vaus:spreading_educational_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_difusion_educativa	Número	vaus:spreading_educational_identifier
numero_participantes	Número	vusuk:number_of_members

Tabla 7.31: Difusión educativa

**Doctorados por curso**

Doctorados_por_curso → vusuk:doctorate_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_doctorado	Número	vusuk:doctorate_identifier
alumnos_doctorados	Número	vusuk:number_of_students
porcentaje_mujeres_doctorados	Flotante	vusuk:percentage_women

Tabla 7.32: Doctorados por curso

**Doctorados**

Doctorados → vusuk:doctorate_by_branch_of_knowledge		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_rama	Número	vusuk:branch_of_knowledge_identifier
num_doctorados	Número	vusuk:number_of_members
porcentaje_mujeres_doctorados	Double	vusuk:percentage_women

Tabla 7.33: Doctorados

**Eliminación de residuos**

Eliminacion_residuos → veepr:remove_waste_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_residuo	Número	veepr:waste_identifier
cantidad_eliminacion	Flotante	veepr:metric_tons

Tabla 7.34: Eliminación de residuos

**Empresas vinculadas**

Empresas_vinculadas → vuf:linked_companies		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	ns:academicTerm
empresa_vinculada	Cadena	dcterms:description

Tabla 7.35: Empresas vinculadas

**Enfermedades laborales**

Enfermedades_laborales → vus:illnesses_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_enfermedad	Número	vus:illness_identifier
sexo	Cadena	foaf:gender
numero_enfermos	Número	vus:number_of_sick

Tabla 7.36: Enfermedades laborales



**Equipamientos informáticos**

Equipamientos_informaticos → vui:computer_equipment		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_servicio	Cadena	vui:computer_equipment_identifier
num equipamientos	Número	vusuk:number_of_members

Tabla 7.37: Tabla Equipamientos informáticos

**Estructuras organizativas**

Estructuras_organizativas → vus:university_structures		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_estructuras	Número	vus:university_structures_identifier
tipo_estructura	Cadena	vus:type_university_struture
descripcion_estructura	Cadena	dcterms:description

Tabla 7.38: Estructuras organizativas

**Evolución ayudas planes propios investigación**

Evolucion_ayudas_planes_propios_investigación → vaur:own_research_plans_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_planes_propios_investigacion	Número	vaur:own_research_plans_identifier
num_planes_propios	Número	vusuk:number_of_members

Tabla 7.39: Evolución de ayudas a planes propios

**Evolución investigación**

Evolucion_investigacion → vaur:grants_contacts_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_becas_contratos_investigacion	Número	vaur:grants_contracts_identifier
becas_contratos_investigacion	Número	vusuk:number_of_members

Tabla 7.40: Evolución de la investigación

**Evolución de mujeres**

Evolucion_mujeres → vusuk:women_progress		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
porcentaje_mujeres_estudiantes	Flotante	vusuk:percentage_women_student
porcentaje_mujeres_graduadas	Flotante	vusuk:percentage_women_graduate

Tabla 7.41: Evolución de las mujeres

**Fondos proyectos convocatorias públicas**

Fondos_proyectos_convocatorias_publicas → vuf:project_funding		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_proyecto	Flotante	vaur:project_identifier
fondos_euros	Flotante	vuf:euros

Tabla 7.42: Fondos de proyectos de convocatorias públicas

**Formación pas**

Formacion_pas → vus:pas_training_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_formacion_pas	Número	vus:pas_training_identifier
num_actividades_formativas	Número	vusuk:number_of_members
num_horas	Flotante	vus:number_of_hours
num_asistentes	Número	vusuk:number_of_pas

Tabla 7.43: Formación del PAS

**Formación pdi**

Formacion_pdi → vus:pdi_training_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_formacion_pdi	Número	vus:pdi_training_identifier
num_actividades_formativas	Flotante	vusuk:number_of_members

Tabla 7.44: Formación del PDI

**Fuentes primarias no renovables**

Fuentes_primarias_no_renovables → veepr:primary_sources_nonrenewable		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
fuelle	Cadena	dcterms:description
litros	Double	veepr:liters
gigaJulios	Double	veepr:gigajoules
kg CO2 emitidos	Double	veepr:kilograms_of_CO2

Tabla 7.45: Fuentes primarias no renovables

**Gastos e ingresos**

Gastos_ingresos → vuf:income_expenses_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Flotante	dbpedia:year
id_gastos	Número	vuf:income_expenses_identifier
presupuesto_millones_euros	Flotante	vuf:million_euros

Tabla 7.46: Gastos e ingresos

**Grupo pas**

Grupos_pas → vus:pas_groups		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_grupospas	Cadena	vus:pas_groups_identifier
integrantes	Número	vusuk:number_of_members
porcentaje_mujeres_grupos_pas	Flotante	vusuk:percentage_women

Tabla 7.47: Grupos del PAS

**Horas formación plantilla por sexo pas**

Horas_formacion_plantilla_por_sexo_pas → vus:hours_pas_training		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
tipo_grupo_pas	Cadena	vus:pas_groups_identifier
sexo	Cadena	foaf:gender
num_horas	Flotante	vus:number_of_hours

Tabla 7.48: Horas de formación de plantilla del pas por sexo

**Horas de formación plantilla por sexo pdi**

Horas_formacion_plantilla_por_sexo_pdi → vus:hours_pdi_training		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_figura	Número	vus:staff_identifier
sexo	Cadena	foaf:gender
num_horas	Flotante	vus:number_of_hours

Tabla 7.49: Horas de formación de la plantilla del pdi por sexo

**Igualdad entre sexos**

Igualdad_entre_sexos → vusuk:gender_relations_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_relacion	Cadena	vusuk:gender_relations_identifier
num_relaciones	Flotante	vusuk:number_of_members

Tabla 7.50: Igualdad entre sexos

**Indicadores de deporte**

Indicadores_de_deporte → vaus:signs_sports_activities_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_indicadores_de_deporte	Número	vaus:signs_of_sports_activities_identifier
num_indicadores	Flotante	vusuk:number_of_members

Tabla 7.51: Indicadores de deporte

**Insercción laboral**

Insercion_laboral → vusuk:employment_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_inserccion_laboral	Número	vusuk:employment_identifier
porcentaje	Flotante	dbpedia:statisticValue
promocion	Cadena	vusuk:promotion

Tabla 7.52: Inserción laboral por curso

**Medidas de conciliación con vida laboral**

Medidas_conciliacion_con_vida_laboral → vus:conciliation_measures_working_life_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_medidas_conciliacion	Número	vus:conciliation_measures_working_life_identifier
num_medidas	Número	vusuk:number_of_members

Tabla 7.53: Medidas de conciliación con la vida laboral

**Nuevas instalaciones**

Nuevas_instalaciones → vui:new_infrastructure		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
lugar	Cadena	vui:place
estado_desarrollo	Cadena	vui:state_of_development
nuevas_instalaciones	Texto	dcterms:description

Tabla 7.54: Nuevas instalaciones

**Número de actividades**

Numero_actividades → vusuk:university_activities		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_otroestudios	Número	vusuk:non-formal_education_identifier
tipo_presencia	Cadena	vusuk:presence
actividades	Número	vusuk:number_of_members

Tabla 7.55: Número de actividades

**Oferta grupos de investigación**

Oferta_grupos_investigacion → vaur:university_research_groups		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
tipo_grupo_investigacion	Cadena	vaur:type_of_research_group
puntuacion_paidi	Flotante	vaur:points_paidi
numero	Número	vusuk:number_of_members
investigadores	Número	vaur:number_of_researchers

Tabla 7.56: Oferta de grupos de investigación

**Ofertas titulaciones por cursos**

Ofertas_titulacion_por_cursos → vusuk:university_degrees		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_titulacion	Número	vusuk:degre_identifier
id_rama	Número	vusuk:branch_of_knowledge_identifier
ofertas_cursos	Número	vusuk:number_of_members

Tabla 7.57: Ofertas de titulación por cursos

**Origen de financiación**

Origen_financiacion → vuf:university_funding		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_financiacion	Número	vuf:funding_identifier
total	Número	vuf:total_budget
porcentaje	Flotante	dbpedia:statisticValue

Tabla 7.58: Origen de financiación

**Papel**

Papel → veepr:paper_waste		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_campus	Número	vusuk:campus_identifier
tonelada_metrica	Flotante	veepr:metric_tons
kg_recogido_por_usuario	Flotante	veepr:kilograms_per_user

Tabla 7.59: Residuos del papel

**Pas**

Pas → vus:pas		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_figura	Número	vus:staff_identifier
plantilla	Número	vus:number_of_pas

Tabla 7.60: Personal de administración y servicios



**Perfil de la universidad**

Perfil_universidad → vusuk:university_profile		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
nombre_universidad	Texto	foaf:name
localizacion_sede_principal	Texto	dbpedia:headquarter
nombre_rector	Texto	dbpedia:president
punto_de_contacto	Texto	foaf:workplaceHomepage

Tabla 7.61: Perfil de la Universidad

**Prácticas de empresa**

Practicas_empresa → vuf:business_practices		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_rama	Número	vusuk:branch_of_knowledge_identifier
num_practicas	Número	vusuk:number_of_members

Tabla 7.62: Prácticas de empresa

**Premios y reconocimientos**

Premios_reconocimientos → cerif:PriceAwards		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
titulo	Cadena	cerif:abstract
premio	Cadena	cerif:price
responsable	Cadena	semcerif:awarded
tipo_persona	Cadena	vusuk:university_community

Tabla 7.63: Tabla de Premios y reconocimientos

**Presupuesto**

Presupuesto → vuf:university_budget		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
millones_euros_presupuesto	Flotante	vuf:thousand_euros
subvenciones_gobierno_miles_euros	Flotante	vuf:thousand_euros_research_investment
inversion_investigacion_miles_euros	Flotante	vuf:thousand_euros_subvention

Tabla 7.64: Presupuesto

**Profesorado**

Profesorado → vus:teaching_staff		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_figura	Número	vus:staff_identifier
tiempo	Cadena	vus:type_work_time
num_profesores	Número	vus:number_of_teachers

Tabla 7.65: Tabla de Profesorado

**Programas salud laboral**

Programas_salud_laboral → vaus:occupational_health_programs_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_campus	Número	vusuk:campus_identifier
num_programas	Número	vusuk:number_of_members

Tabla 7.66: Programa de salud laboral

**Programa erasmus**

Programas_erasmus → vuf:erasmus		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
estudiantes_acogidos	Número	vusuk:incoming_students
estudiantes_salientes	Número	vusuk:outgoing_students
plazas_ofertadas	Número	vuf:places_offered
profesores	Número	vus:number_of_teachers
pas	Número	vus:number_of_pas
pdi	Número	vus:number_of_pdi

Tabla 7.67: Programa Erasmus

**Protección de resultados de investigación**

Proteccion_resultados_investigacion → vaur:results_research_protection_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Número	ns:academicTerm
id_proteccion_resultados_investigacion	Número	vaur:results_research_protection_identifier
num_proteccion_resultados_investigacion	Número	vusuk:number_of_members

Tabla 7.68: Protección de resultados de investigación por año

**Proyectos concedidos en convocatorias públicas**

Proyectos_concedidos_convocatorias_publicas → vaur:granted_projects_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_proyecto	Número	vaur:project_identifier
concedidos	Número	vusuk:number_of_members

Tabla 7.69: Proyectos concedidos de convocatorias públicas

**Proyectos docentes**

Proyectos _docentes → vaur:teaching_innovation_projects		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
tipo_proyecto	Cadena	vaur:project_identifier
solicitados	Número	vaur:projects_requested
aceptados	Número	vaur:projects_accepted
implicados	Número	vus:number_of_teachers

Tabla 7.70: Proyectos docentes

**Residuos eléctricos y electrónicos**

Raees → vepr:electrical_electronic_equipment_waste		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
tm_aparatos_electronicos	Flotante	vepr:metric_tons

Tabla 7.71: Residuos eléctricos y electrónicos

**Rama**

Rama → vusuk:branch_of_knowledge		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_rama	Número	vusuk:branch_of_knowledge_identifier
rama	Texto	dterms:description

Tabla 7.72: Ramas de conocimientos

**Recursos de investigación**

Recursos_investigacion → vusuk:research_resources_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_recursos_investigacion	Cadena	vaur:research_resources_identifier
num_ayudas	Número	vusuk:number_of_members
inversion_miles_euros	Número	vuf:thousand_euros

Tabla 7.73: Recursos de investigación

**Rendimiento por curso**

Rendimiento_por_curso → vusuk:academic_performance_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_rendimiento	Cadena	vusuk:academic_performance_identifier
id_titulacion	Número	vusuk:degree_identifier
indicador_rendimiento	Flotante	dbpedia:statisticValue

Tabla 7.74: Rendimiento por curso

**Retribución mínima**

Retribucion_minima → vus:minimum_salary_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
beneficiarios	Cadena	vusuk:university_community
salario_minimo	Flotante	vus:minimum_interprofessional_wage

Tabla 7.75: Retribución mínima

**Satisfacción con servicios**

Satisfacion_con_servicios → vaus:satisfaction_with_university_services		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
satisfacion_pas	Flotante	vaus:pas_satisfaction
satisfacion_pdi	Flotante	vaus:pdi_satisfaction
satisfacion_alumnos	Flotante	vaus:students_satisfaction

Tabla 7.76: Satisfacción con servicios

**Satisfacción educativa por curso**

Satisfaccion_educativa_por_curso → vusuk:educational_satisfaction_by_course		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
id_satisfaccion_educativa	Número	vusuk:educational_satisfaction_identifier
valoracion	Flotante	vaus:satisfaction_with_service

Tabla 7.77: Satisfacción educativa por curso

**Servicios biblioteca**

Servicios_biblioteca → vaus:library_service_by_year		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_servicio	Número	vaus:university_services_idenitifier.
num_servicios	Número	vusuk:number_of_members.

Tabla 7.78: Servicios de biblioteca

**Servicios de información**

Servicios_de_informacion → vaus:information_services		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
año	Número	dbpedia:year
id_servicio	Número	vaus:university_service_identifier
num_servicios	Número	vusuk:number_of_members

Tabla 7.79: Servicios de información

**Tesis doctorales**

Tesis_doctorales → semcerif2:DoctoralThesis		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_curso	Cadena	ns:academicTerm
tesis_doctorales	Número	vusuk:number_of_members
tesis_mencion_doctor_europeo	Número	vaur:mention_of_european_doctor
tesis_defendidas_mujeres	Número	vaur:thesis_by_women

Tabla 7.80: Tesis doctorales

**Tipos de accesos universitarios**

Tipos_accesos → vusuk:university_access		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_acceso	Número	vusuk:university_access_identifier
descripcion_acceso	Texto	dcterms:description

Tabla 7.81: Tipos de accesos universitarios

**Tipos de acciones sociales**

Tipos_acciones_sociales → vaus:social_plans_actions		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_accion_social	Número	vaus:social_plans_actions_identifier
descripcion_accion_social	Texto	dcterms:dscrition

Tabla 7.82: Tipos de acciones sociales

**Tipos de actividades comunicación**

Tipos_actividades_comunicacion → vaus:communication_activities		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_activ_comunicacion	Número	vaus:activity_identifier
actividades_comunicacion	Texto	dcterms:description

Tabla 7.83: Tipos de actividades de comunicación

**Tipos de alumnos extranjeros**

Tipos_alumnos_extranjeros → vusuk:foreign_students		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_alumnos_extranjeros	Número	vusuk:foreign_students_identifier
descripcion_tipo_alumno_extranjero	Texto	dcterms:description

Tabla 7.84: Tipos de alumnos extranjeros



**Tipos de ayudas planes propios de investigación**

Tipos_ayudas_planes_propios_investigacion → vaur:own_research_plans		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_planes_propios_investigacion	Número	vaur:own_research_plans_identifier
descripcion_planes_propios_investigacion	Texto	dcterms:description

Tabla 7.85: Tipos de ayudas a planes propios de investigación

**Tipo de bau**

Tipo_bau → vaus:user_mailbox_attention		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_bau	Número	vaus:user_mailbox_attention_identifier
descripcion_bau	Texto	ns:academicTerm

Tabla 7.86: Tipos de bau

**Tipo de becas y contratos e investigación**

Tipo_becas_contratos_investigacion → vaur:grants_contracts		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_becas_contratos_investigacion	Número	vaur:grants_contracts_identifier
tipo_becas_contratos	Texto	dcterms:description

Tabla 7.87: Tipos de becas o contrato de investigación

**Tipo de becas de movilidad**

Tipo_becas_de_movilidad → vuf:mobility_grants		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_beca	Número	vuf:mobility_grants_identifier
beca	Cadena	dcterms:description
pais_destino	Cadena	vuf:country_destination_of_grants

Tabla 7.88: Tipos de becas de movilidad

**Tipo de compromiso social**

Tipo_compromiso_social → vaus:social_compromise		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_compromiso	Número	vaus:social_compromise_identifier
compromiso_social	Texto	dcterms:description

Tabla 7.89: Tipos de compromiso social

**Tipo de convenios**

Tipo_convenios → vuf:university_agreements		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_convenio	Número	vuf:university_agreements_identifier
descripcion_convenio	Texto	dcterms:description

Tabla 7.90: Tipos de convenios

**Tipo de difusión educativa**

Tipo_difusion_educativa → vaus:spreading_educational		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_difusion_educativa	Número	vaus:spreading_educational_identifier
descripcion_dif_educativa	Cadena	dcterms:description

Tabla 7.91: Tipos de difusión educativa

**Tipo doctorado**

Tipo_doctorado → semcerif:Doctorate		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_doctorado	Número	vusuk:doctorate_identifier
doctorado	Texto	dcterms:description

Tabla 7.92: Tipos de doctorados

**Tipos enfermedades**

Tipo_enfermedades → vus:illnesses		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_enfermedad	Número	vus:illness_identifier
enfermedad	Texto	dcterms:description

Tabla 7.93: Tipos de enfermedades laborales

**Tipo de figura contratada**

Tipo_figura_contratada → vus:staff		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_figura	Número	vus:staff_identifier
figura_contratada	Texto	dcterms:description

Tabla 7.94: Tipos de figura contratada

**Tipo de financiación**

Tipo_financiacion → cerif:Funding		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_financiacion	Número	vuf:funding_identifier
financiacion	Texto	dcterms:description

Tabla 7.95: Tipos de financiación

**Tipo de formación del pas**

Tipo_formacion_pas → vus:pas_training		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_formacion_pas	Número	vus:pas_training_identifier
descripcion_formacion_pas	Cadena	dcterms:description

Tabla 7.96: Tipos de formación del PAS

**Tipo de formación del pdi**

Tipo_ formacion_pdi → vus:pdi_training		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_ formacion_pdi	Número	vus:pdi_training_identifier
descripcion_ formacion_pdi	Texto	dcterms:description

Tabla 7.97: Tipos de formación del PDI

**Tipo de gastos e ingresos**

Tipo_gastos_ingresos → vuf:income_expenses		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_gastos_ingresos	Número	vuf:income_expenses_identifier
descripcion	Texto	dcterms:description
gastos_ingresos	Cadena	rdf:type

Tabla 7.98: Tipos de gastos e ingresos

**Tipo de idioma**

Tipo_idioma → lingvo:Language		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_idioma	Número	vusuk:language_identifier
descripcion_idioma	Texto	dcterms:description

Tabla 7.99: Tipos de idiomas

**Tipo de igualdad entre sexos**

Tipo_igualdad_entre_sexos → vusuk:gender_relations		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_relacion	Número	vusuk:gender_relations_identifier
descripcion_relacion	Texto	dcterms:description

Tabla 7.100: Tipos de igualdad entre sexos

**Tipo de indicadores de deporte**

Tipo_indicadores_de_deporte → vaus:signs_of_sports_activities		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_indicadores_de_deporte	Número	vaus:sign_of_sports_activities_identifier
descripcion_indicadores_de_deporte	Texto	dcterms:description

Tabla 7.101: Tipos de indicadores de deporte

**Tipo de inserción laboral**

Tipo_insercion_laboral → vusuk:employment		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_insercion_laboral	Número	vusuk:employment_identifier
inserccion_laboral	Cadena	dcterms:description

Tabla 7.102: Tipos de inserción laboral

**Tipo de medidas de conciliación con vida laboral**

Tipo_medidas_conciliacion_con_vida_laboral → vus:conciliation_measures_working_life		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_medidas_conciliacion	Número	vus:conciliation_measures_working_life_identifier
descripcion_medidas_conciliacion	Cadena	dcterms:description

Tabla 7.103: Tipos de conciliación con vida laboral

**Tipo de modalidad formativa**

Tipo_modalidad_formativa → vusuk:non-formal_education		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_otrosetudios	Número	vusuk:non-formal_education_identifier
otros_estudios	Texto	dcterms:description

Tabla 7.104: Tipo de modalidades formativas

**Tipo de protección de resultados de investigación**

Tipo_proteccion_resultados_investigacion → vaur:results_research_protection		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_proteccion_resultados_investigacion	Número	vaur:results_research_protection_identifier
descripcion_proteccion_resultados	Cadena	dcterms:description

Tabla 7.105: Tipos de protección de resultados de investigación

**Tipo de proyectos de convocatorias públicas**

Tipo_proyectos_convocatorias_publicas → cerif:Project		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_proyectos	Número	vaur:project_identifier
proyectos	Cadena	dcterms:description

Tabla 7.106: Tipos de proyectos de convocatorias públicas

**Tipo de recursos de investigación**

Tipo_recursos_investigacion → vaur:research_resources		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_recursos_investigacion	Número	vaur:research_resources_identifier
descripcion_recursos_investigacion	Texto	dcterms:description

Tabla 7.107: Tabla de Tipos de recursos de investigación

**Tipo de rendimiento**

Tipo_rendimiento → vusuk:academic_performance		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_rendimiento	Número	vusuk:academic_performance_identifier
redimiento_academico	Texto	dcterms:description

Tabla 7.108: Tipos de rendimiento



**Tipo de residuos eliminados**

Tipo_residuos → veepr:remove_waste		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_residuo	Número	veepr:waste_identifier
descripcion_residuo	Cadena	dcterms:description
metodo_tratamiento	Cadena	veepr:treatment
operaciones_valorizacion Eliminación	Cadena	veepr:operation_with_waste

Tabla 7.109: Tipos de residuos eliminados

**Tipo de satisfacción educativa**

Tipo_satisfaccion_educativa → vusuk:educational_satisfaction		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_satisfaccion_educativa	Número	vusuk:educational_satisfaction_identifier
descripcion_satisfaccion_educativa	Cadena	dcterms:description

Tabla 7.110: Tipos de satisfacción educativa

**Tipo de servicios universitario**

Tipo_servicios_universitarios → vaus:university_services		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_servicio	Número	vaus:university_services_identifier
servicio_universitario	Texto	dcterms:description

Tabla 7.111: Tipos de servicios universitarios

**Tipo de titulaciones**

Tipo_titulaciones → vusuk:degree		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_titulacion	Número	vusuk:degree_identifier
titulacion	Texto	dterms:description

Tabla 7.112: Tipo de titulaciones

**Titulaciones por rama y campus**

Titulaciones_por_rama_campus → vusuk:degree_by_campus_knowledge		
	<b>Tipo</b>	<b>Término en el vocabulario</b>
id_campus	Número	vusuk:campus_identifier
id_rama	Número	vusuk:branch_of_knowledge_identifier
id_titulacion	Número	vusuk:degree_identifier
centro	Cadena	vui:center
nombre_titulacion	Cadena	dterms:description

Tabla 7.113: Titulaciones por rama y campus

## 7.4. Publicación de datos RDF

Como se ha mencionado anteriormente, uno de los componentes del sistema de información, D2R-Server, es una herramienta cuya misión es publicar contenido de bases de datos relacionales en la Web Semántica. Este servidor hace uso de HTML y RDF para permitir a los navegadores (como Google Chrome, Mozilla, Opera, etc.) navegar por el contenido de la base de datos de la Web Semántica y también permite hacer consultas usando SPARQL usando una interfaz web<sup>5</sup>.

La plataforma D2RQ usa un lenguaje descriptivo<sup>6</sup> que nos permite describir mapeos entre el esquema de una base de datos relacional de uso específico y una ontología RDF/OWL. Gracias a este lenguaje podemos:

- Realizar consultas SPARQL a una base de datos que no es RDF.
- La base de datos se transforma en un grafo de sólo lectura. De esta forma, a través de las llamadas a la API de Jena se reescriben las consultas en SQL y realizan la consulta a los datos de la base de datos de la aplicación.

## 7.5. Aplicación de consulta de datos

Se diseñó e implementó una aplicación de consulta, *Linked Data Viewer*<sup>7</sup>, como muestra de una aplicación de explotación de los datos almacenados en el sistema e integración con datos procedentes de fuentes de terceros.

### 7.5.1. Diseño de la interfaz de usuario

En esta sección se especifican las interfaces entre el sistema y el usuario, detallando el aspecto y el comportamiento de las diferentes pantallas e informes, de acuerdo con el entorno tecnológico definido.

Al comienzo del proyecto se creó un prototipo de la aplicación de consulta con la idea de poder mostrar la funcionalidad básica y poder identificar fallos de navegación. A continuación se ofrece una imagen de dicho prototipo, cuyo diseño fue descartado en fases posteriores en favor de uno diferente:

---

<sup>5</sup><http://lod.uca.es/d2rq/snorql>

<sup>6</sup><http://d2rq.org/d2rq-language>

<sup>7</sup><http://lod.uca.es/viewer>

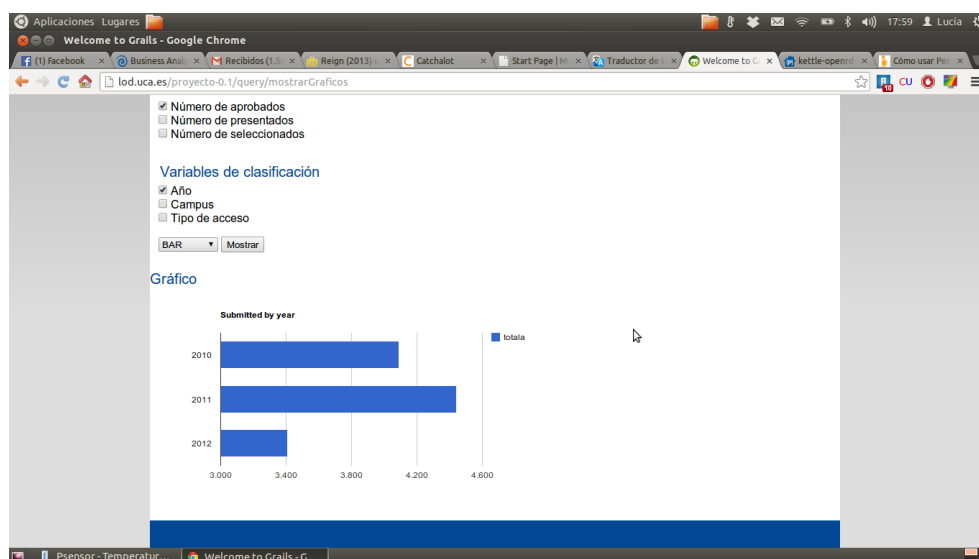


Figura 7.3: Prototipo de la aplicación

La interfaz gráfica de la aplicación ha sido pensada con la simplicidad como elemento fuerte, decidiéndose mostrar la navegación principal del sitio web en una cuadrícula con etiquetas identificativas y logotipos descriptivos con el fin de que el usuario pueda identificar a simple vista las diferentes categorías de los datos publicados.



Figura 7.4: Menu principal

Para la pantalla de consulta de datos, se optó por un diseño a dos columnas, con la columna principal izquierda más ancha para albergar los gráficos y tablas mostrados como resultado de la consulta del usuario. La columna izquierda permite, a través de un simple formulario, modificar

los parámetros de la consulta en función de las variables a seleccionar y el formato del gráfico en el que se desea presentar los resultados.

Principal / Datos de Alumnos / Alumnos que realizan las pruebas de acceso a la universidad

Selecciones las variables y el tipo de visualización

**UCA**  
Universidad de Cádiz

**Variables numéricas**

- Número de presentados
- Número de seleccionados

**Variables de clasificación**

- Año
- Campus
- Tipo de acceso

**Tipo de visualización**

Tabla de datos

Figura 7.5: Formulario de Consulta

Dicho formato se repitió para el escenario de comparación de datos externos, añadiendo un campo de tipo fichero en el que el usuario deberá proporcionar el archivo RDF que contiene los datos externos que se desean comparar con los albergados en el almacén de datos de la Universidad de Cádiz.

Inicio Portal de usuarios Portal para desarrolladores Comparativas con otras universidades

Principal / Comparativa / Alumnos que realizan las pruebas de acceso a la universidad

Selecciones las variables y el tipo de visualización

**UCA**  
Universidad de Cádiz

**Subida de fichero**

Seleccionar archivo Ningún archivo seleccionado

**Variables numéricas**

- Número de presentados
- Número de seleccionados

**Variables de clasificación**

- Año
- Tipo de acceso

**Tipo de visualización**

Tabla de datos

Figura 7.6: Formulario de Comparación

### 7.5.2. Arquitectura de diseño

La aplicación de consulta de datos *Linked Data Viewer*, se ha desarrollado siguiendo un patrón arquitectónico Modelo-Vista-Controlador (MVC, *Model-View-Controller*) dado que desde el punto de vista del desarrollo y mantenimiento futuro del software proporciona una clara separación entre las capas de presentación, lógica y acceso a datos, facilitando la adición de nuevas características y reduciendo el acoplamiento de las distintas funcionalidades que componen la aplicación.

El patrón MVC es ampliamente usado en todo tipo de software, pero es en las aplicaciones con fuerte uso de interfaces de usuario, como las aplicaciones web, donde dicho patrón es el preferido, por su diferenciación de los cometidos de cada parte. Como su nombre indica, el patrón MVC se compone de tres tipos de componentes, claramente diferentes:

- **Modelo:** Es una representación de los datos accedidos por la aplicación. Gestiona el acceso, modificación, creación y eliminación de datos, abstrayendo su fuente real (sea una base de datos, ficheros, servicios web, etc.).
- **Vista:** Es una abstracción de la presentación de la aplicación. Se encarga de gestionar toda la capa de presentación, mostrando al usuario los resultados de la petición en un formato acorde con las necesidades de cada caso.
- **Controlador:** Se encarga de orquestar la aplicación, siendo el intermediario directo entre el modelo y la vista. Es invocado por las acciones del usuario y coordina las acciones del sistema para responder a la petición del usuario de forma correcta.

Como se puede comprobar, los componentes que forman el patrón básico de cualquier aplicación basada en el modelo MVC se corresponden con las diferentes capas del modelo teórico de desarrollo de aplicaciones, separando cada una de las capas en función de su responsabilidad. Esto es una clara ventaja a la hora del mantenimiento del software, puesto que permite reutilizar una cantidad ingente de código, evitando el crecimiento innecesario de líneas de código a revisar, probar y mantener, así como facilita a los desarrolladores la inclusión de nuevas funcionalidades de forma rápida y sencilla.

La aplicación de consulta es una aplicación Java que sigue el patrón Modelo-Vista-Controlador (MVC) en la que se encuentran todos los diferentes componentes que propone la arquitectura de diseño. En nuestro caso, al utilizar un paradigma de programación imperativa orientada a objetos únicamente disponemos de una clase `QueryController` que actúa de controlador principal de la aplicación, una clase `Sparql` que implementa un servicio de acceso a datos a través de SPARQL correspondiente con el componente del Modelo, y un componente de Vista, implementados en cada uno de los ficheros `.gsp` de Grails (ver estructura de ficheros en sección 8.2.1).

## Capítulo 8

# Implementación del Sistema

Este capítulo trata sobre todos los aspectos relacionados con la implementación del sistema, haciendo uso de un determinado entorno tecnológico.

### 8.1. Entorno tecnológico

#### 8.1.1. Interfaz con el usuario

Debido al carácter informativo del producto final, la elaboración de una interfaz de usuario simple, usable y amigable ha sido uno de los requisitos fundamentales a la hora de desarrollar la aplicación, para evitar frustraciones y confusiones a usuarios, intentando que cualquier persona, independientemente de su procedencia o experiencia, pueda usar la aplicación y obtener la información que desea.

Esto ha llevado a la utilización de estándares y tecnologías probados en la industria que ayuden a alcanzar tal fin de la forma más sencilla de cara al desarrollador. Para el desarrollo de la interfaz de usuario, nos hemos basado en las siguientes tecnologías:

#### Lenguajes de marcas de hipertexto, HTML

HTML (*HyperText Markup Language*) es el pilar básico de la *World Wide Web*. Fue desarrollado por Tim Berners-Lee [16] durante su periodo de contratista en el Centro de Investigación Nuclear (CERN) como un lenguaje de definición de documentos cuyo objetivo sería facilitar la creación de documentos enlazados para compartir documentación y todo tipo de información para la organización. Posteriormente se estandarizó y es el estándar actual de definición de estructura y contenidos de un sitio web.

Existen diferentes versiones del estándar HTML. *Linked Data Viewer* usa y respeta el estándar HTML en su quinta versión, y su sintaxis ha sido validada por la herramienta oficial de validación del W3C (*World Wide Web Consortium*), como se detalla en la sección 9.

#### Lenguaje de Hojas de Estilo en Cascada, CSS

CSS (*Cascading Style Sheets*) es un estándar publicado por el W3C para definir un lenguaje de descripción de estilos y formatos asociados a un documento web. Surge para dar respuesta a la necesidad de separar la estructura y contenido de un documento web con la información asociada

a su presentación (posición, color, formato, etc.) en los diferentes dispositivos, que resulta cuanto menos problemático en sitios grandes o documentos complejos. CSS permite definir estilos que luego serán aplicados a diferentes elementos dentro de uno o varios documentos HTML, permitiendo así la reutilización de dichas definiciones para documentos diferentes, lo cual facilita el mantenimiento del sitio web.

Haciendo uso de CSS para la maquetación de la aplicación aumentamos la reutilización de código, a la vez que facilitamos a los desarrolladores la modificación futura de la misma, llegando incluso al extremo de poder modificar por completo la apariencia del sitio web sin necesidad de cambiar la estructura o el contenido de los documentos HTML gracias a la independencia del contenido con los estilos usados en su presentación.

Para el desarrollo de este proyecto, se ha hecho uso de Bootstrap<sup>1</sup> en su versión 3, un framework CSS inicialmente desarrollado para su uso en la popular red social Twitter, con el objetivo de agilizar el desarrollo y maquetación del producto, partiendo de una base consolidada y conocida, lo cual contribuye tanto a la calidad final de la aplicación como al tiempo y costes de desarrollo.

## Javascript

Javascript es un lenguaje de scripting inicialmente concebido para dotar de dinamismo en el lado del cliente a los sitios web. Es una de las tecnologías más usadas hoy día en ámbitos muy diferentes, pero usado en su inmensa mayoría como lenguaje de scripting ejecutado los navegadores para modificar de forma dinámica el contenido de los sitios web, mejorando la interactividad de las aplicaciones y la experiencia final del usuario.

En nuestra aplicación se ha usado el framework JQuery<sup>2</sup> en su versión 1.10 para realizar el desarrollo del código Javascript, ya que agiliza en gran medida el trabajo con este lenguaje, aportando funcionalidades avanzadas de forma estándar que con el lenguaje tradicional supondrían muchas líneas de código y tiempo invertido en desarrollar funcionalidades muy comunes que vienen de serie con JQuery. Asimismo se ha usado la librería Select Picker para mejorar la usabilidad de los cuadros de texto desplegados existente en la aplicación.

## Google Charts

Google Charts<sup>3</sup> es un producto de la empresa Google Inc. ofrecido como servicio de forma gratuita que permite la creación de gráficos de distintos tipos en tiempo real. Gracias a este servicio se evita tener que desarrollar un componente que integre una librería de gráficos externa, reduciendo el tiempo y coste de desarrollo, a la vez que se facilita el mismo por la naturaleza de servicio web del producto.

El resultado del uso de este producto es unos gráficos ricos en funcionalidad y compatibles con un amplio abanico de navegadores, a la vez que se reduce el tiempo de carga de la aplicación puesto que los gráficos son generados en formato vectorial (SVG) en lugar de una imagen, por lo que la cantidad de datos transmitidos es menor, tanto para el cliente (menor tiempo de carga

---

<sup>1</sup><http://getbootstrap.com/>

<sup>2</sup><http://jquery.com/>

<sup>3</sup><https://developers.google.com/chart/>



total) como para el servidor de nuestra aplicación (descarga de trabajo hacia los servidores de Google, resultando en menor consumo de recursos y tráfico).

## Iconos

Los gráficos e iconos de una aplicación contribuyen en gran manera a la usabilidad y riqueza visual de la misma, aportando un alto grado de información no textual a los usuarios. Para el desarrollo de esta aplicación, se ha usado la biblioteca de iconos Flaticon<sup>4</sup>, disponibles gratuitamente para uso personal y comercial, un conjunto de iconos en formato vectorial (SVG) con distinta temática. Al ser formato vectorial, es posible escalar los iconos a las dimensiones necesarias en cada momento sin perder calidad, a la vez que reduce el tiempo de carga de la página al ser transferidos como texto en lugar de imágenes, redundando así en una mejor experiencia de usuario final.

### 8.1.2. Lógica de servidor

Para el conjunto de procesos ejecutados en el servidor como parte de la aplicación, se han usado otras tecnologías diferentes adecuadas a la tarea y a la infraestructura existente sobre la que se ejecuta.

## Groovy/Grails

Para el desarrollo de la aplicación web de este proyecto, hemos empleado el lenguaje Groovy<sup>5</sup>, un lenguaje de programación dinámico diseñado sobre la plataforma Java, en conjunción con el framework Grails<sup>6</sup>. El combo Groovy+Grails proporciona una potencia y velocidad de desarrollo muy elevada, adhesión al patrón de desarrollo MVC (7.5.2), y la existencia de un entorno de desarrollo completo basado en el entorno de desarrollo integrado Eclipse junto a un servidor web preconfigurado de pruebas, reduciendo el tiempo efectivo de configuración y adaptación a la plataforma.

La decisión de la plataforma además se vio fuertemente influenciada por la necesidad de disponer de un paquete tecnológico probado, eficiente y escalable, para dotar de la capacidad de crecimiento necesaria al proyecto si se llegase a necesitar en un futuro, además de la anterior experiencia y comodidad de la autora con la plataforma Java y extensa documentación de todos los paquetes mencionados.

La elección de una aplicación Java en contraposición a otra tecnología facilita, además, la configuración de toda la infraestructura necesaria, la cual se reduce enormemente, pues D2RQ, el software utilizado para la publicación de datos RDF, es igualmente una aplicación Java, por lo que se reduce el número de intérpretes y entornos de ejecución, siendo ambas gestionadas por el servidor de aplicaciones Tomcat.

---

<sup>4</sup><http://www.flaticon.com/>

<sup>5</sup><http://groovy.codehaus.org/>

<sup>6</sup><https://grails.org/>

A continuación se muestra un diagrama explicativo de la plataforma de desarrollo usada en el presente proyecto:

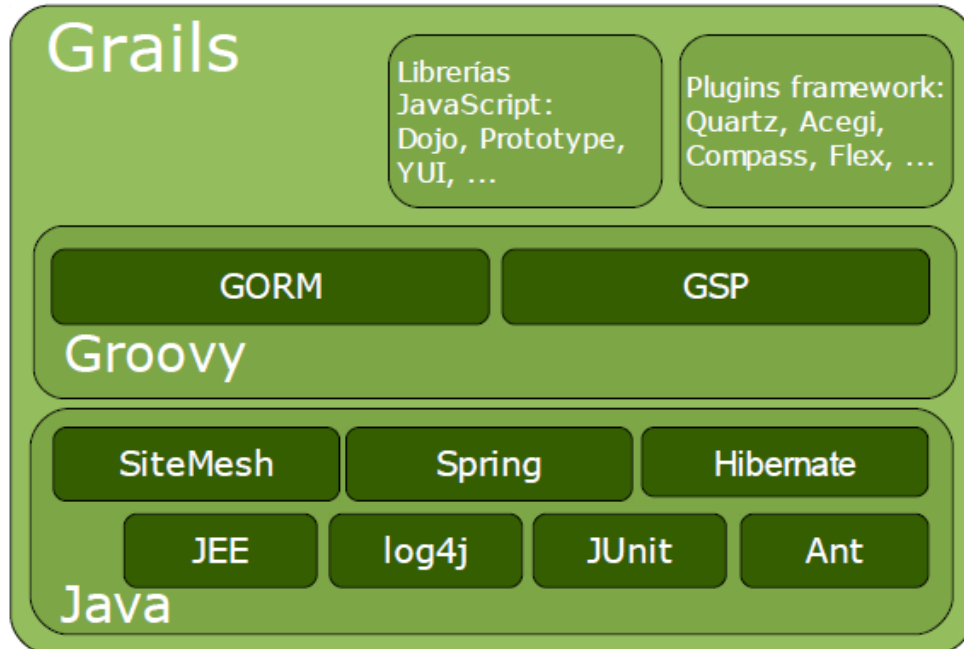


Figura 8.1: Estructura framework Grails

### 8.1.3. D2R-Server

Para establecer las correspondencias entre la base de datos y la ontología, se crea un archivo de mapeo, el cual es un documento RDF que sigue la sintaxis N3, en nuestro caso el fichero se llama `mapping.n3`. Para realizar el esquema se usan las siguientes estructuras:

- **Espacio de nombres**

El espacio de nombres es la declaración de todos los vocabularios RDF que se van a utilizar, que deben realizarse al principio del fichero.

- **d2rq:Database**

Esta estructura se encarga de definir una conexión JDBC a una base de datos relacional local y especifica el tipo de las columnas de la base de datos utilizada. Para definir dicha conexión se utilizan las siguientes estructuras:

- *d2rq:jdbcDriver*. Indica el URL JDBC de la Base de datos.
- *d2rq:jdbcDSN*. Especifica el nombre del Driver JDBC para la base de datos.
- *d2rq:username* y *d2rq:password*. Son el usuario y contraseña de la base de datos.
- *d2rq:autoReconnect* y *d2rq:zeroDateTimeBehaviour*. Son propiedades propias del driver JDBC para MySQL.

```

1  map:database a d2rq:Database;
2      d2rq:jdbcDriver "com.mysql.jdbc.Driver";
3      d2rq:jdbcDSN "jdbc:mysql://localhost:3306/memoria_uca";
4      d2rq:username "root";
5      d2rq:password "*****";
6      jdbc:autoReconnect "true";
7      jdbc:zeroDateTimeBehavior "convertToNull";
8  .

```

#### ▪ **d2rq:ClassMap**

Esta estructura representa una clase o grupo de clases similares de una ontología OWL o un esquema RDFs. Esta palabra identifica las instancias de la clase de la ontología que se está mapeando. La aplicación se conecta a un `d2rq:Database` y tiene un conjunto de `d2rq:PropertyBridges` que asignan propiedades a las instancias. Algunas de las propiedades de esta estructura son:

- *d2rq:dataStorage*. Hace referencia a un `d2rq:Database` donde son almacenados los datos de las instancias.
- *d2rq:Class*. Especifica una clase RDFS u OWL de la ontología.
- *d2rq:uriPattern*. Especifica un patrón URI que será utilizado para identificar las instancias de `d2rq:ClassMap`. Un patrón URI se crea insertando los valores de ciertas columnas de la base de datos, donde la parte de las @@ indican las columnas de la base de datos al igual que `Tabla.Columna`.
- *d2rq:classDefinitionLabel*. Especifica una etiqueta o `rdfs:label` para todas las definiciones de las clases asociadas.

#### ▪ **d2rq:PropertyBridge**

Esta estructura relaciona las columnas de las tablas de la base de datos con las propiedades RDF. Se usa para asociar propiedades a los recursos RDF definidos como `d2rq:ClassMap`. Estas propiedades pueden ser literales, URIs o nodos vacíos que relacionen un recurso con otro. Algunas de las propiedades de esta estructura son las siguientes:

- *d2rq:belongsToClassMap*. Indica que el `d2rq:PropertyBridge` pertenece a un `d2rq:ClassMap`.
- *d2rq:property*. Especifica una propiedad RDF que conecta con el `d2rq:ClassMap` con el objeto o literal creados por el enlace de `d2rq:PropertyBridge`.
- *d2rq:column*. Indica la columna de la base de datos que contiene los valores literales.
- *d2rq:datatype*. Indica el tipo de datos RDF de las literales.
- *d2rq:propertyDefinitionLabel*. Especifica una etiqueta `rdfs:label` para todas las definiciones de propiedad asociadas.

## 8.2. Código fuente

A continuación se describe la jerarquía de ficheros y directorios del código fuente, describiendo la utilidad de los diferentes archivos y su distribución en paquetes o carpetas. La plataforma Java permite el empaquetamiento de dicho sistema de archivos en un fichero con extensión WAR (*Web application ARchive*) como producto de la compilación y enlazado del código fuente, facilitando así su despliegue.

### 8.2.1. Estructura de una aplicación Grails

- grails-app
  - + conf → Ficheros de configuración de la aplicación.
  - + controllers → Las clases que actúan de controlador siguiendo el patrón modelo-vista-controlador (MVC), se encargan de gestionar las peticiones que provienen desde la vista y preparan la respuesta de la aplicación.
  - + domain → son las clases que estarán asociadas con entidades o tablas en la base de datos.
  - + i18n → Archivos necesarios para la internacionalización de la aplicación.
  - + services → Es la capa de servicios de la aplicación, aquí se implementan todas aquellas funciones que necesitamos para la lógica de la aplicación.
  - + taglib → Son las librerías de etiquetas que vayamos a usar.
  - + utils → Clases comunes.
  - + views → Vistas de la aplicación (ficheros .gsp).
- lib → Contiene las distintas librerías que necesita el proyecto.
- scripts → Contiene los scripts.
- src → Son aquellas clases que no sean servicios, un ejemplo de esto podría ser una configuración de la interfaz.
- test → Son las pruebas de la aplicación.
- web-app → Raíz de la aplicación web.

### 8.2.2. Scripts auxiliares de transformación de datos

Se han desarrollado dos scripts, uno en PHP para cargar los datos en la base de datos de la memoria y agilizar el proceso de carga durante las fases de desarrollo y depuración, y otro script en Python para automatizar el cambio del vocabulario que genera el fichero de mapping inicial al vocabulario que nosotros hemos generado en Neologism, con la misma idea de poder refinar los vocabularios durante el proceso de investigación.

El script PHP se creó con el objetivo de cargar los datos desde ficheros CSV a la base de datos de la Memoria, ya que tenía que estar cargando datos continuamente en la base de datos y a la larga era más útil crear un script que cargase cada CSV en la tabla que le corresponde de la base de datos. Para ejecutar el script en PHP se puede ejecutar por consola o desde el servidor accediendo a la URL del script.

El script en Python surgió tras generar el fichero de mapping de la aplicación del D2RQ, al tener que realizar las correspondencias entre el término que genera por defecto el D2RQ y las clases y propiedades que hay que sustituir con el vocabulario para darle la semántica a cada dato referenciado y recurso.

### 8.3. Procesos ETL

La información se encuentra desestructurada por lo que habrá que realizar el siguiente proceso de análisis y tratamiento de datos disponibles:

- **Fase 1:** Me puse en contacto con el departamento que desarrolla la Memoria académica de la Universidad y me facilitaron información en formato Word para poder manejar mejor la información que ya esta disponible en la Web.
- **Fase 2:** Cargué los datos en formato CSV para que estén en formato libre y posteriormente diseñe un script que permite la carga de datos desde los ficheros CSV a la base de datos. La carga de información se realizó teniendo en cuenta que si hay cambios en las siguientes memorias como trabajo futuro se pueda diseñar una interfaz para automatizar la carga de esos datos y así reducir el trabajo manual.
- **Fase 3:** En la tercera fase podemos nombrar el proceso ETL 7.2 que se ha creado con ayuda de la herramienta de Pentaho, donde se han realizado transformaciones desde bases de datos externas que permitan carga los datos en base de datos de forma automática, aunque estos datos no ha sido posible incluirlos en la versión final ya que son datos de carácter sensible y no se ha obtenido los permisos necesarios para que su publicación posterior en el D2RQ Server.
- **Fase 4:** Finalmente como última fase, para verificar los datos que hemos cargado en base de datos de la Memoria de la UCA son correctos y que no tengan ninguna errata se ha utilizado Open Refine para limpieza de los datos. Principalmente, se ha abierto un proyecto en Open Refine y se ha ido clasificando por variables de clasificación como año, id, etc... Para esto se ha utilizado las distintas funciones de la aplicación, principalmente Text Facet y Numeric Facet. Si se desee consultar otras funciones de Open Refine para su utilización en profundidad consulte esta guía de limpieza de datos.

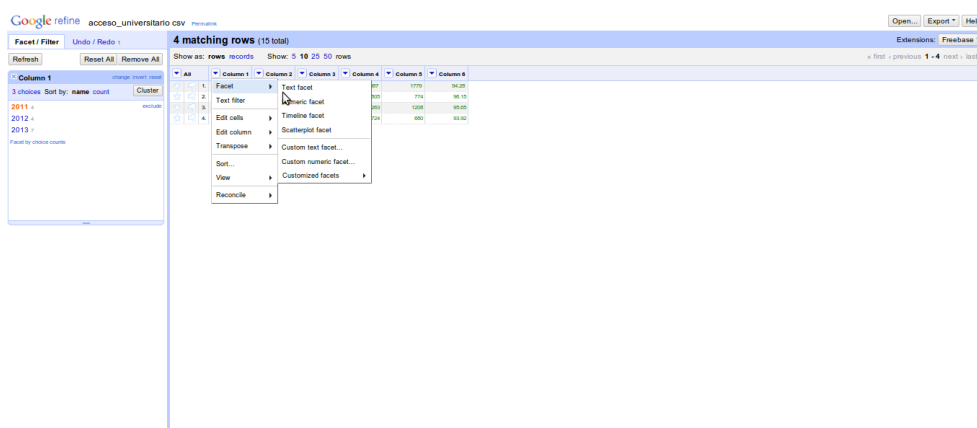


Figura 8.2: Limpieza de datos con Open Refine

Se ha desarrollado un proceso ETL (extract, transform and load) donde utilizamos Spoon de Pentaho para implementar el proceso y Oracle Developer<sup>7</sup> como ayudante para la conexión a la

<sup>7</sup><http://www.oracle.com/technetwork/developer-tools/developer-suite/overview/index.html>

base de datos. Una vez realizada la conexión podremos crear la primera conexión a los distintos servidores de la Universidad que contiene alguna base de datos sobre infraestructura. Como realmente la información que nos interesaba consultar de las bases de datos de la UCA son realmente 4 tablas, realizaremos tres transformaciones distintas que uniremos entre sí en un fichero de trabajo.

El proceso de desarrollo del ETL con la herramienta Spoon será siempre el mismo, primero se crea la configuración con las bases de datos 8.3 origen y destino, después se eligen los pasos que se deseen utilizar para las transformaciones, si se desea más información, consultar la Wiki [2]. En nuestro caso, casi todas las transformaciones van a utilizar un paso de entrada de tabla que extrae información de una tabla de la base de datos y el segundo paso será salida de tabla donde se carga la información en la base de datos de salida. Una vez realizada todas las transformaciones necesarias se unirán en un trabajo para ejecutarlas secuencialmente.

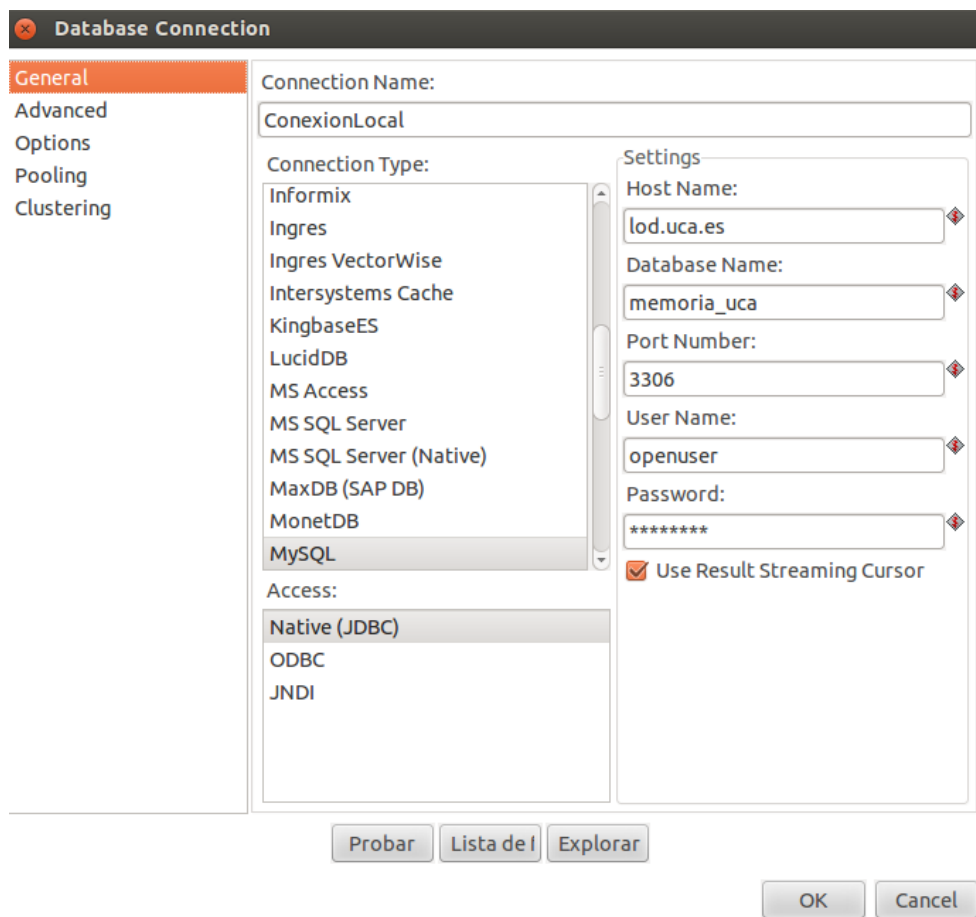


Figura 8.3: Configuración de la base de datos

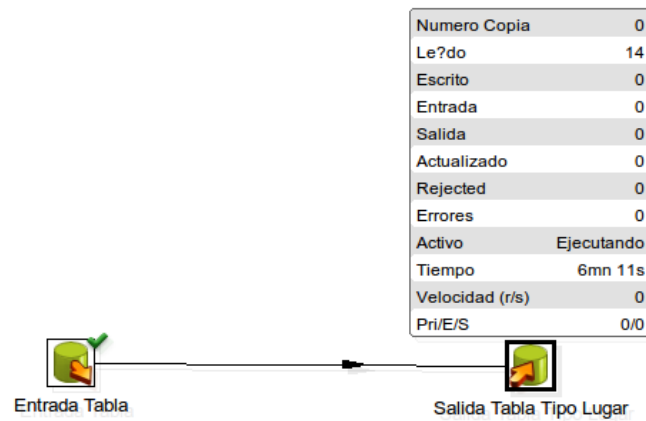


Figura 8.4: Ejemplo de transformación

Las cuatro transformaciones del trabajo 1 son:

- Transformación de Campus.
- Transformación de Tipo de Lugar.
- Transformación de Edificio.
- Transformación de Aula.

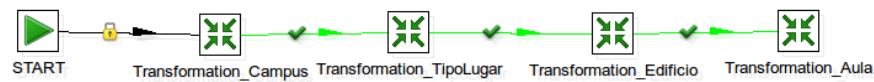


Figura 8.5: Trabajo 1

Toda esta información pasará por una base de datos intermedia (figura 8.6) que diseñamos y en el segundo trabajo se realiza la transformación desde la base de datos intermedia, que en nuestro caso se encontraba en el equipo local, a la base de datos del servidor de D2RQ.

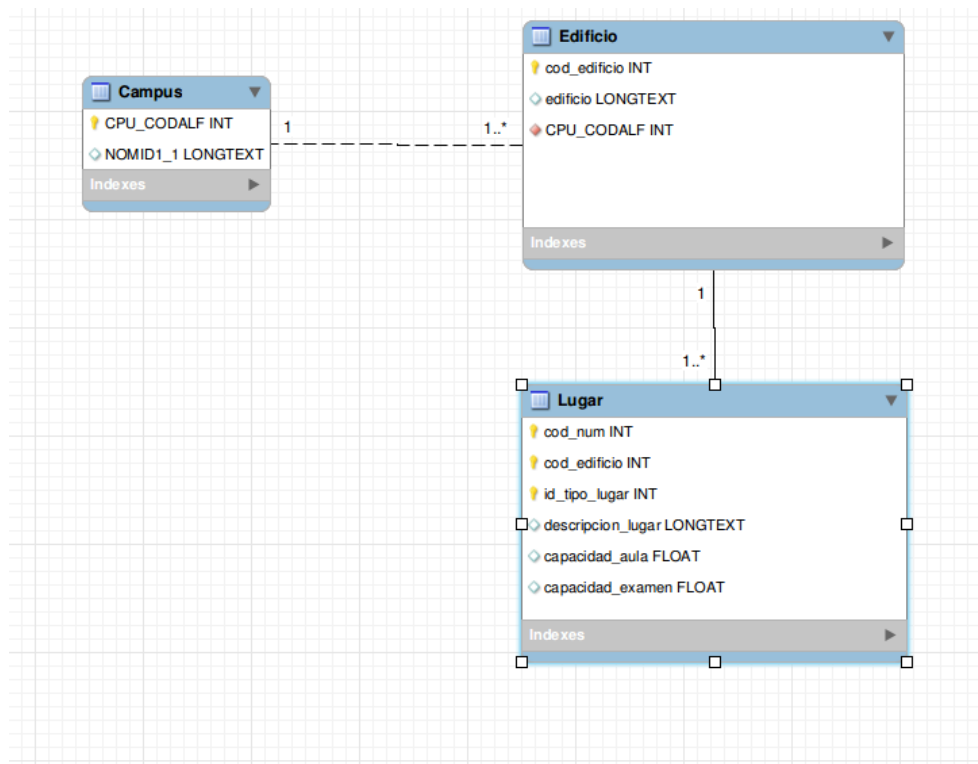


Figura 8.6: Base de datos de infraestructura intermedia

Las tres transformaciones del trabajo 2 son:

- Transformación de Edificios.
- Transformación de Lugares.
- Transformación de Aulas.

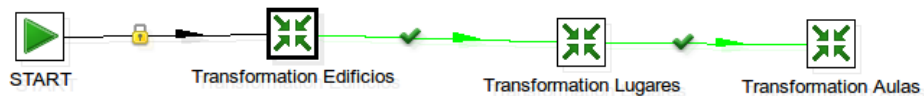


Figura 8.7: Trabajo 2

Debido al acuerdo de confidencialidad, estos datos no se han podido publicar en el resultado final, con lo cual se ha decidido documentar el trabajo hecho como demostración de lo que se podría hacer con la herramienta Spoon de Pentaho de cara a los futuros programadores que van a organizar la información.



## Capítulo 9

# Pruebas del Sistema

En este capítulo se documentan los diferentes tipos de pruebas que se han llevado a cabo, ya sean manuales (mediante listas de comprobación) o automatizadas mediante algún software específico de pruebas.

### 9.1. Pruebas unitarias y de integración

Para este proyecto las pruebas unitarias que se hicieron fueron manualmente, probando que los datos que están en la memoria académica que coinciden con los datos mostrados en la aplicación web. Para realizar este proceso, se realizará de forma visual a través de los gráficos. A continuación vamos a mostrar un ejemplo:

Primero, buscamos una tabla de datos de ejemplo en la memoria.9.1

Evolución de la estimación de número de estudiantes en equivalencia a tiempo completo

Curso	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13
Alumnos de 1er y 2º ciclos	15.982	15.660	14.668	11.318	7.901	4.907
Alumnos de grado	-	-	1.403	5.304	9.082	11.768
Alumnos de másteres oficiales	326	439	935	946	913	875
<b>TOTAL equivalentes a tiempo completo</b>	<b>16.308</b>	<b>16.099</b>	<b>17.006</b>	<b>17.568</b>	<b>17.896</b>	<b>17.550</b>

Figura 9.1: Datos de la Memoria Académica 2012-2013

Y después, probamos que la consulta sale correctamente en el endpoint del D2RQ, como la aplicación va a realizar dicha consulta. El resultado de la consulta en el endpoint y los resultados del gráfico de la imagen 9.2 será exactamente lo mismo.

The screenshot shows a web application interface with a breadcrumb trail: 'Principal / Datos de Alumnos / Alumnos a tiempo completo'. Below this is a table with two columns: 'Curso académico' and 'Alumnos'. The table contains the following data:

Curso académico	Alumnos
2007-2008	16308
2008-2009	16099
2009-2010	17006
2010-2011	17568
2011-2012	17906
2012-2013	17550

To the right of the table, there is a section titled 'Variables de clasificación' with two radio buttons: 'Curso académico' (selected) and 'Titulación'. Below this is a section titled 'Tipo de gráfico' with a dropdown menu set to 'Tabla de datos' and a 'Mostrar' button.

Figura 9.2: Gráfico de la Aplicación

Dada la arquitectura altamente modular en la que se basa este proyecto, la prueba de integración consiste en comprobar que la comunicación e interacción entre los mismos se efectúan de forma satisfactoria para garantizar el correcto funcionamiento del conjunto de la plataforma.

Las pruebas de integración se han efectuado de una forma manual, siempre disponiendo de la información más actualizada de todo el conjunto de datos publicado cedido por la Universidad de Cádiz, que se supone correcta.

El funcionamiento y las pruebas de la plataforma de publicación de vocabulario Neologism se realizó al mismo tiempo que se generaron los vocabularios que se importarían en D2RQ para generar el modelo de datos del conjunto de datos provisto por la Universidad. Dado que esta importación se realizaría usando un conjunto de ficheros con formato RDF generados por Neologism, se revisaron cuidadosamente los ficheros para comprobar que el modelo creado fuese el correcto.

Para validar los datos presentados por la aplicación de visualización *Linked Data Viewer*, se ha comparado las cifras mostradas en los gráficos con su correspondiente resultado fruto de ejecutar la misma consulta SPARQL contra el endpoint local generada por la aplicación y obtenida por rutinas de depuración del programa en ejecución, a través de la interfaz de ejecución de consultas que provee D2RQ<sup>1</sup>, corroborando que el resultado obtenido es idéntico.

A su vez, dado que D2RQ genera un grafo virtual del modelo de datos almacenado en una base de datos MySQL. Para verificar que la información que obtiene es correcta, se comprueba que los resultados de la consulta SPARQL generada con el caso anterior coincide con el dato almacenado en la tabla correspondiente a la propiedad del vocabulario necesario, siguiendo la relación presentada en el punto 7.3.

## 9.2. Pruebas de sistema

En esta actividad se realizan las pruebas de sistema de modo que se asegure que el sistema (tanto el software como el hardware) cumple con todos los requisitos establecidos: funcionales, de almacenamiento, reglas de negocio y no funcionales. Se suelen desarrollar en un entorno específico para pruebas.

<sup>1</sup><http://lod.uca.es/d2rq/snorql/>

### 9.2.1. Pruebas funcionales

Con estas pruebas se analiza el buen funcionamiento de la implementación de los flujos normales y alternativos de los distintos casos de uso del sistema.

#### Pruebas exploratorias

Para la elaboración de este proyecto, al poseer un componente importante de investigación, resulta más óptimo disponer de una metodología de pruebas que se pueda adaptar con facilidad en función de los avances de la investigación. Como hemos elegido una metodología de desarrollo ágil e iterativa que permite cambios rápidos en función del curso del proyecto y los avances de la investigación, en lugar de elegir una metodología de pruebas unitarias basadas en código, que nos obligaría estar modificando y añadiendo nuevos casos de pruebas constantemente, se ha optado por un enfoque exploratorio donde se han diseñado diferentes sesiones de pruebas realizadas por diferentes testers procedentes de campos con diferentes niveles de experiencia, con el objetivo de amoldar las pruebas de las diferentes exigencias de las pruebas de desarrollo.

El enfoque de pruebas exploratorias es una aproximación a la prueba del software que unifica simultáneamente el aprendizaje de la aplicación, diseño de pruebas y ejecución de las mismas. Según su propio creador, Cem Kaner:

"(El testing exploratorio) es un estilo de prueba de software que enfatiza la libertad personal y la responsabilidad del tester como individuo para optimizar de forma continua la calidad de su trabajo, tratando el aprendizaje derivado de sus pruebas, el diseño de las pruebas, la ejecución de las pruebas y la interpretación de resultados como actividades que se soportan mutuamente, realizándose en paralelo durante el ciclo de vida del software."[\[14\]](#)

En esencia, el testing exploratorio permite al probador usar el conocimiento adquirido acerca del diseño y funcionamiento de la aplicación, junto a su propia creatividad y experiencia en el campo de la ingeniería del software, para diseñar y ejecutar pruebas que se adapten a cada caso. Esto dota al probador de una flexibilidad sin límites para decidir qué o cómo debe probar un caso de uso o funcionalidad concreta, y agiliza la detección y corrección de fallos, permitiendo una comunicación fluida ente probador y desarrollador.

El resultado obtenido es que se han descubierto y corregido fallos de muy diversa naturaleza, a la vez de obtenerse un flujo continuo de sugerencias muy valiosas para la mejora de la usabilidad de la interfaz, punto que se enfatizó durante las sesiones de pruebas debido a la orientación a la facilidad de uso universal del proyecto. Toda esta información proporcionada por los expertos en pruebas ha sido determinante para el resultado de la aplicación, la cual no se hubiese podido obtener con una batería de pruebas tradicional.

### 9.2.2. Pruebas no funcionales

Estas pruebas pretenden comprobar el funcionamiento del sistema, con respecto a los requisitos no funcionales definidos en la etapa de análisis: rendimiento, accesibilidad, etc.

## Validación HTML

Para corroborar que el contenido de los diferentes documentos HTML generados por la herramienta de visualización son conformes al último estándar HTML promulgado por el *World Wide Web Consortium* (W3C), tal y como se indica en el apartado de soluciones tecnológicas usadas para la implementación de la interfaz de usuario, se ha sometido la susodicha herramienta a un proceso de validación de sintaxis y estructura, usando para ello el *W3C Markup Validation Service*, una herramienta publicada por el propio W3C<sup>2</sup>, cuya misión es analizar y reportar fallos encontrados en la estructura de un documento HTML que difieran de las reglas que se encuentran en el texto del estándar.

El sitio se validó contra el estándar HTML5, con resultados satisfactorios a excepción de una advertencia en un atributo *rel* establecido por el plugin Select Picker a la hora de realizar transformaciones en los cuadros de texto desplegados, y que parece ser causado porque no se ha tenido en cuenta en la validez del código HTML producido por el mismo.

## Pruebas de estrés

Estas pruebas se conocen como “stress testing”, el objetivo principal de esta prueba es obtener datos como resultado de la generación de carga en el sistema para comprobar su funcionamiento, con la finalidad de comprobar el número de usuarios concurrentes que es capaz de soportar hasta su completa degradación.

En estas pruebas influye tanto el rendimiento que pueda alcanzar la aplicación, como la potencia del hardware subyacente y la óptima configuración de todos los servicios del sistema, por lo que no se puede considerar una prueba absoluta, sino más bien una referencia inicial con la que comenzar a trabajar para la optimización de la plataforma en conjunto (software + hardware).

La prueba de estrés se ha realizado usando la herramienta Apache Bench<sup>3</sup>, una utilidad que permite realizar una batería de peticiones concurrentes a un sitio web, midiendo parámetros como el tiempo empleado en servir la página, el número de peticiones erróneas, el total de datos transferidos, etc.

Apache Bench forma parte del paquete *apache2-utils*, disponible en los repositorios de Debian. Para instalarlo, basta con ejecutar

```
apt-get install apache2-utils
```

La ejecución del programa se realizó de la siguiente forma, con una batería de 1000 iteraciones, cada una realizando 20 peticiones concurrentes, escribiendo los resultados en un fichero para su posterior análisis:

```
ab -n 1000 -c 20 -g output.plot http://lod.uca.es/viewer
```

Los resultados de la prueba de estrés fueron los siguientes:

---

<sup>2</sup><http://validator.w3.org/>

<sup>3</sup><https://httpd.apache.org/docs/2.2/programs/ab.html>

This is ApacheBench, Version 2.3 <\$Revision: 655654 \$>  
 Copyright 1996 Adam Twiss, Zeus Technology Ltd, <http://www.zeustech.net/>  
 Licensed to The Apache Software Foundation, <http://www.apache.org/>

Benchmarking lod.uca.es (be patient)

```

Server Software:      Apache-Coyote/1.1
Server Hostname:     lod.uca.es
Server Port:         80

Document Path:       /viewer
Document Length:     0 bytes

Concurrency Level:   20
Time taken for tests: 13.181 seconds
Complete requests:   1000
Failed requests:     0
Write errors:        0
Total transferred:   223000 bytes
HTML transferred:    0 bytes
Requests per second: 75.87 [#/sec] (mean)
Time per request:    263.613 [ms] (mean)
Time per request:    13.181 [ms] (mean, across all concurrent requests)
Transfer rate:       16.52 [Kbytes/sec] received

```

#### Connection Times (ms)

	min	mean[+/-sd]	median	max
Connect:	73	125 21.0	120	229
Processing:	91	135 22.0	137	247
Waiting:	91	133 22.2	135	247
Total:	187	260 25.9	258	469

#### Percentage of the requests served within a certain time (ms)

50%	258
66%	264
75%	271
80%	274
90%	282
95%	293
98%	308
99%	396
100%	469 (longest request)

A continuación se muestra un gráfico del tiempo de respuesta de las peticiones realizadas por la prueba de estrés durante el transcurso de la batería de pruebas:

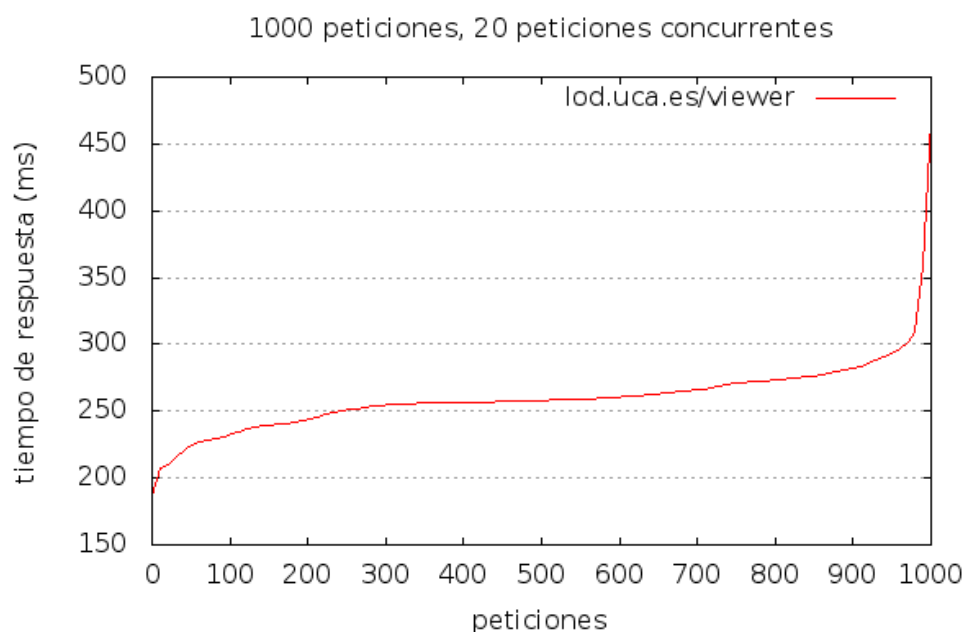


Figura 9.3: Gráfico comparativo de tiempo de carga durante la prueba de estrés

Como se puede ver, tanto en la salida de la prueba como en el gráfico, es que el sistema se comporta de una forma bastante estable con el aumento de peticiones, hasta el momento en el que el sistema alcanza el techo de rendimiento, momento en el que se satura y comienza a degradarse, elevándose el tiempo de respuesta de forma rápida. Cabe decir que un benchmark por sí mismo no tiene sentido, sino que sólo se podrá interpretar en el contexto de la aplicación, puesto que dependiendo del público objetivo deberemos estimar un mayor o menor número de visitas y peticiones concurrentes hechas a la plataforma. Dado que no se dispone información con respecto al número de visitantes que esta plataforma podría tener, se ha realizado una prueba con unas cifras acordes a las dimensiones de la plataforma sobre la que se ha realizado las pruebas (ver A.2), puesto que, de igual forma, esta podrá comportarse de otra manera completamente diferente en otro sistema.

### 9.3. Pruebas de aceptación

El objetivo de estas pruebas es demostrar que el producto está listo para el paso a producción. Suelen ser las mismas pruebas que se realizaron anteriormente pero en el entorno de producción. En estas pruebas, es importante la participación del cliente final, en nuestro caso, el tutor del proyecto y el departamento al que se adscribe el mismo, quienes mostraron su conformidad con el trabajo realizado y aceptando los resultados obtenidos, expuestos en la presente memoria.

## 9.4. Caso de estudio

Hoy en día, todas las universidades, al igual que las administraciones públicas, deben publicar sus datos gracias a la Ley de Transparencia. Hasta ahora, muchas de ellas lo hacían igual que nuestra universidad, publicando los datos en una página web o en archivos que eran difícilmente accesibles y bastante arduos de leer. Con la iniciativa de Linked Open Data lo que se pretende es que las Universidades publiquen sus datos entrelazados para su reutilización por entidades externas y no sólo personal cercano a la Universidad.

### 9.4.1. Universidad Pablo de Olavide

La Universidad Pablo de Olavide de Sevilla posee un portal de datos abiertos y enlazados<sup>4</sup> como una iniciativa dentro de un plan de Reutilización de Datos (RISP) [13]. La universidad puso en marcha un portal web a través del cual ofrece un repositorio con todos los datos, permitiendo su descarga en formatos RDF y CSV y proporcionando métodos de búsqueda y filtrado para los mismos. El portal incluye bastante documentación acerca tanto de la iniciativa como de los conceptos más usados sobre Linked y Open Data.

A nivel técnico, el portal implementa un sistema basado en Apache Marmotta<sup>5</sup>, una plataforma de código libre basada en Java que proporciona una base inicial sobre la que desarrollar el portal. En cuanto al número de dataset en el CKAN que forma parte de nuestro sistema cuenta con un número de conjunto de datos mayor, y el de la UPO solo tiene en total 11 conjuntos de datos. Además, como puede verse en la imagen 9.5, el catalogo de datos de la UPO es de 4 estrellas [7] ya que existen datos enlazables a través de URIs pero no existen enlaces a otros datos.

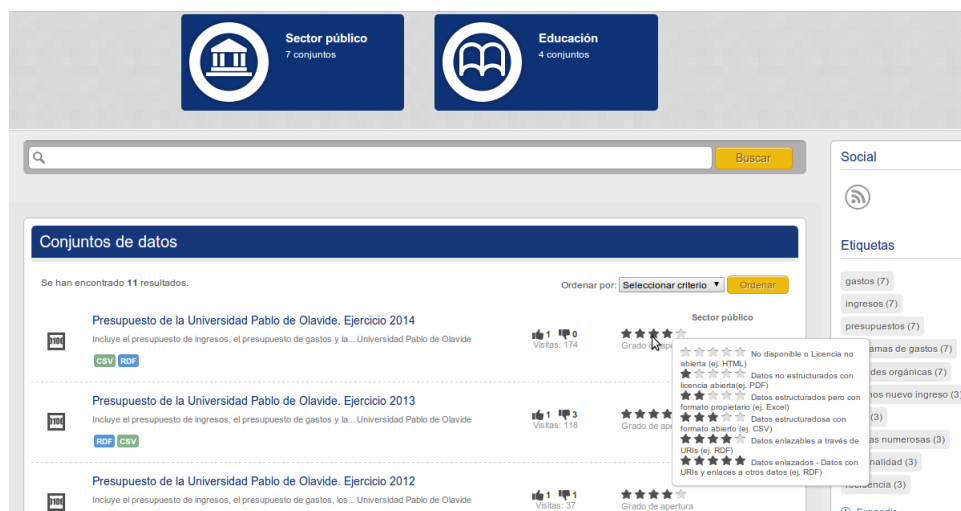


Figura 9.4: Conjunto de datos de la Universidad Pablo de Olavide

Con esta iniciativa, la universidad ha publicado información acerca de sus memorias anuales y datos relevantes sobre sus infraestructuras, titulaciones y presupuestos, logrado un mayor nivel de transparencia para con el resto de ciudadanos y centralizado en un portal la adquisición de

<sup>4</sup><http://datos.upo.gob.es/>

<sup>5</sup><https://marmotta.apache.org/>

toda esta información, proporcionándoles el reconocimiento y la visibilidad que una institución como la universidad debe exigirse en el ejercicio de todas sus funciones.

Con respecto a los planes de futuro, la entidad planea seguir reforzando su apuesta en materia de Open Data, con un compromiso a seguir abriendo su información al público, recogido en el III Plan Estratégico UPO 2014-2016 [13].

### 9.4.2. Comparación de la UPO con la UCA

Al comparar los datos de la UPO con los que hemos publicado en este proyecto, nos damos cuenta de que el catálogo de datos disponible en la Universidad de Cádiz es mucho mayor. Además, los gráficos disponibles ya vienen hechos y no pueden ejecutarse en tiempo real, tal y como puede realizarse con el producto final Linked Data Viewer.

Al utilizar vocabularios generalizados, que pueden usarse para todas las universidades, si una universidad publica sus datos reutilizando los vocabularios RDF entonces habrá más datos entrelazados y por lo tanto podemos comparar dicho datos sacando estadísticas, evaluaciones o un gráfico comparativo como es el caso de la aplicación web *Linked Data Viewer*. En la siguiente imagen se muestra un ejemplo que se ha implementado en la aplicación con el fin de demostrar las ventajas y privilegios que puede ofrecernos la tecnología Linked Data.

En este ejemplo, comparamos los datos sobre presupuestos que hemos encontrado en el portal UPO, subiendo un fichero RDF a la aplicación web y seleccionamos el tipo de variable “Presupuestos”.

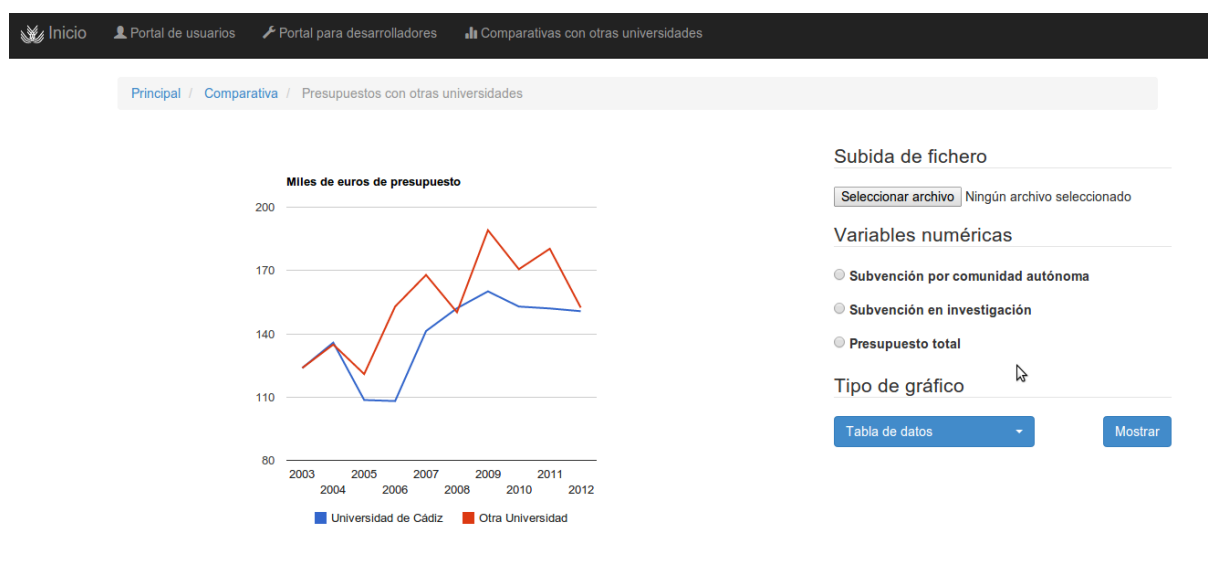


Figura 9.5: Comparativa de Presupuestos de la UCA con la UPO en barras



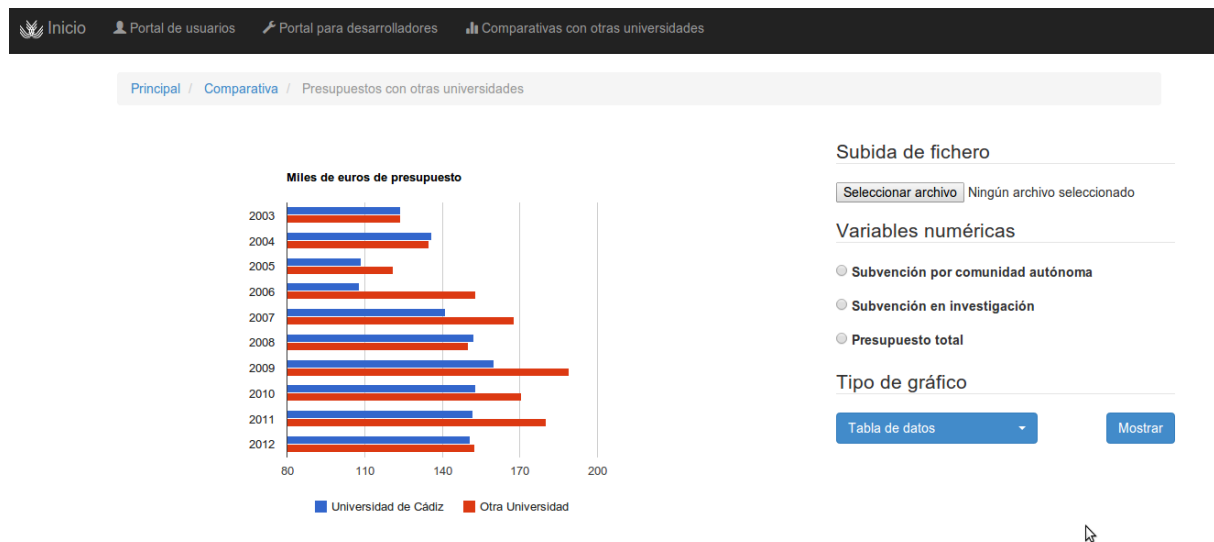


Figura 9.6: Comparativa de Presupuestos de la UCA con la UPO en líneas

La idea de Linked Open data es que además de publicar estos datos haya aplicaciones que los consuman y en el caso de portal de la UPO, no se ha asociado al portal ninguna aplicación en la que se pueda consumir los datos publicados.



Parte III

Epílogo



# Capítulo 10

## Conclusiones

En este último capítulo se detallan las lecciones aprendidas tras el desarrollo del presente proyecto y se identifican las posibles oportunidades de mejora sobre el software desarrollado.

### 10.1. Objetivos cumplidos

A modo de conclusión, a continuación se enumeran los objetivos generales y específicos alcanzados con el proyecto detallado en los capítulos anteriores.

- Obtener una visión general de la evolución de la web hasta el desarrollo de la Web Semántica, dada la estrecha relación del grupo de investigación con la *World Wide Web*, resulta interesante no sólo mantener sino ampliar todo lo relativo a la evolución de la red de redes desde sus inicios hasta nuestros días a fin de poder identificar tendencias futuras y nuevas líneas de investigación y mejora.
- Introducir el concepto de datos abiertos enlazados (LOD, *Linked Open Data*) y trabajar con ellos, por el especial interés de la Universidad por el software libre y el conocimiento abierto.
- Analizar las distintas ontologías, que sobre un mismo tema, definen diferentes fuentes de datos, llegando a diseñar e introducir nuevos vocabularios que puedan ser usado a nivel global para describir datos que puedan ser compartidos y tratados de forma uniforme por todos.
- Usar URIs y vincular los datos para obtener mejores búsquedas, cumpliendo los requisitos 5 estrellas de Linked Data.
- Estudiar los vocabularios propuestos por las instituciones que definen el desarrollo de la Web Semántica y los que usan distintas fuentes de datos, reutilizando así el conocimiento ya desarrollado y aunando esfuerzos con otras entidades, en lugar de dividirlo proponiendo una solución alternativa.
- Analizar fuentes de datos, estudiar su documentación, modelo de datos, problemas, virtudes, etc., como forma de identificar nuevas fuentes de conocimiento, atesorando un poderoso *know-how* que pueda servir a otros investigadores en sus futuros esfuerzos.
- Desarrollar un almacén de datos donde centralizar la publicación de datos académicos de la UCA.

- Poner en práctica todos los conocimientos adquiridos mediante la realización de una aplicación real, a modo de prueba de concepto y aproximación inicial a una iniciativa completa por parte de la Universidad donde se involucre todo el personal y recursos.
- Como resultado de lo anterior, obtener una aplicación capaz de realizar comparaciones con otras universidades, como una herramienta valiosa para mostrar el potencial de esta tecnología y los beneficios para toda la sociedad que supone poder acceder a la información de entidades públicas de forma transparente, rápida y sencilla.

## 10.2. Lecciones aprendidas

Tras el desarrollo de este proyecto he podido aprender como analizar y diseñar un sistema de datos estructurados bastante amplio y las buenas prácticas que deben de llevarse a cabo a la hora de implementarlo y la redacción de su documentación. Además he tenido la suerte de poder aprender tecnologías semánticas que me han hecho darme cuenta del potencial que tiene la Web de los Datos y la capacidad de crecimiento y oportunidades que supone la misma.

Para el desarrollo del proyecto he tenido que investigar, aprender y utilizar los siguientes conceptos, tecnologías y herramientas:

- La Web Semántica, su concepto, filosofía y entorno.
- Diseño de ontologías y vocabularios semánticos, analizando los ya existentes y reutilizando la información necesaria.
- Lenguajes de consulta y descripción de datos semánticos que desconocía completamente, como SPARQL y RDF.
- Instalación y configuración de un sistema de información de mediana complejidad, con diferentes servicios y aplicaciones que dependen las unas de las otras.
- Desarrollo de una aplicación usando el lenguaje Groovy, y frameworks como Bootstrap para maquetar la web y Apache Jena para trabajar con datos semánticos usando las ontologías creadas.

A nivel personal, este proyecto me ha permitido realizar un trabajo original y adquirir conocimientos interesantes que pueden ayudarme de cara a mi futuro profesional, además de sentirme realizada habiendo conseguido todos los objetivos propuestos al inicio del proyecto y por la aportación que mi trabajo pueda hacer en un futuro a la comunidad universitaria.

## 10.3. Trabajo futuro

- En el futuro, si la iniciativa Linked Data sigue adelante, se tendría que hablar con el departamento que lleva a cabo el diseño de la memoria de la universidad, con la idea que sea posible el acceso a los datos que se desean publicar y así evitar el problema de confidencialidad con los datos que ha surgido durante el presente proyecto. Para esto, la administración de la Universidad debería de apoyar la iniciativa, facilitando los datos para su publicación.

- Integrar el dataset UCA al Linked Data Cloud<sup>1</sup>, el repositorio centralizado donde se publica la información, la cual pasa a formar parte del grafo mundial de datos abiertos e interconectados. Se ha dejado como trabajo futuro, ya que debido al proceso de verificación al que son sometidos todos los datos que se envían, resulta imposible su finalización para la fecha de exposición de este proyecto.
- También se ha contactado con la organización Linked Open Vocabularies para que publiquen los vocabularios en su buscador<sup>2</sup> y así cumplir el principio de reutilización de vocabularios por otras universidades, al igual que hemos hecho nosotros para crear el nuestro. De nuevo, este proceso no es inmediato y requiere la aceptación previa del conjunto de vocabularios por parte de los responsables de la iniciativa.
- En cuanto a la ampliación del sistema de información, me ha parecido muy interesante la aproximación usada por EduBase, proyecto que se comentó junto con otros sistemas basados en Linked Data (ver 3.2). Ofrecer una API REST en conjunción con el endpoint SPARQL ofrece una interfaz de consulta distinta, lo cual facilita el trabajo a los desarrolladores que, por preferencia o limitaciones de diseño, decidan consultar la información por otra vía, lo cual ayudará a la difusión de este tipo de proyectos.

---

<sup>1</sup><http://lod-cloud.net/>

<sup>2</sup><http://lov.okfn.org/dataset/lov/>





Parte IV  
Apéndice



## Apéndice A

# Manual de instalación y explotación

En esta sección se presenta un manual de despliegue y configuración del proyecto con los pasos más importantes y acciones a tener en cuenta para la correcta puesta en funcionamiento del conjunto de aplicaciones que componen el sistema de información desarrollado en este trabajo.

### A.1. Introducción

El sistema Linked Open Data está formado por un conjunto de componentes independientes que, funcionando en conjunto, forman una plataforma tecnológica sobre la que se desarrolla un sistema de almacenaje, publicación y consulta de datos relativos a la memoria académica de la Universidad de Cádiz mediante Linked Data. Ofrece la posibilidad de consultar los datos publicados de forma automática mediante el lenguaje de consultas SPARQL, así como a través de una aplicación web pensada para ciudadanos no expertos en la materia, permitiendo representar la información de forma gráfica y realizar comparaciones con otras universidades que publiquen sus datos como linked data.

### A.2. Requisitos previos

El proyecto se ha realizado en un servidor de integración provisto por el departamento, cuyas características técnicas son las siguientes, a la espera de su publicación definitiva en un entorno de producción:

Características	
Procesador	AMD Athlon(tm) 64 X2 Dual Core Processor 5200+
Memoria RAM	2009MB
Sistema Operativo	Ubuntu Server 10.04 LTS (Linux Kernel 2.6.32-38-generic)

### A.3. Inventario de componentes

Lista de los componentes software que se incluyen en la versión del producto.

1. **Apache.** Servidor web HTTP, que nos permite visualizar la Web y el D2R server que muestra la base de datos mapeada.

2. **Apache Tomcat.** Es el servidor web utilizado para soportar los servlets que se usan en esta plataforma.
3. **D2R Server.** Es una aplicación web que permite el acceso a la base de datos a través del protocolo SPARQL, con Linked Data y una simple interfaz HTML.
4. **Aplicación web.** Esta aplicación web es la que va a consumir los datos de la memoria que podemos visualizar gracias al D2RQ Server.
5. **CKAN.** Es un portal que nos permite publicar el dataset de la Memoria de la Universidad de Cádiz y añadir otros existentes.
6. **Neologism.** Es una aplicación web que sirve como editor de ontologías.

## A.4. Procedimientos de instalación

### A.4.1. D2RQ

Para la instalación del D2RQ y la aplicación Linked Data Viewer necesitamos previamente haber instalado Apache y Tomcat donde colocaremos en la carpeta de Tomcat (`/var/lib/tomcat6/webapps`) el WAR comprimido.

1. Descargamos el `d2rq.war` del disco de instalación proporcionado y lo insertamos en la carpeta del Tomcat `/var/lib/tomcat6/webapp`.
2. Después desplegaremos en el directorio de Tomcat. Por esta razón debemos de reiniciar Tomcat.

```
sudo /etc/init.d/tomcat6 stop
sudo /etc/init.d/tomcat6 start
```

Para poder visualizar Tomcat por el protocolo HTTP tendremos que instalar Apache y configurarlo mediante estas líneas en el fichero `httpd.conf`, usando el módulo ProxyPass de Apache.

#### Httpd.conf

```
ServerName lod.uca.es
```

```
ProxyPass /d2rq http://lod.uca.es:8080/d2rq
ProxyPassReverse /d2rq http://lod.uca.es:8080/d2rq
JkMount /d2rq apj13
```

```
ProxyPass /viewer http://lod.uca.es:8080/viewer
ProxyPassReverse /viewer http://lod.uca.es:8080/viewer
RedirectMatch ^/$ /viewer
```

```
<Proxy *>
    Order deny,allow
    Allow from all
</Proxy>
```

## CKAN

Para instalar CKAN debemos de seguir los siguientes pasos, recogidos a modo de resumen. La información más completa y actualizada se encuentra en el sitio web oficial de CKAN [3]:

```
sudo apt-get update
sudo apt-get install -y wget
echo "deb http://apt.ckan.org/ckan-1.8 lucid universe" | sudo tee /etc/apt/sources.list.d
/ckan.list
wget -q0- "http://apt.ckan.org/packages_public.key" | sudo apt-key add -
sudo apt-get update
sudo apt-get install -y ckan postgresql-8.4 solr-jetty
```

Una vez instalado, dejaremos que CKAN configure automáticamente los almacenes de datos y crearemos una instancia ingresando el nombre de la instancia (`std`), el nombre de dominio (`lod.uca.es`) y que deseamos que se configure automáticamente una conexión a la base de datos local (`yes`):

```
sudo ckan-setup-solr
sudo ckan-create-instance std lod.uca.es yes
```

Activaremos el VirtualHost de la instancia creada anteriormente que la instalación de CKAN ha creado automáticamente:

```
sudo a2ensite std
sudo service apache2 reload
```



# Apéndice B

## Manual de usuario

En esta sección se especifica el manual de usuario de la aplicación que hemos desarrollado como resultado final de este Proyecto Fin de Carrera.

### B.1. Introducción

El objetivo es proporcionar al usuario final un portal web desde donde consultar de forma dinámica información de la Universidad de Cádiz.

Además se han realizado un prototipo de dos ejemplos en los que se pueden comparar los datos de nuestra universidad con otra universidad. De esta forma se ha conseguido cumplir los requisitos establecidos por la filosofía Open Data que además de disponer los datos en formato no propietarios, es decir mostrar los datos en RDF, se permiten realizar un ejercicio de transparencia de la Universidad como entidad administrativa hacia el ciudadano.

El alcance que se desea conseguir con el desarrollo de este software es conseguir que en un futuro se pueda dar un paso más en la publicación de datos según la escala Open Data, pasando de las tres estrellas en las que se encuentra ahora mismo la web de la memoria académica que abusa del contenido estático en tablas y tiene poco contenido gráfico a una Web Semántica donde la ciudadanía pueda acceder a los datos de una manera más ágil y rápida que consultado un simple PDF o una web con demasiado contenido.

### B.2. Requisitos previos

Los requisitos previos que necesita el usuario para el correcto uso de la plataforma, es un navegador web actualizado puesto que toda la información esta disponible en formato web y la interacción con el sistema se realiza por esta vía.

### B.3. Utilización

Se ofrece a continuación una descripción del área de trabajo del sistema y las instrucciones concretas para hacer uso de las funcionalidades básicas más comunes del mismo usando las diferentes aplicaciones de las que se compone la plataforma de Linked Open Data.

### B.3.1. CKAN

CKAN<sup>1</sup> es una herramienta para construir sitios web de datos abiertos.

En CKAN, los datos se publican empaquetados en unidades llamadas "datasets". Un dataset es un paquete de datos, por ejemplo, podría ser las estadísticas del crimen para una región, las cifras de gasto para un departamento del gobierno, o las lecturas de temperatura de varias estaciones meteorológicas. Cuando los usuarios buscan datos, los resultados de búsqueda que ven serán conjuntos de datos individuales.

#### Buscar un dataset

Para encontrar datasets en CKAN, escriba cualquier combinación de palabras de búsqueda en el cuadro de búsqueda en cualquier página. CKAN muestra la primera página de los resultados en esta categoría. Si hay un gran número de resultados, los filtros pueden ser muy útiles, ya que se puede combinar filtros, añadir de forma selectiva y modificar y repetir la búsqueda con los filtros existentes todavía en su lugar.

#### Publicar un dataset

Para poder publicar un dataset es preciso estar identificado en el sistema con una cuenta con privilegios de publicación. Consulte a su administrador del sistema en caso de duda.

Para publicar un nuevo dataset, diríjase al enlace "Add a dataset" en el menú principal de la aplicación. CKAN le pedirá la siguiente información acerca de sus datos:

- Título. Este título será único para todos los datasets existentes en CKAN.
- Descripción. Puede agregar una descripción más larga del dataset aquí, incluyendo información como de dónde se han obtenido los datos y cualquier información que la gente necesite saber cuando se utilizan los mismos.
- Licencia. Es importante incluir información de la licencia para que la gente sepa cómo se pueden utilizar los datos. Este campo es una caja desplegable. Si necesita utilizar una licencia no en la lista, póngase en contacto con el administrador del sistema.
- Recursos. En este campo es donde se añaden los recursos que contienen los datos para este dataset. Elija un archivo o enlace de su recurso de datos y seleccione la opción adecuada en la parte superior de la sección: subir un fichero desde su ordenador o importarlo a través de una URL o API.

---

<sup>1</sup><http://lod.uca.es/ckan>



### Add a Dataset

**Title**

**Uri**

Type at least two characters...  
2+ characters, lowercase, using only 'a-z0-9' and '-\_'

**License**

(Don't worry if you don't know which license the data has been released under).

**Description**

You can use **Markdown formatting** here.

**Groups** Add to:

---

### Add Resources

Upload or link data files, APIs and other materials related to your dataset.

[Link to a file](#) [Link to an API](#) [Upload a file](#)

Figura B.1: Creación de dataset

## Descargar recursos

Una vez localizado un recurso de interés dentro de un dataset, es posible proceder a su descarga a través del botón “Download” situado en la cabecera de la página, junto al título del recurso.

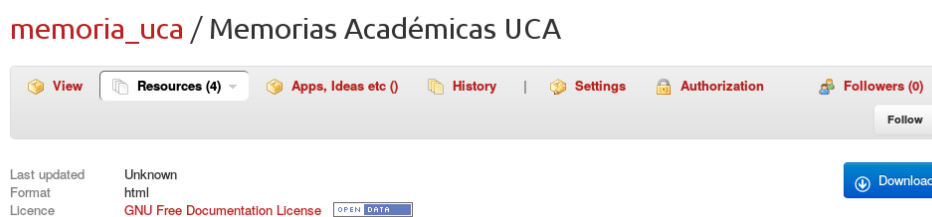


Figura B.2: Descarga de un recurso

### B.3.2. D2R Server

D2R Server<sup>2</sup> es una aplicación que permite interactuar con el servidor D2RQ que publica datos en RDF de la base de datos interna del sistema. Es posible usar una interfaz HTML de ejecución

<sup>2</sup><http://lod.uca.es/d2rq>



The image shows a SPARQL query interface. At the top, it says "SPARQL:". Below that is a list of prefixes with their corresponding URIs. The query itself is:
 

```
SELECT DISTINCT ?class
WHERE { [] a ?class }
ORDER BY ?class
```

 At the bottom, there are three buttons: "Results:", "Browse" (with a dropdown arrow), "Go!", and "Reset".

Figura B.4: Interfaz Snorql de consulta

### B.3.5. Aplicación Web

La aplicación web consta de un menú principal que se divide en:

- Datos de Alumnos. Recoge todos los datos de la memoria académica acerca de los alumnos como por ejemplo, alumnos matriculados, de nuevo ingresos, graduados, etc.
- Datos de Empleados. Recoge todos los datos de la memoria académica sobre los empleados de las universidad tanto del personal de administración y servicios como del personal investigación.
- Datos de Financiación. Recoge no solo los datos de financiación de la Universidad, gastos e ingresos sino también la financiación que recibe cada curso académico para investigación.
- Datos de Servicios Universitarios. Recoge todos los datos acerca de los servicios universitarios que pone a disposición de sus usuarios(Estudiantes. Personal de administración y servicios y Personal docente investigador)
- Datos de Investigación. Recoge todos los datos de grupos de investigación de la UCA y proyectos desarrollados.
- Datos de Infraestructura. Recoge los datos de infraestructuras de la Universidad de Cádiz, tal y como se especifico en el apartado de la transformación de datos aquí únicamente se recogen los datos disponibles en la web [10] ya que se firmo un acuerdo de confidencialidad

donde no se pueden públicas los datos de carácter sensible que nos han sido facilitados. Estos datos se utilizarán solo en el apartado de transformación.

- Datos de Energía y Recursos materiales. Recoge todos los datos sobre el gastos de agua, papel, electricidad y la recogida de residuos.

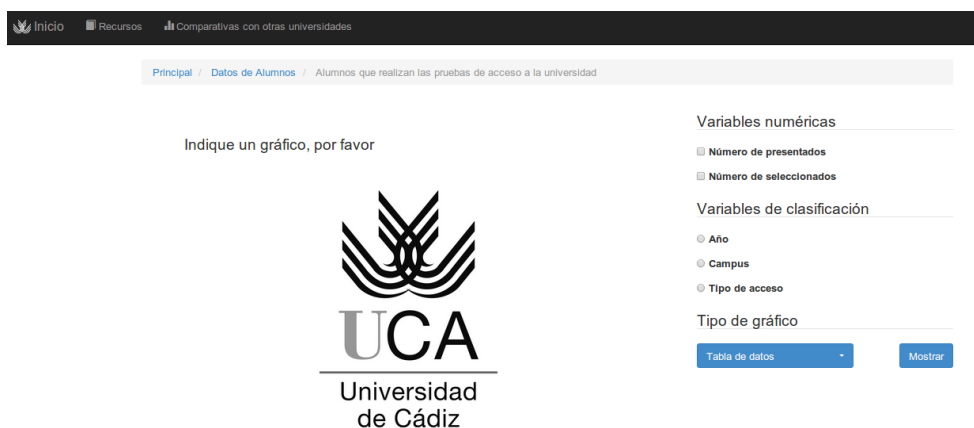


Figura B.5: Menu principal de la aplicación

El usuario deberá de seleccionar el campo del menú que le interese consultar y le redirigirá a otro menú donde podrá elegir cada una de las temáticas de los datos que desee consultar.

Una vez que seleccione el tópico de datos que desee consultar le aparecerá el siguiente formulario, donde le pedirá que seleccione según el caso una variable de tipo numérico que puede tener selección múltiple o no y una variable de clasificación que en este caso únicamente podrá elegir una. También deberá elegir en el menú desplegable el tipo de gráfico que desee crear que serán los siguientes:


- Tabla de datos. Para poder visualizar los datos en tablas.
- Gráfico de Barras. Permite comparar valores discretos (valor numérico, eje vertical) organizados por categorías (como años, eje horizontal).
- Gráfico de Líneas. Permite visualizar de datos que tienen lugar durante un período continuado de tiempo.
- Gráficos de Sectores. Permite visualizar los datos en un círculo, de modo que la frecuencia de cada valor viene dada por un trozo de área del círculo. Así, el círculo queda dividido en sectores cuya amplitud es proporcional al valor representado.



Inicio Recursos Comparativas con otras universidades

Principal / Datos de Alumnos / Alumnos que realizan las pruebas de acceso a la universidad

Indique un gráfico, por favor



Variables numéricas

- Número de presentados
- Número de seleccionados

Variables de clasificación

- Año
- Campus
- Tipo de acceso

Tipo de gráfico

Tabla de datos

Mostrar

Figura B.6: Formulario para generar la gráfica



# Apéndice C

## GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

<<http://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

### Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “**Document**”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “**you**”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “**Modified Version**” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “**Secondary Section**” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “**Invariant Sections**” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “**Cover Texts**” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “**Transparent**” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “**Opaque**”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, TeXinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “**Title Page**” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “**publisher**” means any person or entity that distributes copies of the Document to the public.

A section “**Entitled XYZ**” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “**Acknowledgements**”, “**Dedications**”, “**Endorsements**”, or “**History**”.) To “**Preserve the Title**” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.



## 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

## 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>. Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

## 11. RELICENSING

“Massive Multiauthor Collaboration Site” (or “MMC Site”) means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A “Massive Multiauthor Collaboration” (or “MMC”) contained in the site means any set of copyrightable works thus published on the MMC site.

“CC-BY-SA” means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

“Incorporate” means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is “eligible for relicensing” if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

### **ADDENDUM: How to use this License for your documents**

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with ... Texts.” line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

# Bibliografía

- [1] Documentación de CKAN.  
<http://ckan.readthedocs.org/en/ckan-1.8.2/>, 2012.
- [2] Manual de Spoon.  
<http://wiki.pentaho.com/display/EAI/.01+Introduction+to+Spoon>, 2012.
- [3] CKAN Data Management System Documentation.  
<http://docs.ckan.org/en/ckan-1.8/install-from-package.html>, 2014.
- [4] OWL Web Ontology Language .  
<http://www.w3.org/TR/owl-ref/>, 2014.
- [5] RDF Current Status.  
[http://www.w3.org/standards/techs/rdf#w3c\\_all](http://www.w3.org/standards/techs/rdf#w3c_all), 2014.
- [6] RDF Schema 1.1.  
<http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>, 2014.
- [7] Tim Berners-Lee. Cinco estrellas Open data.  
<http://5stardata.info/>, 2009.
- [8] Chris Bizer. D2RQ Platform.  
<http://d2rq.org/>, 2012.
- [9] BOE. Tablas salariales.  
<http://www.boe.es/boe/dias/2009/04/04/pdfs/BOE-A-2009-5688.pdf>, 2009.
- [10] Universidad de Cádiz. Memoria académica 2012-2013 de la Universidad de Cádiz.  
<http://memoria2013.uca.es/>, 2013.
- [11] Gobierno de Estados Unidos de América. data.gov.  
<http://www.data.gov/>, 2014. [Internet, consultado 5 de Septiembre del 2014].
- [12] Gobierno de Nueva Zelanda. data.govt.nz.  
<https://data.govt.nz/>, 2014. [Internet, consultado 5 de Septiembre del 2014].
- [13] Universidad Pablo de Olavide. Plan de medidas de impulso de la reutilización de la información.  
[http://datos.upo.gob.es/wp-content/uploads/2014/07/UPO\\_PlanRISP.pdf](http://datos.upo.gob.es/wp-content/uploads/2014/07/UPO_PlanRISP.pdf), 2014. [Internet, consultado 20 de Agosto del 2014].
- [14] Falk Kaner and Nguyen. Testing computer software (second edition), 1993.

- [15] LinkedData.org. Linked data - connect distributed data across the web.  
<http://linkeddata.org/>, 2014. [Internet, consultado 20 de Agosto del 2014].
- [16] Wikipedia. Definición de html.  
<http://es.wikipedia.org/wiki/Interoperabilidad>. [Internet, consultado 14 de Agosto del 2014].
- [17] Wikipedia. Definición de interoperabilidad.  
<http://es.wikipedia.org/wiki/Interoperabilidad>. [Internet, consultado 14 de Agosto del 2014].