

**UNIVERSIDAD CATÓLICA DE VALENCIA
SAN VICENTE MÁRTIR**



**INTEGRACIÓN DE CONTENIDOS
SEMÁNTICOS EN UN PORTAL WEB DE
CIENTÍFICOS Y HUMANISTAS
VALENCIANOS: VESTIGIUM**

TESIS DOCTORAL

Presentada por: D^a. Alicia García García

Dirigida por: Dra. D^a. Fernanda Peset
Dra. D^a. Antonia Ferrer Sapena
Ponente: Dra. D^a. Asunción Gandía

Valencia, 2014

Dra. D^a. Fernanda Peset Mancebo, profesora del Departamento de Comunicación, Documentación e Historia del Arte de la Universidad Politécnica de Valencia.

Dra. D^a. Antonia Ferrer Sapena, profesora del Departamento de Comunicación, Documentación e Historia del Arte de la Universidad Politécnica de Valencia.

CERTIFICAN:

Que la presente tesis doctoral titulada Integración de contenidos semánticos en un portal web de científicos y humanistas valencianos ha sido realizada por Alicia García García bajo nuestra dirección, en el Programa de Doctorado Investigación y Desarrollo para la obtención del título de Doctor por la Universidad Católica de Valencia San Vicente Mártir.

Para que así conste a los efectos legales oportunos, se presenta esta tesis doctoral y se extiende la presente certificación en Valencia a 23 de septiembre de 2014.


Fdo.: *Fernanda Peset*


Fdo.: *Antonia Ferrer*

AGRADECIMIENTOS

En primer lugar quiero agradecer a mis directoras, la prof. Dra. Fernanda Peset y la prof. Dra. Antonia Ferrer, su implicación, la confianza que han depositado en mí durante estos años y, especialmente el apoyo que me han dado. Sin ellas no habría sido posible este trabajo. He aprendido mucho de vosotras.

También quiero agradecer a la prof. Dra. Asunción Gandía la oportunidad de realizar este trabajo en la UCV. Y al profesor Dr. Rafael Aleixandre por estar ahí todos estos años. He aprendido mucho con este proyecto.

A Imma Subirats, Johannes Keizer y otros tantos amigos de FAO.

No puedo olvidar a otros profesores (Dr. Enrique Orduña, Daniel Vila, Dra. Yolanda Blasco, Dr. Víctor Agulló, Dr. Juan Antonio Pastor, Lourdes Castelló) que me han brindado su apoyo y su ánimo en algún momento de este largo recorrido.

Por supuesto, a mi familia, a mi madre y a mi padre por sus ánimos, por su apoyo, a mi hermana Vanesa por su alegría, y a Mateusz que siempre ha estado ahí, y a mis amigos, por sus ánimos y por seguir ahí.

A todos ellos, mi agradecimiento más sincero.

ABSTRACT

HOW TO PUBLISH SEMANTIC CONTENTS ON A VALENCIAN SCIENTIFIC AND HUMANIST WEB SITE

The purpose of this paper is to study the needs of developing a digital library containing information on Valencian Researchers and Humanists, compatible with semantic web technologies, to publish data and share information according to the recommendations of the World Wide Web Consortium. The purpose of the Vestigium website is to gather, unite and preserve information on the most important Valencian researchers and humanists that, thanks to their academic and professional career, have become icons in their areas of expertise. The website shall focus mainly on personalities that have worked during the twentieth century, as progress was made in all areas of science in this period, without forgetting the contribution of illustrious Valencian people in terms of scientific, academic and humanist development.

Objectives

The main goal of this study is to investigate the best way of applying semantic web technologies in order to develop a website that may preserve the history of science, through

biographies of Scientists and Academics. This paper also deals with secondary objectives, which are as follows:

- To describe the evolution of the dissemination of scientific and technical information.
- To appreciate the status of bibliographic description standards in libraries.
- To identify concepts and technologies arising from the emergence of the semantic web.
- To study the influence linked data are experiencing in the digital libraries context.
- To establish the type of data that can be published on Vestigium.
- To design the Vestigium web site using the analysed protocols and tools.
- To identify the process of structured content publication in Vestigium.
- To ascertain how data from external sources relates with Vestigium.

Methodology

The methodology used in this study was as follows:

1. Review of academic and professional literature obtained by searching in the following sources: E-Lis Repository:

E-prints in Library and Information Science; Lisa: Library and Information Science Abstracts; CSIC databases: ISOC Social Sciences and Humanities; Web of Science; World Wide Web Consortium, RECOLECTA (FECYT): open access literature from OA Spanish repositories; DOAJ: Directory of Open Access Journals; Dialnet. REBIUN: Spain university libraries network; BNE: Biblioteca Nacional Española; ISBN database; and TESEO and TDX databases.

2. To thoroughly review the World Wide Web Consortium (W3C) website to analyse information regarding linked data applicable to libraries:
 - Reports, guidelines and recommendations.
 - Work groups.
 - Vocabulary guidelines.
3. To analyse Linked Open Data repositories and Linked Open Data banks registries:
 - OpenDOAR, which is an international directory of open access repositories.
 - The Data Hub, which is an international registry of linked datasets.
 - Linked Open Data Project, which is a diagram to represent the linked datasets.
 - Library Linked Data Incubator Group: datasets, value vocabularies and metadata element sets.

4. Once data sources have been analysed, we used the descriptive methodology to ascertain which sources could be suitable to use in Vestigium.
5. To analyse information needs, website planning, structure design and website content; select the appropriate content management system, download and monitor the implementation of the website.

Structure

This paper broaches the next aspects:

Access to academic and scientific information

Nowadays, Internet is a global network that provides access and exchange of all types of information, be it cultural, leisure-related, academic or scientific. Most users, including researchers, academics and students, begin their information and bibliographic research using general search engines such as Google, Bing or Yahoo. However, these search engines do not retrieve information from every single existing library catalogues or scientific databases, although they do retrieve complete open access (OA) texts from OA repositories.

In order to incorporate contents from databases in the website environment, OA publishing procedures have come to mind, which has also come to revolutionise the use of scientific

information since the beginning of the XXI century. OA practices are based on the standardisation of documents description and formats. OA also combines library document description rules with web technologies. All of these practices are oriented towards facilitating access, preservation and interoperability of information between different systems (libraries, repositories, web sites).

Document description

Cultural institutions, especially libraries, are responsible for the preservation of knowledge and the dissemination thereof. Libraries, as memory institutions, have drawn up rules for the management of information resources, called Bibliographic Processing Cataloguing Rules. These rules have secured the control of heritage, access thereof and their present and future communication, as a hallmark of cultural communication. In this paper, we have reviewed the main international and national cataloguing rules and metadata schemes for the interoperability of information between systems. The authorities responsible for these regulations have adjusted them to describe the information of the new media that has emerged in the digital realm in recent decades.

A traditional bibliographic description process is nowadays adapted to new digital contents and their current requirements. Library tools, such as description systems, become metadata

schemas (DC) or metadata vocabularies (Schema.org). Terminology control tools (thesaurus, vocabularies and authority registries) become ontology languages (FOAF Languages) and value vocabularies (LEM).

Linked Data Web

The web has evolved somewhat since its inception. At the beginning, in 1945 it was only an idea related to hypertext links, which became the World Wide Web in 1990. Through numerous changes and adaptations according to user needs, it became a social web called Web 2.0 at the beginning of the XXI century. The evolution continued and we now talk about Web 3.0, or in other words, semantic web or linked data web.

This paper identifies the concept of linked data in the context of digital libraries. Linked data are the set of techniques and good practices recommended by W3C for presenting, sharing and connecting data on the semantic web using URIs and RDF language. The four principles on linked data are based on how to describe data using URIs and RDF language to identify unambiguous digital resources.

Ontologies, metadata and value vocabularies are necessary in order to publish this information as a linked data. This paper therefore identifies the concept of ontology in the semantic web context. We thus reviewed the main ontology languages,

metadata schemes and the main value vocabularies used on the web. Information related to the digital object connects creating a data network thanks to semantic technologies.

Vestigium: a digital library of Valencian Scientists and Humanists

There are numerous content management systems on the market to develop a website but we chose Drupal 7 from among others, to develop the Vestigium website. For this, we analysed the types of content we could publish on Vestigium and studied the information structure of the website. We also examined the Drupal 7 modules that are necessary to publish content as a linked data, using RDF and URI, as well as ontology languages, metadata schemes and value vocabularies that best describe the types of content to be published on Vestigium.

We considered the different institutions which present their data as linked open data and identified how to link datasets with these external resources.

Conclusions

The following conclusions are a summary of the entire study:

One. Nowadays, a digital library with open access content is not enough. The web is saturated with all kinds of content and it has become increasingly difficult to get the attention of the users. Semantic technologies are necessary to provide an informative context for the information provided. Persistent links connect related data between different sources from the semantic web to increase the information on the main resource. Libraries do not usually provide the context information from other sources. Semantic technologies therefore enable linking bibliographic collections with geographic and authority data, thus building a network to browse beyond the first source originated by the library.

Two. Content management systems make it easier to create digital information objects (for instance, documents or author biographies) and to publish it in the form of linked data. The Drupal 7 content management system is an effective tool to publish content following the four principles of good practice of linked data. This content management system has the necessary elements to describe data with RDF language. However, additional plugins (such as the RDF and SPARQL modules) are necessary to advance in the integration of other external content in one's own library.

Three. The content published on Vestigium are mainly biographical pages of Valencian Researchers and Humanists, which are linked to documentary pages (articles and thesis) prepared by the researchers and humanists themselves. The same happens with bibliometric data taken from studies that have been conducted on their publications, opinion articles and informative pages. All the above has been published as structured content, described in RDF language with persistent URIs that identify the digital object and all of their alternative language variants. The digital object characteristics are described using metadata vocabularies and ontologies compatible with the semantic web. In short, we have reached LOD principles of good practice and are able to provide informative context links that navigate towards related data from other sources.

Four. Vestigium contents are linked with external information sources through persistent links. Many of the sources that publish data as LOD are national libraries from all over the world (mainly from Anglo-Saxon countries), cultural organizations and international research centres. For this reason the context information to complete Vestigium's own content is fully verified.

Five. Vestigium reuses monitoring tools that are being used on the web, as promoted in the semantic environment. Metadata vocabularies that best describe the characteristics of Vestigium contents have been selected. Vestigium has also reused ontologies and value vocabularies (Schema.org, FOAF, DC, OWL and SIOC) to describe the content.

Six. Access, interoperability and reuse are the benefits of using semantic web technologies and the consequences are visible in the connection between the information from different professional groups. The same technologies to describe, communicate and link the information would be necessary to connect information objects from all over the world. The standardisation is the main element to guarantee that every information source describes the data in the same language. Describing the contents as semantic data enables software applications (semantic index, semantic engines, RDF query applications), which facilitates the integration of data, improves SEO positioning, visibility and access to the actual resource. Until now, semantic technologies are still consolidating, as happened with open access. The use of semantic technologies was established due to the influence of W3C, their effectiveness and the collaboration of cultural institutions. Given many cultural institutions use semantic technologies, the amount of data available on semantic webs is increasing day by day.

Final reflexion

Vestigium is part of the semantic web. The linked data network is an international cooperation project intended to establish links between information objects from different sources, to ultimately provide informational context. Vestigium links with LOD sources and provides contents about Valencia's scientific and humanist heritage.

Future work

This paper offers an approach to the current state of the publication of library information as linked data. The results of this paper could be considered a departure point for further work at much broader levels.

A deeper approach of this kind would address issues that have been left out of this particular study, such as future development and improvements in relation with the presentation of information to the end-user, the development of graphical interfaces and data visualization.

So far, the study carried out has been more focused on linking information and sharing sources. Although, at present, many institutions are a part of the semantic web network, but in the future, when information silos are connected, which up to now have been isolated, we will be heading in the right direction to attain knowledge commons.

ÍNDICE

INTRODUCCIÓN	37
Objetivos	48
Metodología	49
Estructura.....	52
DATOS BIBLIOGRÁFICOS EN EL SECTOR DE LAS BIBLIOTECAS: HACIA SU INTEROPERABILIDAD	55
Normas de descripción bibliográfica.....	59
Reglas de catalogación internacionales	59
Reglas de catalogación nacionales	80
La información científica	84
Publicación <i>Open Access</i> o <i>Gold road</i>	88
Depósito en repositorios abiertos o <i>Green road</i>	92
Interoperabilidad	111
Los esquemas de metadatos para bibliotecas	127
Machine Readable Cataloguing (MARC).....	141
Metadata Object Description Schema (MODS).....	152
Metadata Authority Description Schema (MADS).....	154
Dublin Core Metadata Element Set (Dublin Core).....	157
Metada Enconding and Transmission Standard (METS).162	
Text Enconding Intiative (TEI).....	165

LA WEB DE LOS DATOS ENLAZADOS, LA WEB	
SEMÁNTICA.....	169
Generalidades de la web semántica.....	179
Publicación y descripción en el entorno web.....	184
HyperText Markup Language (HTML)	184
eXtensible Markup Language (XML).....	188
Resource Description Framework (RDF)	192
Vocabularios y ontologías para describir datos	202
Fundamentos de las ontologías	208
Lenguajes de definición de ontologías	217
Vocabularios de valores	227
Fuentes linked data y su consulta	259
Publicación en LOD	259
Fuentes disponibles	264
Consulta de datos enlazados.....	281
LA IMPLEMENTACIÓN EN VESTIGIUM	289
Justificación	294
Metodología	301
Objetivos y público	317
Diseño e implementación.....	319
Usuarios.....	325
Estructura	327
Organización de la información	329
Áreas de contenido	334
Elementos de navegación y recuperación	339

Conjuntos de elementos de metadatos	345
Vocabularios de valores	351
Tipos de contenido	358
Módulos.....	364
Apariencia.....	382
Contenidos propios y externos.....	388
Evaluación de Vestigium	407
CONCLUSIONES	413
BIBLIOGRAFÍA	425
SIGLAS Y ACRÓNIMOS	471
GLOSARIO.....	477
ANEXOS.....	485
Anexo 1: Materias por las que clasifica WOS.....	487
Anexo 2: Modelo de documentos de acuerdo de cesión de derechos con los archivos	493
Anexo 3: Protocolo de trabajo para las fuentes orales ...	494

Índice de imágenes

Ilustración 1: Gráfico de evolución de las normas de descripción bibliográfica.....	69
Ilustración 2: Reglas de catalogación españolas, edición 2010.	83
Ilustración 3: Logotipo del Open Access	92
Ilustración 4: Interfaz de consulta del registro ROAR de repositorios OA.....	101
Ilustración 5: Gráfico del crecimiento de los repositorios españoles en la base de datos OpenDOAR.....	102
Ilustración 6: Gráfico de las tipologías de repositorios de acceso abierto en España en los años 2012, 2013 y 2014	103
Ilustración 7: Gráfico de la arquitectura de consulta de un harvester a múltiples proveedores de datos	116
Ilustración 8: Relación entre identificador, recurso y representación.....	124
Ilustración 9: Registro de la base de datos AGRIS (FAO).....	134
Ilustración 10: Visualización en RDF/XML de un registro de la base de datos AGRIS (FAO)	134

Ilustración 11: Visualización del registro en formato MARC.	145
Ilustración 12: Visualización del registro en formato de ficha para usuarios	146
Ilustración 13: Diferencias de los lenguajes de marcas HTML y XML	148
Ilustración 14: Extracto de un ejemplo de Marcado con MARC Bibliographic DTD (XML), desde la cabecera hasta el campo de control 008	150
Ilustración 15: Extracto de un registro del esquema MARC XML Schema.....	151
Ilustración 16: Ejemplo de los campos titulo de un registro MODS y equivalencia con los campos del registro MARC....	153
Ilustración 17: Ejemplo de campos de autor en formato MADS y los campos equivalentes en formato MARC 21	155
Ilustración 18: Ejemplo de un registro de autoridad para un autor personal en formato MADS.....	155
Ilustración 19: Visualización de registro en Dublin Core	160
Ilustración 20: Ejemplo de documento HTML.....	186
Ilustración 21: Visualización del documento HTML en la web	187

Ilustración 22: Ejemplo de documento escrito en XML con etiquetas de definición de la estructura pero sin etiquetas de estilo para la visualización.....	190
Ilustración 23: Arquitectura de la web semántica según el W3C de 2000	191
Ilustración 24: Ejemplo de enlaces RDF de un triplete	194
Ilustración 25: Representación gráfica del triplete RDF de la ilustración 24	195
Ilustración 26: Representación gráfica de un triplete y representación gráfica de un enlace hipertextual.....	196
Ilustración 27: Grafo compuesto por tres tripletes RDF.....	197
Ilustración 28: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo)	200
Ilustración 29: Ejemplo de axioma clase en lenguaje OWL ...	211
Ilustración 30: Ejemplo de una ontología OWL.....	215
Ilustración 31: Representación gráfica de la ontología owl de la ilustración 30	215

Ilustración 32: Gráfico jerárquico con la definición de propiedades de la clase <i>spouse</i>	218
Ilustración 33: Ejemplo de un extracto de lenguaje owl para representar una clase y una subclase de una ontología sobre vino	221
Ilustración 34: Relación jerárquica entre dos conceptos de un sistema de organización del conocimiento en SKOS	223
Ilustración 35: Representación gráfica de la etiqueta de relación jerárquica entre dos conceptos de un sistema de organización del conocimiento	224
Ilustración 36: Aplicación de los vocabularios LCSH, LEM y RAMEAU	229
Ilustración 37: Interfaz de búsqueda del registro de vocabularios Linked Open Vocabularies de la OKFN.....	233
Ilustración 38: Interfaz del proyecto HIVE	234
Ilustración 39: Ejemplo de la clase 641 con las variantes lingüísticas consultado desde la interfaz para usuarios	236
Ilustración 40: Visualización del concepto Biography de la LCSH como LD.....	238
Ilustración 41: Interfaz de consulta del tesoro AGROVOC de agricultura de FAO de UN.....	242

Ilustración 42: Interfaz de consulta de autoridades WorldCat Identities	244
Ilustración 43: Ejemplo de la ontología BIBO aplicada en la descripción de un libro	250
Ilustración 44: Ejemplo de la ontología BIBO aplicada en la descripción de un artículo de revista	250
Ilustración 45: Grafico de los conceptos clave de descripción de un dataset estadístico descrito con el vocabulario qb	252
Ilustración 46: Extracto de definición de la estructura de los datos del dataset con el vocabulario qb	253
Ilustración 47: Mapa de los elementos de la web semántica...	257
Ilustración 48: Interfaz de la aplicación LOD para la consulta de la colección SOMNI del repositorio de la Universitat de València	261
Ilustración 49: Primer diagrama de LOD Cloud a fecha de 2007	265
Ilustración 50: Diagrama del LOD Cloud en noviembre de 2008	266
Ilustración 51: Diagrama de LOD Cloud en marzo de 2009 ...	266
Ilustración 52: Diagrama de la nube de LOD en 2010	267

Ilustración 53: Diagrama de la nube de LOD en 2011	267
Ilustración 54: Diagrama de la nube de LOD en 2014.....	268
Ilustración 55: Página web de información y acceso a los LOD de DBpedia en la interfaz de consulta de Data Hub	269
Ilustración 56: Interfaz de DBpedia.....	272
Ilustración 57: Interfaz de búsqueda de SPARQL Endpoint de Europeana Linked Open Data.....	274
Ilustración 58: Interfaz de búsqueda de Bne.linkeddata.es en 2011	275
Ilustración 59: Interfaz de búsqueda de datos.bne.es en 2014.	276
Ilustración 60: Interfaz de búsqueda de la Biblioteca Digital Valenciana	278
Ilustración 61: Interfaz de acceso al spaql endpoint de Medline	279
Ilustración 62: Interfaz de búsqueda de AGRIS.....	280
Ilustración 63: Interfaz de búsqueda del índice semántico Sindice	285
Ilustración 64: Interfaz de búsqueda del buscador semántico WolframAlpha.....	286

Ilustración 65: Interfaz del buscador semántico Falcons.....	287
Ilustración 66: Arquitectura de Drupal: Interfaces y capas	323
Ilustración 67: Roles de usuarios y permisos	326
Ilustración 68: Mapa del sitio web Vestigium.....	329
Ilustración 69: Tipos de páginas de Vestigium y áreas de contenido de la estructura interna.....	330
Ilustración 70: Portada de Vestigium	331
Ilustración 71: Ejemplo de página interna de Vestigium.	332
Ilustración 72: Página libre sobre un documento de Vestigium	333
Ilustración 73: Menú de la interfaz de administrador de Drupal	334
Ilustración 74: Bloques de trabajo en el apartado “estructura” de la interfaz de administrador de Vestigium.....	335
Ilustración 75: Bloque de contenido central en la home de Vestigium.....	337
Ilustración 76: Bloques de contenido “tríptico primero, medio y último” en la home de Vestigium.....	339

Ilustración 77: Interfaz de administración de la estructura con los menús activados	341
Ilustración 78: Taxonomías de Vestigium.....	344
Ilustración 79: Ejemplo de un triplete de RDF de un autor de Vestigium.....	346
Ilustración 80: Representación gráfica del triplete RDF de la imagen anterior	346
Ilustración 81: Extracto de algunos campos de descripción que emplean los metadatos del modelo Schema.org en el tipo de contenido Autor	351
Ilustración 82: LEM en la nube LOD	353
Ilustración 83: Interfaz de búsqueda de VIAF.....	354
Ilustración 84: Interfaz de búsqueda de GeoNames	356
Ilustración 85: Ejemplo de un extracto de nombre geográfico de Valencia y algunos de sus nombres alternativos en un registro de GeoNames	356
Ilustración 86: Interfaz de búsqueda de nombres geográficos del GRI	357
Ilustración 87: Vocabulario GeoNames en el diagrama de la nube de LOD	358

Ilustración 88: Vistas que organizan el contenido en cuatro bloques en las páginas internas.....	369
Ilustración 89: Tres vistas en la home de Vestigium.....	369
Ilustración 90: Visualización de la Vista Estudios bibliométricos como página.....	370
Ilustración 91: Módulo de RDF que forma parte del núcleo original de Drupal versión 7	375
Ilustración 92: Ejemplo de enlaces RDF de un triplete de un autor de Vestigium.....	379
Ilustración 93: Representación gráfica del triplete RDF de la imagen 93	379
Ilustración 94: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo)	381
Ilustración 95: Diseño de portada	384
Ilustración 96: Diseño de página interna	385
Ilustración 97: Portada.....	386
Ilustración 98: Página interna	387
Ilustración 99: Logotipo de Vestigium.....	389

Ilustración 100: Pie de página del sitio web con el logotipo y las entidades colaboradoras.....	390
Ilustración 101: Página de un autor	393
Ilustración 102: Página de un documento.....	394
Ilustración 103: Página de un estudio bibliométrico	395
Ilustración 104: Interfaz de consulta SPARQL de DBpedia con la búsqueda de autores nacidos en Alicante	399
Ilustración 105: Resultado de búsqueda de un autor en bne.linkeddata.es	400
Ilustración 106: Gráfico con los resultado de búsqueda de un autor en bne.linkeddata.es	401
Ilustración 107: Resultado de búsqueda de un autor en datos.bne.es.....	402
Ilustración 108: Interfaz de búsqueda de Síndice con los resultados de buscar el autor “Avelino Corma”	403
Ilustración 109: URI del mismo recurso en DBpedia en diferentes idiomas.....	404
Ilustración 110: URI de información relacionada con el recurso en fuentes externas.....	405

Ilustración 111: Interfaz de búsqueda de WolframAlpha con los resultados de la búsqueda de Sorolla.....	406
Ilustración 112: Página sobre un autor en Vestigium.....	408
Ilustración 113: Resultado de Sindice sobre el análisis de la URL de un contenido de autor creado en Vestigium.....	410
Ilustración 114: Resultado de Sindice, predicados RDF empleados en la ficha de autor de Vestigium	411
Ilustración 115: Grafo generado por Sindice sobre el recurso autor Barcia Goyanes en Vestigium	412

Índice de tablas

Tabla 1: Diez entidades clasificadas en tres grupos	74
Tabla 2: Tipos de relaciones entre las entidades FRBR	75
Tabla 3: Grupos de clasificación de los atributos de las entidades FRBR.....	76
Tabla 4: Ejemplo de <i>crosswalk</i> de metadatos DC, EAD y MARC21	137
Tabla 5: Códigos de subcampo del campo 856 - Localización y acceso electrónicos	146
Tabla 6: Ejemplo de etiquetas SKOS para representar el término preferente y el término no preferente de un concepto	224
Tabla 7: Tecnologías de la web semántica	226
Tabla 8: Ejemplo de URI de valor de metadato de vocabulario de valor	230
Tabla 9: Ejemplo de relación entre un trabajo citante y uno citado descrito con la ontología CiTO estructurado en RDF...	255
Tabla 10: Ejemplo de algunas entidades y atributos del lenguaje CiTO para la caracterización de las citas bibliográficas.....	256

Tabla 11: Clasificación temática de los conjuntos de datos representados en el diagrama de la nube de LOD	270
Tabla 12: Producción científica (artículos en WOS) Comunidad Valenciana y España 1990-2009	298
Tabla 13: Ejemplo de vocabulario de metadatos DCMI Terms y DCMI Elements	348
Tabla 14: Esquema de descripción Dublin Core seleccionado	349
Tabla 15: URI	355
Tabla 16: Tipos de contenidos que se visualizan como páginas web dentro del sitio Vestigium	359
Tabla 17: Campos de información del tipo de contenido autor y elementos de metadatos de varios vocabularios	360
Tabla 18: Campos de información del tipo de contenido documentos y elementos de metadatos del vocabulario schema.org	362
Tabla 19: Campos de información del tipo de contenido página informativa y elementos de metadatos de varios vocabularios	363
Tabla 20: Campos de información del tipo de contenido artículos de opinión y elementos de metadatos de varios vocabularios	363

Tabla 21: Vistas de los tres bloques de la home (autores, publicaciones y estudios bibliométricos) con URL de direccionamiento para visualizar como página	370
Tabla 22: Vistas de los cuatro bloques de las páginas interiores	372
Tabla 23: URI del recurso autor en Vestigium y URI del mismo recurso en VIAF	397
Tabla 24: URI del recurso autor en Vestigium y URI del mismo recurso en VIAF y en BNE.....	403
Tabla 25: Campos propuestos por la Dr. Yolanda Blasco para las biografías de Vestigium.....	498

INTRODUCCIÓN

En la última década el uso de Internet ha aumentado de forma exponencial y la tendencia continúa al alza. La web es un entorno digital que se ha convertido en medio de trabajo, de ocio y de negocios, en el que las personas y las organizaciones tienen una forma de promocionarse. Es una herramienta de comunicación habitual en la vida diaria al igual como lo ha sido el teléfono, la televisión y la radio en décadas anteriores (Castells, 2003). Facilita las actividades cotidianas y es un medio de comunicación con un coste económico reducido.

Los datos del Instituto Nacional de Estadística (INE) para el período de 2008 a 2012 revelan que más de la mitad de la población española utiliza diariamente la web, y que su uso va en aumento. Esto se debe a que la web facilita el acceso a información y a servicios, y proporciona además el valor añadido de la inmediatez para consultar y para comunicar. Ello constituye un elemento fundamental en el desarrollo de las actividades y la toma de decisiones en un mundo tan cambiante como el actual.

En la web se encuentra todo tipo de información, desde la científico-técnica y académica hasta los contenidos que crean los propios usuarios sobre aficiones, actividades lúdicas y todo tipo de intereses. Pocos años atrás, para tener acceso a esta información científica, básicamente bibliográfica, era necesario desplazarse a la biblioteca y consultar el catálogo para obtener la

obra físicamente. Ahora, Internet permite la consulta de los catálogos de las bibliotecas, las bases de datos y la versión digital de algunos de sus contenidos. La web ha ido evolucionando, ha mejorado sus características, ha ampliado sus servicios y las herramientas que ofrece se adaptan cada vez más a las necesidades de los usuarios. Hasta alcanzar su desarrollo actual ha sufrido una serie de cambios a lo largo de todo el siglo XX.

En la última década, nuevas aplicaciones y nuevos software permiten a los usuarios crear sus propias páginas web y crear contenidos de forma sencilla. El crecimiento exponencial de contenidos digitales y la masificación de información existente en el entorno web ha derivado en una falta de exhaustividad en la recuperación de información relevante por los motores de búsqueda, y en problemas en conocer la veracidad de las fuentes y la calidad de la información. Algunos de estos problemas son consecuencia de que el contenedor informativo que constituye Internet no está gestionado por normas de catalogación y clasificación como lo está la información recogida en las bibliotecas.

Para solventar los problemas de acceso a la información que se producen en la red, se plantea la necesidad de crear otro tipo de definiciones que identifiquen los recursos electrónicos de una forma más precisa y que permita la recuperación de información

de forma más eficaz. La utilización de metadatos favorece este intento. En los últimos años las tecnologías de la web semántica y los principios de los datos enlazados (*Linked Data*, en adelante LD) surgen como solución para la gestión eficaz de la información en el ámbito digital. Como ocurrió con el protocolo de transferencia de hipertexto HTTP, ha sido Tim Berners Lee quien en 2009 presentó en TED (Technology Entertainment and Design) la idea de web semántica, basada en datos enlazados (Berners-Lee, 2009).

Esta nueva concepción de web está dirigida a que las distintas aplicaciones de la red puedan comprender a nivel semántico el contenido de la información que están mostrando al usuario. Este cambio supone una mejora en la precisión y la relevancia de los resultados de las búsquedas de información que generan los motores de búsqueda actuales. En la web actual los motores de búsqueda recuperan contenido basándose en palabras en lenguaje natural a las que se asignan diferentes ponderaciones, lo que produce errores de recuperación por sinónimos y homónimos, o por el exceso de ruido y silencio, nociones habituales en el ámbito de las ciencias de la documentación. Estos problemas de recuperación se corrigen en la web semántica cuando se introducen descripciones explícitas de la información mediante el uso de metadatos y ontologías, pues con estas descripciones el software es capaz de entender el significado de la información e identificar conceptos y sus

relaciones (Castells, 2003).

La adaptación de los contenidos hacia una definición estructurada de los mismos a través de la publicación como datos enlazados está cobrando fuerza en distintos sectores, especialmente en instituciones culturales y bibliotecas. Por esta razón nos planteamos los beneficios que puede aportar la aplicación de estas tecnologías en un sitio web. Estudiamos especialmente los procedimientos y los movimientos relacionados con el acceso y la visibilidad de los contenidos que serán alojados en un sitio web de estas características y la relación entre los tipos de datos que alberga.

Por tanto, la investigación que se presenta es deudora de varias disciplinas científico-técnicas. El corazón del trabajo se fundamenta en aspectos que han sido tradicionalmente acometidos por la biblioteconomía y la documentación, como son la descripción y difusión de la información o el conocimiento y uso de fuentes de información. Pero los aspectos que ahora se han desarrollado con la aparición de los sistemas automatizados y especialmente la web ha cambiado radicalmente algunos aspectos de la difusión y de la recuperación de la información, lo que afecta a la visibilidad de los contenidos. Por esta razón, este trabajo presenta los estándares que se están utilizando para la descripción y la difusión de la información, por ejemplo, los esquemas de

descripción de metadatos o las ontologías. Sin embargo, el nivel de desarrollo tecnológico no es abordado en profundidad porque hace parte de las disciplinas relacionadas con la informática y las telecomunicaciones. Así también, esta investigación aborda aspectos relacionados con la historia de la ciencia y de los profesores, puesto que será el contenido a distribuir. Sin embargo, esos aspectos no se tratan desde el punto de vista de la generación de nuevos contenidos, de nueva investigación. Esta tesis filtra este tipo de contenidos para construir el producto de información original, una biblioteca digital que reúna las características que desde el punto de vista de la biblioteconomía y documentación más punteras se consideran necesarias para que sus contenidos se puedan convertir en un punto de referencia para otros productos. Y ello entronca, por último, con los aspectos de difusión de la información científico-técnica más actuales, los movimientos de apertura de información y de promoción de la reutilización. Estos puntos son abordados por muy diversas disciplinas, puesto que son la confluencia de estudios clásicos sobre comunicación científica (biblioteconomía y documentación), desarrollos tecnológicos (informática y telecomunicaciones), estudios sobre la ciencia (sociología, psicología aplicada e historia de las universidades), aspectos de política científica (administración y política) e incluso en ocasiones, para la generación de las ontologías

participan lingüistas e investigadores de las disciplinas para las que se producen.

Por estas razones, la tesis que se presenta tiene como característica esencial la hibridación con otros perfiles, si bien la profundidad en la investigación solo se refiere a su núcleo, que parte de la biblioteconomía y documentación para pasar al sector de las tecnologías de la información y las telecomunicaciones (TIC en adelante). Esta evolución o fusión entre ambas disciplinas todavía no se ha producido en España, pero se percibe a nivel internacional cómo es imposible mantener ambos campos de conocimiento aislados unos de otros pues sus preocupaciones sobre descripción, recuperación y reutilización de la información son actualmente indisolubles. Por esta razón, el campo TIC será mencionado en este trabajo como una categoría que abarque ambas disciplinas. De esta manera, y como tal producto híbrido y colaborativo, su aportación al conocimiento reside en investigar temas limítrofes al campo documental para aplicar de todos ellos lo que resulta beneficioso al diseño del producto de información original, que en definitiva es el resultado de esta tesis. No se ha tratado de generar códigos para desarrollo tecnológicos, ni nuevos estándares sino en investigar y documentar, tal y como hacen las ciencias del diseño, las mejores opciones que permitan producir un recurso de información original en su sector.

Por tanto, el trabajo se enmarca en el estudio de la aplicación de la tecnología semántica para la descripción de contenidos digitales, su recuperación y su reutilización. Pretende revisar estas tecnologías y estudiar los beneficios que puede aportar la utilización de las normas de la web semántica como elemento de cambio en los servicios de publicación y difusión de la información del sitio web *Vestigium: Biblioteca digital de Científicos y Humanistas Valencianos* a toda la comunidad de usuarios de Internet. A lo largo del trabajo se recogen políticas de acceso a la información científica, e informes que proponen lenguajes de marcado de texto, lenguajes formales de ontologías, lenguajes de consulta, tecnologías para la implementación de un sitio web y tecnologías para la interoperabilidad y la reutilización de los contenidos.

El desarrollo del portal *Vestigium* trata de recuperar el patrimonio científico y humanista valenciano reciente, haciéndolo accesible en Internet mediante protocolos que aseguren su preservación, visibilidad y reutilización de los contenidos. Esta tesis se desarrolla en el marco de trabajo de un grupo de investigación interdisciplinar dedicado al estudio de científicos y humanistas valencianos. El patrimonio es un tesoro de valor incalculable para cualquier país. La aportación de las personalidades definidas como valencianas (nacidos en la comunidad o que hayan ejercido en ella su profesión) es parte de la cultura identitaria de la zona y por extensión de la comunidad

científica. La elección geográfica de estos estudios viene avalada por ser esta comunidad la cuarta en producción científica en nuestro país, por lo que la presencia en Internet de sus investigadores y humanistas puede ser relevante para la preservación patrimonial y científica.

En suma, el sitio web se enfoca como un repositorio que gestiona, aglutina y disemina los productos resultantes de la actividad investigadora sobre académicos valencianos del siglo pasado, cuya actividad científica, académica o humanística tiene repercusión social o cultural, o supone una aportación a la historia de la ciencia en su ámbito de conocimiento. Este recurso de información se plantea con el objetivo de proporcionar acceso a información biográfica, bibliográfica y documental (documentos, videos, material sonoro y audiovisual). Definir los datos bibliográficos y de autoridades como datos enlazados mejoraría los problemas identificados en la búsqueda de información en la actual web. Es en la actual web donde los datos de bibliotecas, tanto bibliográficos (geográficos, fechas) como los datos de autoridades (instituciones, personas), se almacenan en bases de datos que, aunque disponen de interfaces de búsqueda en la web, no están muy integrados con otras fuentes de información, incluso a veces son silos invisibles a los buscadores. La descripción de los datos de biblioteca como datos enlazados favorece el intercambio de estos datos con otros que se encuentran alojados en diferentes depósitos, lo que

permite ofrecer un valor añadido a los contenidos propios. En este sentido, nos planteamos la importancia de que Vestigium publique los contenidos como datos enlazados para formar parte de la web semántica y pueda enriquecer los propios con otras fuentes.

La implementación de Vestigium siguiendo los principios LD es una apuesta dirigida a enriquecer los datos que albergue gracias a la información de contexto aportada por otras fuentes con las que se enlace, evitando silos de información en páginas web diseminadas por Internet sin conexión entre ellas. Así también, Vestigium especialmente mejora la visibilidad de esta información en Internet ya que enlaza con otras fuentes. Para Vestigium como recurso de información digital es una oportunidad para formar parte de esta nueva tecnología a la que están adaptando sus contenidos un número de instituciones culturales, patrimoniales y de investigación cada vez mayor.

Objetivos

El objetivo general del trabajo es investigar la mejor forma de aplicar las tecnologías de la web semántica al campo de la historia de la ciencia para desarrollar contenidos sobre personas y sus trayectorias, tomando como caso de estudio Vestigium. El trabajo se propone alcanzar los siguientes objetivos secundarios:

- Detallar la evolución que ha sufrido la diseminación de la información científico-técnica.
- Conocer el estado de los estándares de la descripción bibliográfica en las bibliotecas.
- Acuñar los conceptos y las tecnologías que surgen a partir de la eclosión de la web semántica.
- Estudiar la influencia que los datos enlazados están asumiendo en el ámbito bibliotecario.
- Fijar los tipos de datos interoperables que puedan ser mostrados en Vestigium.
- Diseñar el sitio web Vestigium aplicando los instrumentos y protocolos estudiados.
- Identificar el proceso de publicación de contenidos estructurados en Vestigium.

- Conocer cómo se consumen datos de otras fuentes externas con Vestigium.

Metodología

La metodología utilizada para realizar este trabajo es la siguiente:

1. Revisión bibliográfica de la literatura académica y profesional obtenida de las búsquedas en las siguientes fuentes para artículos de revista, monografías y tesis:
 - El repositorio especializado en documentación, E-Lis (*Eprints inLibrary and Information Science*).
 - LISA (*Library and Information Science Abstracts*).
 - Sistema de información de las Bases de Datos CSIC: ISOC de Ciencias Sociales y Humanidades.
 - *Web Of Science*.
 - Plataforma Recolecta de la Fundación Española para la Ciencia y la Tecnología (FECYT), que permite el acceso libre a toda la producción

científica depositada en abierto en los repositorios científicos españoles.

- Directorio DOAJ (*Directory of Open Access Journals*) que recientemente permite la consulta de artículos.
- Dialnet.
- Catálogo de la Biblioteca Nacional de España y la Red de bibliotecas universitarias REBIUN.
- Base de datos de libros editados en España ISBN, del Ministerio de Cultura.
- Bases de datos de tesis doctorales TESEO y TDX.

2. Revisión exhaustiva del sitio web del World Wide Web Consortium (W3C) para analizar la información relacionada con los datos enlazados aplicables a bibliotecas:

- Informes, pautas y recomendaciones.
- Grupos de trabajo.
- Guías sobre vocabulario de descripción de contenidos y de consulta.

3. Estudio de inventarios de repositorios y bancos de datos que puedan almacenar fuentes semánticas en formato LOD para ser consumidas por Vestigium:
 - OpenDOAR. Es un registro internacional de bancos de datos y repositorios de acceso abierto. Permite buscar depósitos de datos enlazados.
 - The Data Hub. Es un registro de datos que contiene el mayor número de conjuntos de datos descritos en formato de datos enlazados.
 - Linked Open Data Project. Representa gráficamente mediante un diagrama cómo están conectadas las fuentes y los paquetes de datos enlazados.
 - Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets <<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>>.

4. Una vez identificadas las fuentes de datos, se utiliza el método descriptivo para conocer en detalle las que pueden ser adecuadas para nuestra investigación.

5. Estudio de las necesidades de información, planificación del sitio web, diseño de la estructura y de los contenidos, selección del gestor de contenidos, descarga y

seguimiento de la implementación informática del sitio web.

Estructura

Este trabajo se estructura en tres capítulos principales, junto con un preliminar y las conclusiones finales. En la presente introducción, preliminar al cuerpo de la investigación, se recogen las motivaciones para realizar este trabajo, los objetivos planteados y la metodología seguida.

En el primer capítulo se plantea de forma general los antecedentes de la web y la situación actual, el acceso a la información científica en Internet, revisando las principales formas de publicación en acceso abierto y el depósito en repositorios dado que el sitio web que ha de diseñarse va a distribuir este tipo de información. La publicación en acceso abierto es una de las formas de interoperabilidad que más extensión ha adquirido en el sector de las bibliotecas, pero no es la única. Por ello, en este capítulo se examina el proceso de interoperabilidad en términos amplios y el papel de los metadatos. Por último, en este capítulo se exponen las principales herramientas de descripción bibliográfica que nos permitirán conocer las bases del trabajo con documentos y apreciar la evolución que han sufrido en los últimos años.

En el segundo capítulo se aborda el concepto de datos enlazados y de las tecnologías que se emplean en la web semántica, en dirección hacia la mejora del tratamiento de la información de Internet. La modelización de los contenidos divididos en pequeñas piezas de información supone una revolución que tan solo es posible en un entorno web, en el que las máquinas son capaces de interpretar los datos individuales y presentar la información que reside en otros sitios. Así como en el capítulo anterior veíamos la difusión de información científica en registros que contenían varios metadatos sobre un ítem, la aplicación de LD atomiza estas unidades para dotarlas de un nuevo sentido aplicable a muchos otros contextos. Así pues, la publicación de información estructurada de este nuevo modo, siempre normalizado, se dirige a solventar los problemas de recuperación exhaustiva apoyada por un contexto informativo procedente de diversas fuentes contrastadas y de calidad. En este sentido, la información que reside en bibliotecas cuenta con una validación interna, la certificación de que es información cierta. Por esta razón, los productos de información procedentes de cualquier institución de la memoria son valiosos en el mundo web, donde cualquier persona puede estar emitiendo información no contrastada.

El tercer capítulo pone en práctica las cuestiones examinadas en los capítulos anteriores. Se expone el proceso de planificación y desarrollo del sitio web Vestigium. Para ello se recopilan los

contenidos que van a servir de base de testeo, se diseña la estructura contando con los aspectos habituales en la construcción de sitios web con CMS, como pueden ser los de usabilidad y como resultado de la aportación de esta tesis se implementan los principios de buenas prácticas de datos enlazados.

En el capítulo cuarto se recogen las conclusiones obtenidas tras el análisis de la situación actual, el estudio de los conceptos abordados en el desarrollo del proyecto y la implementación del sitio web.

Por último, un anexo recoge la bibliografía utilizada para realizar esta tesis utilizando el formato Norma UNE-ISO 690:2013 para elaborar las referencias.

**DATOS BIBLIOGRÁFICOS EN EL
SECTOR DE LAS BIBLIOTECAS: HACIA
SU INTEROPERABILIDAD**

“El carácter cíclico de la información, es decir, el hecho de que ésta se base siempre en otra información previa, no hace sino poner de manifiesto la importancia y necesidad de su transferencia, condición indispensable para el progreso social, y por ende científico” (Pinto Molina, 1989). La Biblioteconomía y la Documentación como disciplinas se basan en esta realidad y para alcanzarla se han encargado a lo largo de su historia de desarrollar normas y estándares para normalizar la información de manera que permita garantizar el acceso a una masa de documentos creciente y cada vez mayor, con el objetivo de asegurar el ciclo de transferencia informativa.

El tratamiento controlado de la información se ha encargado de la descripción de documentos. En su forma tradicional se define como catalogación y en la descripción de documentos electrónicos se acuña el término asignación de metadatos (García-Marco, 2013). La catalogación es parte del control bibliográfico, que adopta una serie de reglas que son necesarias para procesar la información de forma que permita identificar adecuadamente el contenido de cualquier soporte informativo y facilitar así su recuperación. La asignación de metadatos responde al mismo objetivo en un entorno en el que los recursos de información son digitales. La asignación de metadatos es el término utilizado para catalogar recursos web (Vaishnav y Sonwane, 2007).

Actualmente se está produciendo una convergencia entre las formas tradicionales de procesamiento de la información y de descripción de recursos, con la gestión de contenidos digitales de la web (Saorín-Pérez y Pastor-Sánchez, 2012). Esta situación viene causada porque en los últimos diez años, el número de documentos publicados en formato electrónico originalmente ha crecido de forma exponencial y necesitan algún tipo de control. Los profesionales que habitualmente han tratado documentos para su preservación, recuperación y difusión han sido los gestores de información de las instituciones de la memoria: bibliotecas, archivos, museos. Por esta razón, en los países anglosajones en los que esta profesión es más sólida, se ha integrado en los procesos de organización del conocimiento que reside en la web y en el establecimiento de normas que van más allá de la comunidad bibliotecaria.

Normas de descripción bibliográfica

El control bibliográfico llevado a cabo por las bibliotecas nacionales se sostiene sobre una serie de principios y normas que se centraban principalmente en la catalogación de los materiales impresos y en las versiones comerciales, y que, con el avance de las nuevas tecnologías, se han ampliado para contemplar los nuevos soportes. Los inicios de la aplicación de estas normas pueden verse en diversas corrientes. La evolución de sus códigos junto con la reimpresión de sus versiones refleja el esfuerzo que ha supuesto a lo largo de la historia acordar unas normas de nivel internacional. La International Federation of Library Associations and Institutions (IFLA), la Library of Congress (LC), la American Library Association (ALA), la Library Association (LA) y la Online Computer Library Center (OCLC), asumen funciones que han sentado las bases de la estandarización de la información que distribuye el sector bibliotecario.

Reglas de catalogación internacionales

Los primeros códigos catalográficos surgieron a raíz de que los profesionales que trabajan en las grandes bibliotecas dieran solución a los problemas que se planteaban al confeccionar los catálogos. Los códigos más significativos tienen su origen en las

Rules for the compilation of the catalogue redactadas por Anthony Panizzi en 1841 para elaborar el catálogo del British Museum. Conocidas como *Las 91 reglas de catalogación de Panizzi*, tuvieron gran influencia en la elaboración de códigos posteriores, y se han aplicado a otras bibliotecas nacionales y universitarias como la de la Universidad de Cambridge (Garrido, 2000).

En 1852 Charles Coffin Jewett, bibliotecario de la Smithsonian Institution (Washington DC, EEUU), publicó *On the construction of catalogues of libraries, and of a general catalogue; and their publication by means of separate, stereotyped titles*, un claro antecedente de la ISBD según Estivill (2012). Son el primer código de catalogación americano en el que se recomienda la catalogación cooperativa centralizada, objetivo que veinte años más tarde perseguiría la IFLA a través del programa de Control Bibliográfico Universal (Garrido, 1996).

En 1875 Charles A. Cutter publicó las *Rules for a printed dictionary catalogue*, que poco después se adaptaron al catálogo de fichas *Rules for a dictionary catalog* (1889, 1891, 1904) (Estivill, 2012). Su característica particular era que permitía ya redactar un registro bibliográfico por tres entradas, autor, título y materia, lo que estaba dirigido al ordenamiento científico de las

fuentes para la consulta de investigadores y académicos (Garrido, 1996).

En 1876 se fundó la American Library Association en una convención de bibliotecarios en la Historical Society of Pennsylvania, convocada por bibliotecarios representantes de algunas bibliotecas americanas, como Melvill Dewey y C.A. Cutter entre otros. Constituida como organización profesional, su propósito era promover los intereses de las bibliotecas del país en todo el mundo intercambiando puntos de vista y fomentando la cooperación. Desde entonces ha trabajado sobre muchos de los aspectos a los que se enfrentan las bibliotecas, los bibliotecarios y los usuarios, para darles respuesta. Es la promotora de recomendaciones sobre buenas prácticas bibliotecarias (ALA, 2014). Siguiendo sus pasos, un año después, se fundó la Library Association (LA) de UK en Londres. Esta asociación es el órgano que ha representado a los trabajadores del sector en el territorio inglés, trabajando en los diferentes frentes que afronta la profesión y sus profesionales. Se fusionó con el Institute of Information Scientist en 2002 cambiando de nombre a Chartered Institute of Library and Information Professionals (CILIP) (Munford, 1977).

Fuera del ámbito angloamericano empiezan a surgir códigos nacionales de catalogación, como las *Instruktionen für die*

alphabetischen Kataloge der preussischen Bibliotheken de 1899, conocidas como las instrucciones prusianas, por obra de Otto von Bismarck, que se aplicaron en las bibliotecas universitarias de Prusia, en las alemanas y austriacas (Garrido, 2000). Dejaron una fuerte influencia en varios códigos europeos, como el primer código español, las *Instrucciones para la redacción de los catálogos en las Bibliotecas Públicas del Estado* de 1902, como se verá más adelante.

En 1901 las fichas catalográficas de la Library of Congress de EE.UU. (LC) tuvieron una gran difusión por las bibliotecas de la nación en primer lugar y después a nivel internacional. La LC fue creada en 1800 como biblioteca parlamentaria, posteriormente asumió el papel de biblioteca nacional y en 1897 abrió sus puertas al público. En 1901 se extendió el acceso a investigadores y académicos, se inauguró un servicio de préstamo interbibliotecario, y comenzó la venta y distribución de sus registros bibliográficos a bibliotecas académicas, universitarias y de investigación. La difusión de los esquemas de catalogación y clasificación ayudó a estandarizar los procesos en el resto de bibliotecas de la nación y a fortalecer unos esfuerzos de cooperación entre las mismas. Esta temprana disponibilidad de las fichas de catalogación impulsó a la biblioteca a una posición de liderazgo (LC, 2006). Desde entonces fue adquiriendo relevancia en el ámbito internacional y desde los años sesenta y setenta del siglo XX, ha facilitado el intercambio

de la información y la racionalización de los recursos a través de proyectos de cooperación y trabajo colectivo (Picco y Ortiz, 2012). Hoy en día es uno de los referentes institucionales en el ámbito bibliotecario internacional, como iremos viendo a lo largo de este trabajo.

En 1904 las asociaciones americana y británica de bibliotecarios acordaron la creación de un código conjunto para uniformizar la práctica catalogadora. En 1908 se publicó el trabajo conjunto, el código angloamericano por la ALA y la LA que mejora los códigos precedentes. Está claramente influenciado por Panizzi, las instrucciones prusianas y la perspectiva de Cutter, que trabajó durante varios años en la ALA. Se publicó en dos versiones diferenciadas: *Catalog rules: author and title entries*, para la comunidad de los Estados Unidos, y *Cataloguing rules: author and title entries* para la británica. Entre 1936 y 1939 estas asociaciones preparaban un segundo código conjunto, pero la Segunda Guerra mundial frenó la colaboración y hasta 1967 no se reanudó la cooperación entre estos dos países en materia de catalogación (Estivill, 2012).

En 1927 nació la IFLA. Hoy en día es el principal organismo internacional que constituye una unión mundial de asociaciones nacionales de bibliotecas. Representa a nivel mundial los intereses de los profesionales de las bibliotecas y la documentación, de los servicios y de los usuarios de los

servicios bibliotecarios y de documentación. Es una organización internacional independiente, no gubernamental y sin ánimo de lucro que se fundó en Edimburgo (Escocia) en un congreso internacional. El nacimiento de la IFLA surgió del esfuerzo para organizar una cooperación internacional en el campo bibliotecario. Las primeras iniciativas de cooperación internacional fueron en 1876 y 1877 cuando se fundaron la ALA y la LA. Ambas dirigidas a la organización de conferencias que reunieron a los profesionales del ámbito de las bibliotecas. Pero no fue hasta 1926 en el Congreso Internacional de Bibliotecarios de Praga donde se propuso la creación de un comité permanente internacional de bibliotecas (International Library and Bibliographical Committee). Lo que desembocó en la creación de la IFLA un año después. Aunque no fue hasta 1929 en el Congreso Internacional de Bibliotecarios de Roma cuando el International Library and Bibliographical Committee recibió su actual nombre: International Federation of Library Associations and Institutions (IFLA) (Lacasa y Martínez, 1993).

En 1931 se publicó el Código Vaticano *Norme per il catalogo degli stampati* para la reorganización de la Biblioteca Vaticana. Este código se redactó junto con representantes de tres bibliotecas estadounidenses, entre ellas la Library of Congress, por lo que supone una obra que sintetiza las corrientes europea y norteamericana de catalogación (Garrido, 2000). Tuvieron

impacto en otros códigos nacionales, en España sobre la versión nacional de 1941 (Estivill, 2012).

En 1941 se publicó la primera edición de las reglas ALA *A.L.A. catalog rules: author and title entries*, que recibió numerosas críticas. Tras una minuciosa revisión, en 1949 se publicó una nueva edición, junto a las reglas de la LC que complementaban a las de la ALA. Tras las opiniones recibidas sobre estos últimos códigos, Seymour Lubetzky trabajó en una nueva edición por petición de la ALA y de la LA, que se publicó en 1953 *Cataloguing Rules and Principles; a Critique of the A.L.A. Rules for Entry and Proposed Design for their Revision* y en 1960 el *Code of Cataloguing Rules; Author and Title* (Garrido, 1996).

En 1961 se celebró en París la International Conference of Cataloguing Principles (ICCP) organizada por el Comité de Catalogación de la IFLA, donde se analizaron estos documentos y surgió la Declaración de Principios Internacionales de Catalogación conocida como los Principios de París, que han tenido repercusión a nivel mundial (IFLA, 1963). Suponen un avance en la tarea unificadora, pues han servido como base para la normalización internacional en la catalogación, ya que la mayoría de los códigos de catalogación que se desarrollaron en el mundo desde entonces los siguen, sino estrictamente, al menos en un alto grado (IFLA, 2009). De esta reunión surgió la

necesidad de publicar un código internacional aceptado por todos los países del mundo, para facilitar el intercambio de información bibliográfica. La primera edición de este texto fue en 1967: las *Anglo-American Cataloguing Rules* (AACR). Aunque se retoma la colaboración de 1909, este texto se publicó en dos versiones para las respectivas comunidades (Estivill, 2012).

La LC, que venía desarrollando una importante iniciativa en materia de desarrollo de estándares de catalogación, lo hacía también en los metadatos (López Guillamon, 2004). Tan solo unos meses antes, en noviembre de 1966 la LC puso en marcha el MARC Pilot Project para la catalogación automatizada de los catálogos colectivos. En 1968 el formato MARC (Machine Readable Cataloguing) se hacía operativo y comenzó la distribución de los registros bibliográficos a otras bibliotecas. En 1971 se aprobó como norma nacional ANSI Z39.2 y en 1973 se convirtió en un estándar internacional ISO 2709. Entre 1970 y 1977 aparecieron numerosas versiones de carácter nacional en múltiples países hasta la publicación del UNIMARC internacional, como veremos en el apartado correspondiente de metadatos para bibliotecas (Estivill, 2012).

La Online Computer Library Center (OCLC) contribuyó a la expansión del formato MARC desde sus inicios adoptándolo y mostrando así las ventajas de la normalización para compartir e

intercambiar registros bibliográficos (Estivill, 2012). La OCLC es uno de los grandes servicios bibliográficos desde 1967, cuando se fundó como una entidad de investigación sin ánimo de lucro, formada por bibliotecas que comparten sus recursos. En poco tiempo se convirtió en una red nacional y hoy en día es una cooperativa internacional de bibliotecas de la que forman parte más de 15.000 bibliotecas de 170 países. Su objetivo se dirige a conectar las bibliotecas institucionales a través de la World Wide Web (WWW) para la cooperación, el intercambio de recursos y el acceso universal a la información bibliotecaria, para reducir costes y mejorar los servicios a través de la catalogación compartida en línea. Uno de los mayores servicios que ofrecen es el catálogo WorldCat, iniciado en 1971 como el catálogo colectivo más numeroso, cuando se agregaron los primeros registros bibliográficos procedentes de la Alden Library de la Ohio University (OCLC, 2014).

En 1969 en Copenhague, en el International Meeting of Cataloguing Experts (IMCE) organizado por la IFLA, se puso de manifiesto la viabilidad de normalizar la forma y el contenido de la descripción bibliográfica, y se designó un grupo de expertos para trabajar en la primera versión del texto ISBD (M): *International standard bibliographic description for monographic publications*, que se dio a conocer en 1971. En 1973 fue adoptado por varias bibliotecas nacionales y se publicó

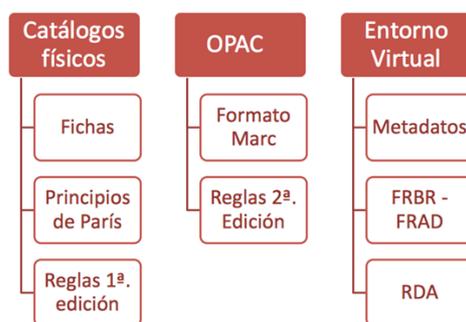
en 1974 como primera edición normalizada. La ISBD (S) *Serials* para publicaciones seriadas, también se publicó ese mismo año (IFLA, 2011). En 1977 salió la ISBD (G), *General International standard bibliographic description for monographic publications*, el esquema principal que debían seguir las versiones especializadas: ISBD (NBM) *non-book materials* y ISBD (CM) *Maps* de 1977; ISBD (M) *Monographs* de 1978; ISBD (A) para publicaciones monográficas antiguas y ISBD (PM) *Music* de 1980 (Estivill, 2012). La norma ISBD especifica los requisitos para la descripción e identificación de los tipos más comunes de recursos publicados que pueden constituir las colecciones de las bibliotecas. Así mismo, asigna un orden a los elementos de la descripción y especifica un sistema de puntuación para la misma (IFLA, 2011; López Guillamón, 2004).

Desde la automatización de los catálogos de bibliotecas en los años sesenta del siglo XX, la tendencia de las normas de catalogación se ha dirigido a contemplar los nuevos soportes digitales. En 1978 se publicaron las AACR2: *Anglo American Cataloguing Rules* de la ALA, que revisaban el anterior texto AACR de 1967 para unificar las versiones del texto británico y del texto americano, para adaptarlo a la nueva normativa de descripción ISBD, así como contemplar la descripción de ficheros de datos legibles por ordenador. Las posteriores

ediciones son de 1988 y 1998, y la de 2002 incluye un capítulo para un nuevo tipo de fuente denominada recursos continuados. La última versión es de 2005, que será sustituida por la nueva normativa RDA: *Resource Description and Access* (Estivill, 2012).

La siguiente ilustración representa gráficamente tres bloques clave en la evolución de los catálogos de bibliotecas, junto con las reglas catalográficas que se han aplicado en cada uno de estos momentos. En el primer bloque se reflejan los catálogos físicos de las bibliotecas, en los que se empleaban las fichas, los Principios de París y posteriormente las primeras reglas de catalogación. En el segundo bloque, en el que la automatización transforma los catálogos en OPAC, se emplea el formato MARC y la segunda edición de las RC. Y por último, en el entorno virtual, los elementos de descripción son los metadatos, la familia FRBR y las RDA.

Ilustración 1: Gráfico de evolución de las normas de descripción bibliográfica



Fuente: Universidad de los Andes. Biblioteca José Enrique Díez. *FRBR y RDA*. <<http://biblioestandares.bn.cl/sites/biblioestandares.bn.cl/files/FRBR%20y%20RDA.pdf>>

En la década de los años ochenta del siglo XX, ya se encontraban consolidados los sistemas de automatización de catálogos de bibliotecas, la catalogación cooperativa e Internet, y continuaban surgiendo nuevos formatos digitales. Así en 1987 la LC publicó la primera edición del documento *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media* con el propósito de ayudar a las bibliotecas y organizaciones que creaban o adquirirían registros MARC; y publicó ediciones subsecuentes en 1990, 1994 y 2000. Este documento ofrece información técnica sobre la estructura de los registros MARC, los conjuntos de caracteres utilizados en ellos, y el formato usado en las cintas magnéticas para distribuir los registros MARC (Furrie, 2003).

Durante los años noventa la IFLA impulsó el trabajo en dos líneas, por un lado, las ISBD incluyeron los nuevos formatos denominados archivos de ordenador en la nueva versión ISBD (CF) *International standard bibliographic description for computer files*. En 1997 los tipos de contenidos digitales se habían ampliado y la nueva edición ISBD (ER) *Electronic Resources* englobaba todos los nuevos soportes. Las ISBD contemplan monografías electrónicas, bases de datos, programas de computación, publicaciones seriadas, artículos y otras contribuciones, boletines electrónicos, listas de discusión y mensajes de correo electrónico, diferenciando entre recursos

electrónicos de acceso local (que contienen un soporte físico) y los recursos de acceso remoto (disponibles mediante el acceso a través de redes de telecomunicación o de sistemas informáticos) (IFLA, 2011).

Los objetivos de la ISBD de 1997 para recursos electrónicos estaban orientados a la conversión de registros bibliográficos a la forma electrónica, a hacer intercambiables los registros entre diferentes organizaciones a nivel internacional para que pudieran ser fácilmente aceptados en catálogos de biblioteca u otras listas bibliográficas en cualquier otro país; y para que los registros producidos por los usuarios de una lengua pudieran ser interpretados por los usuarios de otras lenguas. La edición publicada en 2002 fue ISBD (CR): *International standard bibliographic description for serials and other continuing resources*, para publicaciones seriadas y recursos continuados, que tienen en cuenta los nuevos desarrollos en las publicaciones seriadas, publicaciones electrónicas y en recursos web como bases de datos y sitios web que se actualizan de forma continua. Hasta el año 2007 existían siete ISBD especializadas para cada tipo de recurso, más una general ISBD (G). Ese año la IFLA publicó la edición consolidada preliminar, que unificaba en un solo texto las siete ISBD especializadas y en 2011 publicó la edición de la ISBD consolidada. La ISBD consolidada proporciona una mayor armonización entre los diferentes tipos

de recursos y una mayor adaptación a los requisitos de información establecidos en FRBR (IFLA, 2011).

Por otro lado, la IFLA encargó un estudio que definiera los requisitos funcionales de los registros bibliográficos en relación a las necesidades de los usuarios y a los diversos soportes, en la IFLA Council and General Conference de 1990 en Estocolmo. Para lo que fue nombrado un grupo de trabajo de la Sección de Catalogación de la IFLA entre 1992 y 1997. El resultado fue el informe FRBR *Functional requirements for bibliographic records*, en 1997 (López Guillamón, 2004).

En abril de 1999 se creó el Grupo de Trabajo sobre Requisitos Funcionales y Numeración de Registros de Autoridad (FRANAR) para continuar con el trabajo de FRBR y desarrollar un modelo conceptual para las entidades descritas en los registros de autoridad. En 2009 publicó la extensión del modelo conceptual para datos de autoridad FRAD: *Functional Requirements for Authority Data* (Requisitos Funcionales de los Datos de Autoridad), cuya finalidad era el intercambio internacional de datos de autoridad. El modelo define los datos de autoridad como información sobre una persona, familia, entidad corporativa u obra, como base para un punto de acceso controlado para los registros de un catálogo (IFLA, 2011).

En 2005 se formó el Grupo de Trabajo de la IFLA sobre los Requisitos Funcionales para Registros de Autoridad de Materia (FRSAR) para abordar cuestiones sobre los datos de autoridad de materia. Dio lugar en 2011 a la publicación del tercer modelo conceptual *Functional Requirements for Subject Authority Data* (Requisitos Funcionales para Datos de Autoridad de Materia, FR SAD) que se focaliza en los datos de autoridad de materia (IFLA, 2011).

Estos modelos teóricos conforman la llamada Familia FRBR, que utilizan un modelo entidad-relación en el que se representan los objetos bibliográficos basándose en tres elementos: las entidades, sus atributos y las relaciones. El objetivo que pretende alcanzar esta forma de descripción se basa en las tareas que el usuario desarrolla en el catálogo para recuperar la información. Por lo tanto, esta adaptación a las necesidades de búsqueda del usuario asegura que los elementos y las relaciones que se describen de los objetos digitales sean los básicos y esenciales (IFLA, 2010).

El modelo FRBR para registros bibliográficos establece la conceptualización de los objetos, cómo se describen y cómo se relacionan. FRBR es el ejemplo para entender la proximidad que tienen estos modelos teóricos con la nueva norma de descripción RDA. Este modelo se acerca a la descripción de la información

de acuerdo a los estándares de la web semántica y la hace más próxima al entorno bibliotecario.

FRBR consta de un total de diez tipos de entidades, dos tipos de relaciones, y los atributos o características de las entidades. Estas diez entidades están clasificadas en tres grupos, según sean entidades relacionadas con la descripción, con la autoría o la materia de un objeto bibliográfico. La siguiente tabla resume el modelo (Ríos Hilario y Martín González, 2005).

Tabla 1: Diez entidades clasificadas en tres grupos

Diez ENTIDADES	GRUPO 1 relacionado con la descripción: Obra, Expresión, Manifestación, Ítem	Las entidades relacionadas con el primer grupo, de descripción, son obra (por ejemplo, el Quijote), manifestación (por ejemplo, la novela en español o su traducción en inglés), expresión (por ejemplo, la versión publicada en Madrid en 1995), e ítem (el registro bibliográfico de la BNE o el de la LC).
	GRUPO 2 autoría: Persona, Entidad corporativa	Las entidades relacionadas con el segundo grupo, de autoría, son persona y autor corporativo.
	GRUPO 3 relacionado con la	Las entidades relacionadas con el tercer grupo, el de

	materia: Concepto, Objeto, Acontecimiento, Lugar	materia, son concepto (por ejemplo, Historia), objeto (por ejemplo, Conferencia Internacional sobre Principios y Desarrollo Futuro de las AACR), evento (por ejemplo, congreso o conferencia), y lugar (Toronto).
--	---	---

Fuente: Elaboración propia a partir del modelo FRBR de IFLA, 2004

Las posibles relaciones que se pueden dar entre todas estas entidades son de dos tipos (IFLA, 2004), se muestran en la siguiente tabla.

Tabla 2: Tipos de relaciones entre las entidades FRBR

Relaciones Generales	Entre entidades de distintos grupos: por ejemplo, autor y obra.
Relaciones Específicas	Entre entidades del mismo grupo: dos expresiones de una misma obra, por ejemplo, la versión de una obra impresa en Madrid en 1995 y la versión impresa en Londres en 2000.

Fuente: Elaboración propia extraído del modelo FRBR

Cada una de las diez entidades lleva asociadas un conjunto de características o atributos específicos y diferentes entre sí. Los

atributos son específicos para cada entidad, es decir, son distintos para las obras, expresiones, manifestaciones, ítems, para las personas y autores corporativos, y para los conceptos, objetos, acontecimientos y lugares. Los atributos se agrupan en dos categorías (FRBR, 2004), como presenta la siguiente tabla.

Tabla 3: Grupos de clasificación de los atributos de las entidades FRBR

Primer grupo	Recoge los atributos inherentes a la entidad, como son las características físicas, por ejemplo, soporte y dimensiones, y también características como la información que aparece en la portada y en la cubierta.
Segundo grupo	Engloba todas aquellas características que hacen referencia a una fuente externa, como los identificadores asignados a una entidad, por ejemplo, el registro de una biblioteca, o información contextual como por ejemplo, el contexto político. (Ver listado de atributos para cada entidad en FRBR, 2004).

Fuente: Elaboración propia a partir del modelo FRBR

Esta manera de entender los objetos del mundo bibliográfico ha supuesto una transformación de los principios sobre los que se basa todo el proyecto internacional de control bibliográfico, a través de los distintos programas nacionales e internacionales de codificación y de descripción catalográfica que se han venido desarrollando. Estos modelos teóricos emplean las categorías

lógicas englobadas en el método informático y de teoría de sistemas de análisis entidad/relación (IFLA, 2004), lo que aporta ventajas en la catalogación y en la visualización de los registros en el catálogo. Vincula todos los materiales relacionados de una obra, es decir, todas las variantes de la obra (expresión, manifestación y ejemplar), incluso si estos materiales se catalogan en diferentes lenguas o ediciones. Se puede tomar como ejemplo de aplicación de este modelo el catálogo mundial WorldCat (Santana Arroyo, 2011).

En 2002 empezó la revisión de las AACR2 para redactar el borrador de las AACR3, que debía tener mayor coherencia con el modelo FRBR. Este nuevo borrador cambió el nombre a *Resource Description and Access*, para darle una perspectiva más internacional y no sólo anglo-americana (Tillett, 2009). En 2010 se publicó la nueva normativa RDA: *Resource Description and Access* para la descripción bibliográfica en el entorno web, que sustituye las *Anglo American Cataloguin Rules*, 2ª ed. (AACR2), como se dijo anteriormente. Se publicaron conjuntamente por la ALA, la Canadian Library Association, y CILIP. Son actualizadas y mantenidas por el Joint Steering Committee for Development of RDA (JSC). RDA no son reglas de catalogación, son estándares de representación de datos que se presentan independientes de cualquier formato de almacenamiento (como ISBD o MARC 21). Están basadas en el modelo conceptual FRBR y FRAD para la descripción y el

acceso de los recursos digitales, por lo que la estructura del código se basa también en entidades, atributos y relaciones. Aúna los conceptos catalográficos para diferentes esquemas de codificación y todo tipo de recursos y de contenido en los entornos electrónicos. Permite la compatibilidad con los modelos y normas internacionales actuales que se requieren para la catalogación o con nuevas estructuras que puedan surgir (Tillett, 2009).

El Joint Steering Committee (JSC) quiso evaluar la implantación de las RDA, por lo que en 2010 se realizó una prueba de creación de registros siguiendo las instrucciones RDA, en las tres bibliotecas nacionales de Estados Unidos, la Library of Congress, la National Library of Medicine (NLM) y la National Library of Agriculture (NAL) y en 26 instituciones miembro, con el objetivo de evaluar los resultados y proponer los cambios necesarios (Tillett, 2009). En 2011 la LC, NAL y NLM publicaron los resultados en *Report and Recommendations of the U.S. RDA Test Coordinating Committee* (JSC, 2011) y los principales resultados de las reuniones anuales celebradas en noviembre de 2012 y 2013, en las que se trabaja sobre las propuestas, cambios y acciones a realizar en las RDA que se recogen en los correspondientes *anual reports* (JSC, 2014).

En el ámbito europeo en 2009 surgió el European RDA Interest Group (EURIG) durante la American Library Association

Annual Conference en Chicago, USA. Desde entonces se reúne anualmente para evaluar la situación actual y las especificidades de RDA para analizar su posible adopción por parte de las bibliotecas europeas. Estos resultados quedan reflejados en los informes anuales, el último en abril de 2014 (EURIG, 2014).

La conceptualización de los objetos bibliográficos de RDA va en consonancia con la estructura del modelo de datos enlazados de la web semántica. Ya que como se ha dicho, la estructura de RDA (entidades, atributos y relaciones) basadas en los modelos FRBR, de planteamiento informático, coincide con la estructura del lenguaje de descripción de datos RDF (*Resource Description Framework*), y el uso de URI para identificar los objetos digitales, como se verá en el siguiente capítulo (FRBR, 2004). RDA constituye una herramienta de acercamiento de los sistemas de catalogación y organización de la información del mundo bibliotecario con el mundo de la web, y sus tecnologías semánticas. Esta convergencia puede contribuir a que los datos bibliográficos salgan de los circuitos exclusivos de la comunidad bibliotecaria ampliando su visibilidad a todos los ámbitos en los que tiene alcance la web (Picco y Ortiz, 2012).

Reglas de catalogación nacionales

En el ámbito nacional, la Biblioteca Nacional Española dependiente del Ministerio de Cultura, y la Dirección General del Libro, Archivos y Bibliotecas, constituye el órgano responsable de la elaboración de los instrumentos para el intercambio de la información bibliográfica. Sin estos instrumentos, no sería posible la coherencia en los catálogos ni el acceso sin ruidos ni silencios a la información bibliográfica por parte de los usuarios (Dirección General del Libro, Archivos y Bibliotecas, 1995).

La Biblioteca Nacional Española fue fundada en 1711 por Felipe V como Biblioteca Real, y se abrió al público en 1712 como Real Biblioteca Pública (BNE, 2014). Hacia 1815, Pedro García, bibliotecario, redactó las primeras reglas de catalogación *Reglas que se han de observar para hacer las cédulas para un índice general*. Se basaban en el código francés de 1791, de Jean-Baptiste Massieu. Tuvieron una escasa influencia, pues en 1836 se escribió el primer manual breve de catalogación en español, *Método antiguo de hacer las cédulas para insertarlas después en el índice*. Ese mismo año la Biblioteca Real cambió su denominación por Biblioteca Nacional y pasó a depender del Gobierno (BNE, 2014).

En 1857 Indalecio Sánchez Moreno de Tejada publicó la *Instrucción para formar los índices de impresos existentes en la Biblioteca Nacional*. De 1882 se conoce la *Instrucción para formar los índices de impresos de las bibliotecas administradas por el cuerpo de archiveros, bibliotecarios y anticuarios* con una clara influencia de Panizzi (BNE, 2014). En 1902, se publicó lo que ha sido considerado el primer código español *Instrucciones para la redacción de los catálogos en las Bibliotecas Públicas del Estado* dictadas por la Junta Facultativa de Archivos, Bibliotecas y Museos, ya inspiradas en las Instrucciones Prusianas de 1899 (BNE, 2014; Escolano, 2005).

En la nueva edición de las Instrucciones de 1941, se introdujeron modificaciones para actualizarlas y adaptarlas a la evolución de la técnica catalográfica. La corriente europea de principios de siglo venía representada por el Código Vaticano de 1931, que a su vez sintetizaba la corriente europea y la norteamericana. El código angloamericano de 1908 se basaba en los códigos de Panizzi, Cutter y las Instrucciones prusianas. En 1964 España dio un giro hacia la tendencia angloamericana, que se consolidó con la edición, en ese mismo año, de unas nuevas instrucciones basadas en los Principios de París de 1961 (BNE, 2014; Escolano, 2005).

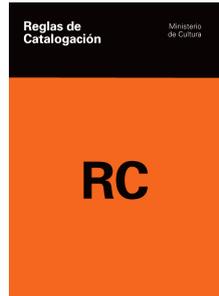
En 1985 apareció la primera versión de las Reglas de Catalogación que se conoce hoy día. Actualmente se maneja la cuarta versión de estas reglas que se publicaron en dos tomos, uno para la descripción general, monografías y títulos de revista, y otro para materiales especiales. Estas reglas se adaptan a los acuerdos sobre descripción bibliográfica procedente de la Reunión de Expertos en Catalogación que se celebró en Copenhague en 1969, las ISBD (M) de 1974 y las AACR2 de 1978. La segunda edición de las AACR establecía una nueva estructura de los códigos nacionales que también se sigue en España: comienza con la descripción bibliográfica y sigue con los puntos de acceso. La edición de 1995 unificó esos dos tomos, e incluyó en un capítulo general la parte de descripción bibliográfica aplicable a todo tipo de documentos, previo a los capítulos específicos con las particularidades de descripción propias de cada material. Las últimas ediciones ya con pocas modificaciones, han sido en 1999 y en 2010 (BNE, 2014; Escolano, 2005).

La Biblioteca Nacional venía empleando junto con las Reglas de Catalogación nacionales el formato IBERMARC que sustituyó por MARC 21 en 2007. Desde 2009 adoptó la ISBD consolidada. Es miembro de EURIG y desde hace varios años viene trabajando en el estudio de todo lo relacionado con las RDA, bajo la coordinación del Departamento de Proceso

Técnico de la BNE. En 2010 participó de las pruebas preparadas por la LC y por el JSC, aunque no han tomado ninguna decisión final sobre la implantación de las RDA (Picco y Ortiz, 2012).

La ilustración 2 presenta la última edición de las Reglas de Catalogación españolas.

Ilustración 2: Reglas de catalogación españolas, edición 2010



Fuente: Ministerio de Educación, Cultura y Deporte, 2012
<<http://www.mecd.gob.es/servicios-al-ciudadano-mecd/catalogo/cultura/mc/reglas-de-catalogacion/presentacion.html>>

La información científica

Este panorama sobre las normas de descripción en instituciones de tipo bibliotecario se ve completado cuando se trata de información de tipo científico en relación con las prácticas que se utilizan para describir los artículos de revista, que en el entorno bibliotecario se han denominado partes componentes. En general estas reglas no han sido muy utilizadas, como veremos en el apartado siguiente, pero poco a poco se ha producido una confluencia entre los métodos de trabajo de ambos sectores, el patrimonial o bibliotecario y el científico centrado en revistas y congresos en amplia medida. Algo que ha venido a modificarse a partir de la aparición de las revistas electrónicas, la web y los movimientos de acceso abierto. Este último factor, el acceso abierto, está suponiendo un cambio de paradigma en el acceso a la información generada por los científicos y humanistas. Como desgranaremos a lo largo de este apartado, las motivaciones para difundirlas han supuesto un cambio abrupto en el panorama existente, aunque la forma de recogerlo, transmitirlo y describirlo no está demasiado lejana de las bibliotecas tradicionales, al menos de las digitales. Algo que sustancialmente es distinto a lo que sucede con las tecnologías semánticas que se proponen para Vestigium, y que suponen un cambio disruptivo en todas sus dimensiones.

El proceso de creación de literatura científica y académica refleja claramente el carácter cíclico de la información que reseñábamos al comienzo del capítulo. Los investigadores y académicos necesitan acceso a la información especializada para generar nuevas investigaciones y nueva literatura. La consulta de estas fuentes primarias que reúnen el conocimiento, las revistas, tiene costes muy elevados, pero la fiabilidad es indiscutible, ya que han pasado por un proceso de evaluación. Los investigadores, tras haber utilizado la información producen nuevo conocimiento que tradicionalmente era difundido por el sector científico editorial, encargado de organizar el *peer review* (la revisión por pares), de la impresión de revistas y libros y de su distribución por canales comerciales. Por estas funciones se transmitía a los editores, que comenzaron siendo las propias sociedades científicas, el copyright sobre los trabajos. Pero la evolución seguida ha venido a modificar el equilibrio existente. La edición de ciertas revistas de alto impacto se convierte en un negocio estrictamente privado, basado en la lógica de mercado. Además, surgen las revistas electrónicas con unos costes supuestamente menores al eliminar el proceso de impresión. Y por último aparece un inexplicable y constante aumento del precio de las revistas. Estas tres razones dan lugar a lo que se conoce como crisis de las revistas en la década de los años 80 y 90 del siglo XX (Andrade, 2009).

Esta situación evidencia una paradoja. En muchas ocasiones las

investigaciones financiadas en parte o en su totalidad por los organismos públicos, cuyos resultados se publican por editoriales científicas que los comercializan. Generalmente los autores no reciben compensación económica por publicar sus trabajos, sino más bien el procedimiento es el inverso, en algunas revistas tienen que pagar para publicar sus artículos. Gastos que se imputan en los presupuestos de las investigaciones. Así mismo, cuando un investigador necesita consultar un artículo, las instituciones de carácter superior y de investigación debe pagar a las editoriales por el acceso a una investigación que previamente han sufragado.

En este contexto se desarrolla un movimiento que promueve el acceso libre al conocimiento denominado *Open Access* (OA). El movimiento OA impulsa el acceso libre a la información y a la ciencia, sin barreras económicas ni restricciones de derechos de autor, con miras a la difusión a un público más amplio y a la reutilización del conocimiento para continuar con el ciclo de la información.

Peter Suber (2004), profesor en la Universidad de Stanford y autor de algunos de los textos más citados sobre el acceso abierto a la ciencia define la literatura científica de OA como: “Open-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions. What makes it possible is the Internet and the consent of the author or copyright-holder. OA is entirely compatible with peer review,

and all the major OA initiatives for scientific and scholarly literature insist on its importance”.

El OA en la información científica es un modelo vinculado a la forma de comunicación de la actividad investigadora impulsado por el desarrollo de las nuevas tecnologías y la generalización del uso de Internet. En la definición de OA se alude a tres características de los contenidos de libre acceso: su carácter digital y su disponibilidad en línea gracias a Internet; su carácter gratuito para los usuarios; y la ausencia de restricciones de derechos de autor. Otro elemento destacable es que se mantiene la revisión por pares igual que en el modelo tradicional de publicación y acceso por suscripción, para asegurar la calidad de la misma (Franklin, 2003).

Existen una serie de declaraciones emitidas en consenso por grupos de expertos y científicos, sociedades profesionales e instituciones internacionales que sostienen y perfilan el movimiento OA. Una de las más importantes es la Declaración de Budapest de 2002 (*Budapest Open Access Initiative*, BOAI) que define como acceso abierto a la científica erudita su disponibilidad gratuita en Internet para que cualquier usuario la pueda leer, descargar, copiar, distribuir o imprimir, con la posibilidad de buscar o enlazar al texto completo del artículo, utilizarlo para cualquier propósito legal, sin otras barreras financieras, legales o técnicas distintas de la fundamental de acceder a la propia Internet. El único límite a la reproducción y

distribución de los artículos publicados, y la única función del copyright en este marco debe ser garantizar a los autores el control sobre la integridad de su trabajo y el derecho a ser acreditados y citados.

Para alcanzar el acceso abierto a la literatura científica se definen dos vías: la vía dorada de modelo de publicación OA (*Gold road*) y la vía verde de autoarchivo en repositorios abiertos (*Green road*).

Publicación *Open Access* o *Gold road*

El modelo de publicación en acceso abierto presenta diversos grados de accesibilidad entre las revistas científicas existentes. Los diferentes niveles de acceso van desde el acceso gratuito, pasando por el establecimiento de periodos de embargo, hasta llegar al pago por publicación. Estos grados de accesibilidad definen diversos ejemplos financieros para superar los gastos derivados del proceso de publicación de la actividad investigadora. Por lo que en este modelo de publicación de acceso abierto según la vía dorada surgen cuatro tipos en los que se puede encuadrar la literatura disponible y cosechable en la web (Melero y Abad, 2008).

- Gratuitas y libres para lectores y autores: los costes son

asumidos por los editores que suelen ser departamentos universitarios, centros de investigación subvencionados u organismos públicos. Este modelo suele estar financiado por fuentes públicas o son revistas en las que la suscripción de la versión en papel proporciona los ingresos suficientes para que la versión en línea sea absolutamente gratuita.

- Pago por publicación en revistas OA: es el autor el que asume los costes de la revisión del material y su publicación. Esta modalidad se fundamenta en que en escasas ocasiones es el autor el que verdaderamente paga las tasas, siendo más habitual que asuma los gastos la institución a la que está adscrita o la biblioteca. Este podría ser el modelo económico contrapuesto al de pago por suscripción.
- Pago por publicación de artículos en OA en revistas de pago por suscripción: es un modelo denominado híbrido porque coexisten artículos abiertos con artículos de acceso por suscripción. Se trata de la posibilidad que ofrecen las editoriales comerciales de publicar trabajos en OA mediante el pago de unas tasas por parte de los autores. Surge por las recomendaciones de los organismos financiadores de hacer disponibles en acceso abierto los resultados de las investigaciones financiadas. Grandes editoriales como Springer y Oxford University

Press cuentan con él. De esta forma las editoriales mantienen sus ingresos por suscripción e incorporan ingresos por publicación, mientras publican artículos en acceso abierto.

- Acceso gratuito con periodo de embargo: se trata de revistas comerciales que siguen el sistema tradicional de acceso por suscripción, tanto en la versión en papel como en la digital, que tras un periodo de embargo ofrecen acceso gratuito a la publicación. El periodo de embargo oscila entre los seis meses y los dos años. Aunque hay que destacar que este modelo no cede los derechos de explotación con lo que no está permitido copiar o distribuir los contenidos. Este modelo no cumpliría con la característica de publicación de acceso abierto de forma inmediata.

Estas formas de publicación se han ido implantando en los últimos años y van ganando mercado, pues cada vez son más las editoriales que ofrecen la opción de publicar en la versión OA de las revistas. Aunque, a pesar de su expansión, existe discrepancia entre los autores, pues no todos están de acuerdo en ofrecer sus resultados en acceso libre. Algunos consideran que la limitación en el uso de la obra no es elemento clave para el desarrollo de esta filosofía, sino que el triunfo del OA puede depender de que se asegure la realización de *peer review*, para asegurar la calidad de la publicación (Franklin, 2003). Pero este

tipo de publicaciones pasa por el mismo proceso de revisión por pares que las publicadas en el modelo tradicional, lo que garantiza su calidad al mismo nivel.

El modelo de negocio OA a priori supone una competencia al statu quo de la edición científica. En principio genera pérdidas financieras en las editoriales académicas en cuanto a lo referido a la suscripción de sus clientes tradicionales (universidades, instituciones gubernamentales, empresas privadas). Los costos editoriales se deben a que estas grandes editoriales invierten recursos en los procesos de revisión por pares y en mantener sus archivos de preservación, que quedan cubiertos por las suscripciones a publicaciones periódicas y bases de datos. Estos elementos de retorno, la suscripción, no se contemplan en los modelos de publicación OA (Arriola, 2011).

Por el momento convive el modelo tradicional de acceso exclusivo por suscripción a las publicaciones en las que las editoriales poseen los derechos exclusivos de sus artículos, junto con el modelo de publicación en OA. Éste último proporciona numerosas ventajas, entre las que se destaca el aumento de la visibilidad de los trabajos y como consecuencia de sus autores, junto a la reducción de costes que supone la publicación con este sistema. Tal y como destaca la declaración de Budapest (2002) la publicación en OA es económicamente viable, los costes totales para dar acceso abierto en línea son mucho más bajos que los costes de las formas tradicionales de difusión. Con la

oportunidad de ahorrar costes y al mismo tiempo ampliar el ámbito de la difusión proporciona a los lectores acceso a literatura relevante, y brinda a los autores y a sus trabajos mayor visibilidad, un nuevo impacto, y un público más amplio. La ilustración 3 presenta el logotipo del OA.

Ilustración 3: Logotipo del Open Access



Fuente: Modificación de Córdoba González, Saray. Acceso abierto al conocimiento en la Universidad de Costa Rica. Universidad de Costa Rica, 2011<<http://www.vinv.ucr.ac.cr/docs/multimedia/repositorios-ucr.pdf>>

Depósito en repositorios abiertos o *Green road*

Para alcanzar el OA en la literatura, la vía verde (*Green road*) establece el depósito de las publicaciones en repositorios en línea de acceso abierto. La Declaración de Bethesda de 2003 (*Bethesda Statement on Open Access Publishing*) apoya esta vía verde estableciendo que una versión completa de la obra y todos los materiales suplementarios debe depositarse de forma inmediata a la publicación inicial en al menos un repositorio en línea apoyado por una institución académica, una sociedad de intelectuales o una agencia gubernamental, para facilitar el acceso abierto, la distribución sin restricciones, la

interoperabilidad y el archivado a largo plazo.

En concordancia con esta declaración, se redacta en el Max Planck Institute la Declaración de Berlín en octubre de 2003 (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*), promovida por diferentes representantes políticos y científicos principalmente de varias instituciones europeas. La Declaración pone de relevancia la importancia del movimiento internacional de OA desde el punto de vista de la política científica para impulsar el desarrollo de iniciativas y proyectos de apoyo por parte de las instituciones miembro.

Posteriormente se han ido sucediendo declaraciones internacionales de apoyo a este movimiento, como la Declaración del Wellcome Trust de 2003, *The Valparaíso declaration for improved scientific communication in the electronic medium* de 15 de enero de 2004, la *IFLA statement on open access to scholarly literature and research documentation* de 24 de febrero de 2004, la *Washington D.C. principles for free access to science* de 16 de marzo de 2004. En noviembre de 2004 treinta universidades italianas firmaron su adhesión a la Declaración de Berlín en la denominada Declaración de Messina, que ahora cuenta con más de setenta universidades. En la XII Asamblea General de REBIUN (Red de Bibliotecas Universitarias), de 2004, la CRUE: Conferencia de Rectores de Universidades Españolas se adhiere también a esta declaración.

Con miras al objeto de investigación de esta tesis sobre la consulta de producción académica y científica de autores valencianos, cabe destacar que instituciones académicas valencianas firmaron la adhesión a la Declaración de Berlín: la Universidad de Alicante, la Universidad Miguel Hernández de Elche, la Universidad Jaume I en 2005, la Universitat de València Estudi General firmó en septiembre de 2008 y la Universitat Politècnica de València en octubre de 2011. Lo que hace disponible la producción científica y académica de docentes e investigadores de estas instituciones.

La vía verde de depósito en repositorios de acceso libre se fomenta en Europa con las recomendaciones europeas que marca el European Research Advisory Board (EURAB), comité asesor creado por la Comisión Europea para la implementación de la política científica europea. Según el informe final sobre las políticas de acceso abierto para las publicaciones científicas de diciembre de 2006 *Scientific publication: Policy on Open Access* (EURAB, 2006), se recomienda que todos los investigadores financiados bajo el Séptimo Programa Marco de Investigación (FP7) estén obligados a depositar los resultados publicados de su investigación en repositorios digitales de acceso abierto en el plazo de seis meses (Comisión Europea, 2008).

Otras políticas y estrategias que secundan el desarrollo del OA a la información científica en el Sur de Europa son la Declaración

de la Alhambra sobre acceso abierto que surgió en Granada en 2010 en el seminario *Policies for the Development of OA in Southern Europe*, organizado por la Fundación Española de Ciencia y Tecnología (FECYT) y el consorcio de bibliotecas del Sur de Europa (Southern European Libraries Link, SELL) (Anglada, 2013).

El programa de financiación europea Horizon 2020 (H2020) fomenta el avance hacia la denominada Ciencia Abierta (*Open Science*) estableciendo la obligación de acceso abierto de forma gratuita de todos aquellos artículos de revista publicados que deriven de proyectos financiados por el programa H2020. Para tal efecto han implementado la infraestructura OpenAIRE, que proporciona un repositorio para publicaciones y datos. Este programa de financiación de la investigación e innovación sustituye a los conocidos Programas Marco vigentes desde 1984, con su primera convocatoria el 11 de diciembre de 2013. La Comisión Europea, concretamente el European IPR Helpdesk, ha publicado una breve guía de preguntas frecuentes para resolver las principales cuestiones sobre el acceso abierto en el programa H2020 respecto a dos categorías, en las publicaciones científicas revisadas por pares, y en los datos de investigación generados en el desarrollo del proyecto (European Commission, 2014). Esta nueva hoja informativa es complementaria a la *Guidelines on Open Acces to Scientific Publications and*

Research Data in Horizonte 2020, de fecha 11 de diciembre de 2013.

Una acción que va a tener una gran repercusión en el futuro inmediato del OA a nivel mundial es el anuncio, en mayo de 2014, de dos políticas de acceso abierto por parte de las agencias chinas National Natural Science Foundation of China (NSFC) y Chinese Academy of Sciences (CAS). Estos mandatos obligan a depositar las publicaciones de las investigaciones financiadas en repositorios en línea y de hacerlas accesibles públicamente tras un periodo de embargo de no más de doce meses desde su publicación. Por el momento recomiendan el depósito en repositorios institucionales a la espera del repositorio NSFC en 2016 (Van Noorden, 2014). La puesta en marcha de estas políticas supone un considerable apoyo al movimiento OA al constituir la producción investigadora de autores chinos en revistas internacionales el 13,9% del total de la producción mundial, según datos del 2012 de la base de datos Science Citation Index (SCI) de Thomson Reuters (Zhang, 2014).

Por otro lado, el depósito de conocimiento científico en OA en el ámbito nacional, se ve apoyado con la aprobación de la Ley de la Ciencia y la Tecnología de España (2011) que incluye un artículo sobre la publicación en OA, que dispone que todos los investigadores cuya actividad haya sido financiada a través de los presupuestos generales del estado está obligado a depositar

en acceso abierto una versión digital de la publicación de la investigación. Ley 14/2011, de 2 de junio, de la Ciencia, la Tecnología y la Innovación, que deroga a la antigua Ley 13/1986 de 14 de abril, de Fomento y Coordinación General de la Investigación Científica y Técnica, en el Capítulo II. Transferencia y difusión de los resultados de la actividad de investigación, desarrollo e innovación y cultura científica, tecnológica e innovadora y en concreto en su Artículo 37. Difusión en acceso abierto:

1. Los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación impulsarán el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación, y establecerán sistemas que permitan conectarlos con iniciativas similares de ámbito nacional e internacional.
2. El personal de investigación cuya actividad investigadora esté financiada mayoritariamente con fondos de los Presupuestos Generales del Estado hará pública una versión digital de la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadas o periódicas, tan pronto como resulte posible, pero no más tarde de doce meses después de la fecha oficial de publicación.

3. La versión electrónica se hará pública en repositorios de acceso abierto reconocidos en el campo de conocimiento en el que se ha desarrollado la investigación, o en repositorios institucionales de acceso abierto.
4. La versión electrónica pública podrá ser empleada por las Administraciones Públicas en sus procesos de evaluación.
5. El Ministerio de Ciencia e Innovación facilitará el acceso centralizado a los repositorios, y su conexión con iniciativas similares nacionales e internacionales.
6. Lo anterior se entiende sin perjuicio de los acuerdos en virtud de los cuales se hayan podido atribuir o transferir a terceros los derechos sobre las publicaciones, y no será de aplicación cuando los derechos sobre los resultados de la actividad de investigación, desarrollo e innovación sean susceptibles de protección.

Estas políticas fomentan que los repositorios institucionales nacionales sean de acceso abierto y que se pueda consultar la literatura académica. Al objeto de la presente tesis, resulta especialmente relevante, pues este acceso abierto va a permitir localizar literatura y estudios académicos que hayan realizado los investigadores y humanistas que se estudian en Vestigium; y

a un tiempo, se producen contenidos que están liberados, si bien no se cosecharán mediante el protocolo utilizado por este tipo de bibliotecas, sino que usa tecnologías más avanzadas, compatibles con todo tipo de contenidos de la web semántica.

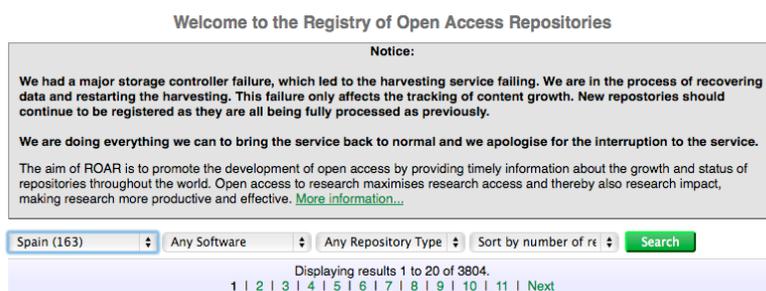
Este conjunto de declaraciones y políticas recomiendan el depósito de los artículos en repositorios electrónicos abiertos. Los repositorios son archivos que almacenan recursos digitales, textuales, de imagen o sonido, que surgen para maximizar la difusión y el impacto de los trabajos científicos (*e-prints*, que incluye los *pre-prints*, los *post-print* y los *reprints*) (Melero, 2005). Se categorizan en dos tipos, los temáticos, que albergan contenidos en función de un área de conocimiento, y los institucionales, que Clifford Lynch (2003) define como “a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution”. Lynch ofrece una visión en la que destaca el compromiso entre la institución y los propios investigadores para preservar la producción en beneficio de ambas partes.

Los repositorios más antiguos citados por el Registro de Repositorios de Acceso Abierto ROAR, son temáticos, Swiss

Federal Institute of Technology, 1991; y ArXiv (Cornell University), también de 1991, que contienen trabajos de una sola disciplina procedentes de múltiples instituciones.

Según Barrueco y Subirats (2003) se trata de una tecnología cuya adopción inicial es paulatina, pero la respuesta positiva a su adopción aumenta de forma exponencial, superando en más de 100 las instituciones que han creado archivos abiertos en 2003. Han sido numerosas las instituciones que han abogado por su uso, y otras lo han respaldado, como la Federación de Bibliotecas Digitales (DLB) y la National Science Foundation USA (NSF) (Gómez y Arias, 2002). Para el objeto de esta tesis nos planteamos conocer las fuentes de información abiertas, puesto que los trabajos de los científicos y humanistas que estamos intentando identificar pueden haber sido recogidos en ellas. La situación ideal sería que estos repositorios, a su vez, pudieran ser fuentes para una biblioteca semántica como la que diseña esta tesis. Desgraciadamente, sólo algunos de los archivos españoles han implementado la tecnología necesaria para ser compatible con la web semántica al utilizar el software de gestión DIGIBIB (DIGIBÍS, 2014). La siguiente imagen muestra la interfaz del registro de repositorios de OA ROAR, que recoge la cifra actual de archivos OA en España, un total de 163 repositorios.

Ilustración 4: Interfaz de consulta del registro ROAR de repositorios OA

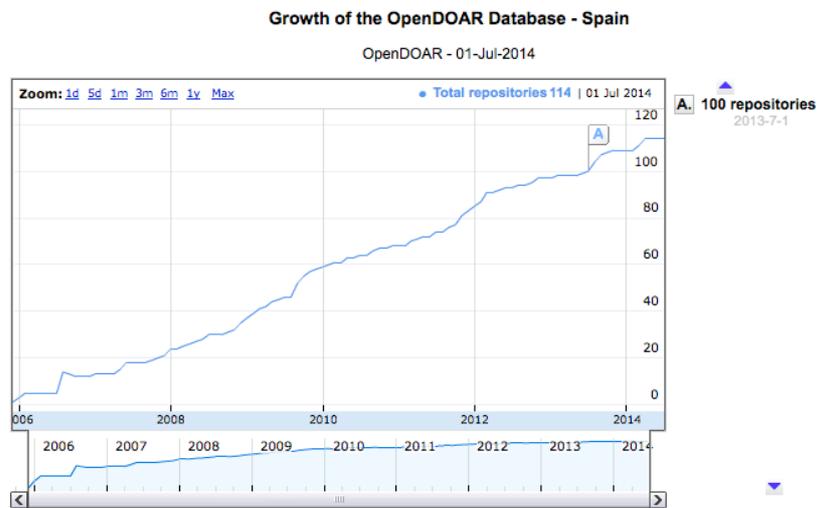


Fuente: OpenROAR, 1 de julio de 2014

En España la mayoría de las universidades cuentan con su propio repositorio institucional. El Registro de Repositorios de Acceso Abierto ROAR, a 1 de julio de 2014, recoge un total de 3.804 repositorios a nivel mundial, de los cuales 163 pertenecen a instituciones y organismos en España.

El Directorio de Repositorios de Acceso Abierto OpenDOAR en la misma fecha, recoge 2.685 repositorios a nivel mundial, de los cuales 114 son de organismos e instituciones españolas, como muestra la siguiente ilustración.

Ilustración 5: Gráfico del crecimiento de los repositorios españoles en la base de datos OpenDOAR



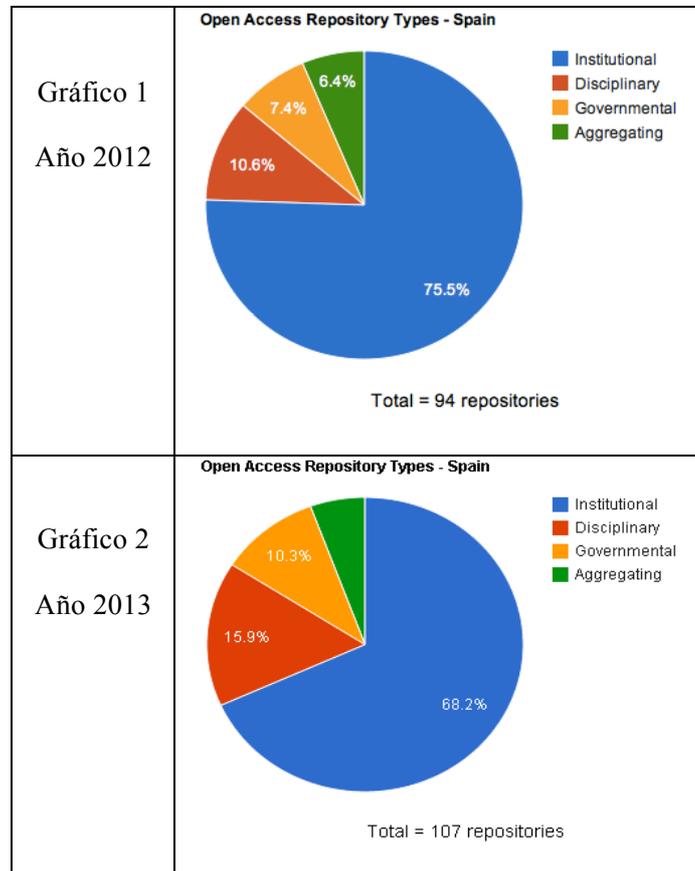
Fuente: OpenDOAR, 1 de julio de 2014

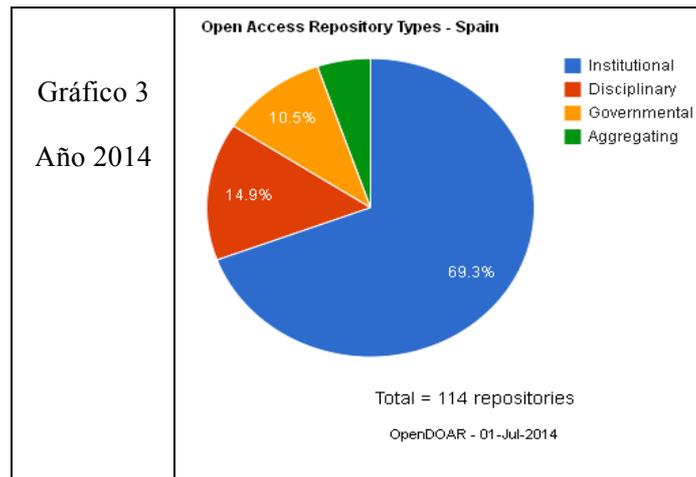
La imagen 5 muestra el crecimiento constante de repositorios españoles en la base de datos del directorio OpenDOAR, desde el año 2006 cuando contaba con cinco repositorios. Se observa un considerable aumento en la creación de repositorios, con 94 en el año 2012, un aumento a 107 repositorios en el año 2013, hasta alcanzar un total de 114 repositorios el 1 de julio el año 2014.

Otra fuente para conocer los repositorios españoles es BuscaRepositorios, un directorio de repositorios institucionales españoles de acceso abierto surgido del proyecto I+D del grupo de investigación de acceso abierto a la ciencia. Con fecha de 1

de julio de 2014 registra un total de 120 repositorios de instituciones académicas, de investigación, y culturales existentes en España.

Ilustración 6: Gráfico de las tipologías de repositorios de acceso abierto en España en los años 2012, 2013 y 2014





Fuente: OpenDOAR, 1 de julio de 2014

En la imagen 6 se puede observar el crecimiento de repositorios OA españoles durante el periodo de 2012 a 2014, que ha aumentado en 20. Los tres gráficos muestran la distribución de los repositorios españoles de OpenDoar por tipología para el periodo de 2012 a 2014.

Los repositorios mayoritarios durante los tres años (75,5% en 2012, 68,2% en 2013, 69,3% en 2014) son los institucionales, principalmente repositorios de universidades y centros de investigación, junto con grandes empresas. Esto puede deberse a que se trata de instituciones con abundante producción académica y científica, y en muchos casos con una profunda tradición en aspectos de presevación, que cuentan con bibliotecas universitarias y personal especializado en gestión de información.

El segundo gran grupo de repositorios son los temáticos o disciplinares (10,6% en 2012, 15,9% en 2013 y 14,9% en 2014), debido posiblemente a la creación de repositorios especializados por áreas de conocimiento, por parte de las editoriales académicas y científicas que se encargan de la preservación de sus publicaciones.

El tercer gran grupo lo constituyen los repositorios gubernamentales (7,4% en 2012, 10,3% en 2013 y 10,5% en 2014), posiblemente creados en respuesta a las políticas gubernamentales que se vienen desarrollando en los últimos años para preservar y difundir los resultados de la investigación subvencionada, que recomiendan y en ocasiones obligan a su depósito en determinados repositorios OA.

Por último, el grupo minoritario (6,4% en 2012, 5,5% en 2013 y 5,3% en 2014) lo conforman repositorios agregados o portales. Este grupo es minoritario por que los repositorios agregados se basan en el cosechado de otros repositorios (Ferreyra, 2009).

Los repositorios de instituciones valencianas recogidos en OpenDoar a fecha de julio de 2014, que pueden albergar contenidos de investigadores y académicos valencianos son:

- Bivaldi: Biblioteca Valenciana Digital.
- RUA: Repositorio Institucional de la Universidad de Alicante.

- RODERIC: Repositori Institucional de la Universitat de València.
- RiuNet: Repositorio Institucional de la Universitat Politècnica de València.
- UJI: Repositorio Univesitat Jaume I.
- REDIUMH: Repositorio Digital de la Universidad Miguel Hernández de Elche.

Otros repositorios nacionales recogidos en OpenDoar que son fuentes para consultar porque pueden albergar documentos relacionados con los autores valencianos son los siguientes:

- Biblioteca Digital Real Academia de la Historia.
- Biblioteca Digital Real Academia Nacional de Medicina.
- Biblioteca Digital Hispánica.
- Biblioteca Virtual del Patrimonio Bibliográfico (Virtual Library of Bibliographical Heritage).
- Biblioteca Virtual Miguel de Cervantes.
- Digital.CSIC.
- Fotografía Sobre España en el Siglo XIX.
- Fundacion Mapfre.
- FHL Virtual Libraries.

- Scientific Electronic Library Online – Spain.
- TDX: Tesis españolas.
- Dialnet.

Estos datos reflejan que la opción de depositar la producción de la actividad investigadora de forma abierta es un servicio utilizado cada vez más por un gran número de investigadores, y que su incremento es progresivo. A ello contribuye que sean las propias instituciones las que pongan a disposición de los académicos un repositorio y establezcan pautas para fomentar su uso. Esto supone la intersección entre las dos vías recomendadas para el acceso abierto, la publicación en revistas y el depósito en archivos. De hecho, algunos editores comerciales internacionales de reconocido prestigio que ofrecen la opción de publicar en OA, incluyen el depósito de los trabajos publicados y los materiales suplementarios en repositorios libres tras un breve periodo de embargo. Además, en ocasiones recomiendan en qué repositorios pueden depositar el trabajo y los materiales, generalmente en función de la tipología. Las características y detalles de estas especificaciones se recogen en las políticas de auto-archivo de las normas para autores de las editoriales.

Algunas de las editoriales que realizan este tipo de recomendaciones son Nature Publishing Group que, dados los mandatos emitidos por instituciones de financiación, promueven

el depósito de los artículos en archivos de acceso público tras un periodo de embargo que varía en función de la revista. Elsevier en las políticas de autoarchivo establece que los artículos se ponen en acceso abierto después de un período de embargo determinado por las características de la revista. Springer establece que los autores pueden autoarchivar la versión aceptada en sus propios sitios web y en cualquier repositorio público doce meses después de su publicación oficial. Oxford Journals y BioMed Central establecen que todas sus revistas tienen políticas específicas de autoarchivo donde detallan las medidas adoptadas para el depósito en repositorios de la versión definitiva del trabajo. Pero la mayoría de las revistas de Oxford cuentan con entre doce y veinticuatro meses de embargo desde la publicación del trabajo y las revistas de BioMed Central entre seis y doce meses.

La política de las distintas editoriales se puede consultar en el recurso ROME0-Sherpa para las publicaciones a nivel internacional y en Dulcinea para las publicaciones a nivel nacional. Estas herramientas identifican y analizan las políticas editoriales de las revistas internacionales y nacionales respectivamente, respecto al acceso a sus textos y archivos, los derechos de copyright sobre los mismos y si permiten archivar el *pre-print* y el *post-print*. Según estas características las clasifican en cuatro categorías: verde, azul, amarillo y blanco (Melero et al., 2009).

- Verde: permite archivar el *pre-print* y el *post-print*
- Azul: permite archivar el *post-print*
- Amarillo: permite archivar el *pre-print*
- Blanco: No se permite el auto-archivo en ningún caso

Además, proporcionan información sobre las condiciones del auto-archivo o depósito de los trabajos en repositorios temáticos, institucionales o páginas web de los autores. Algunas editoriales permiten depositar los trabajos en las páginas web institucionales o bien de los autores, aunque esa práctica no asegura la accesibilidad y perdurabilidad del trabajo siguiendo las características del OA. En cambio, los repositorios tienen la particularidad de asegurar estas dos circunstancias, pues entre las características generales de los repositorios se encuentran (Alonso et al., 2008):

- El auto-archivo. Es el proceso por el que el contenido es depositado por el creador, propietario o una tercera parte en su nombre (editor).
- Libre accesibilidad al texto completo.
- Preservación a largo plazo.
- La interoperabilidad. A través del uso de procesos normalizados se establece la interconexión con otros archivos abiertos para intercambiar información (Protocolo OAI-PMH *Metadata Open Archives Initiative*

– *Protocol Metadata Harvesting*).

Por tanto, apreciamos que es a partir del movimiento para el acceso abierto a la información científica cuando comienzan a liberarse masivamente los contenidos generados por los autores académicos. Este hecho puede ser de utilidad para la biblioteca semántica Vestigium, al facilitar el acceso a las publicaciones científicas de los autores que se recogen en este recurso. Además debemos señalar que los investigadores que se están recopilando en este momento para Vestigium tienen superadas prácticamente todas las limitaciones que imponen las editoriales para el autoarchivo. Los nombres que ocupan los primeros puestos de la lista de Vestigium para incluir su obra en dicha biblioteca son científicos y humanistas que actualmente no se encuentran ya en activo. Quizás uno de los únicos problemas que se prevé para ligar su obra es el que no se encuentre digitalizada o en formato electrónico.

Interoperabilidad

Los repositorios siguen la iniciativa de archivos abiertos Open Archives Initiative (OAI) que proporciona la arquitectura y las especificaciones técnicas necesarias para mejorar la interoperabilidad entre archivos de publicaciones electrónicas, con el propósito de facilitar la diseminación y el acceso a los contenidos que archivan (Van de Sompel y Lagoze, 2008). Pero previamente ya existían tecnologías que hacían interoperable la información.

La interoperabilidad se puede clasificar en varios niveles y las características que la definen varían en función de la especificidad de la misma. En un contexto generalista, en el que probablemente la definición más extendida sea la de ISO y IEEE *Ieee Standard Glossary of Software Engineering Terminology* (1990) la interoperabilidad se define como la habilidad de dos o más sistemas de intercambiar datos, y utilizar mutuamente la información que ha sido intercambiada. Esta definición es clara para establecer en qué consiste el proceso de interoperabilidad.

Otros autores especifican tipos más concretos de interoperabilidad. Es el caso de Folmer (2012) que define la interoperabilidad inter-organizacional como la habilidad de dos o más sistemas de organización para intercambiar información e interpretar la información que ha sido intercambiada. La

matización que se hace con esta definición es la interpretación de la información. En este sentido, Folmer incluye la interoperabilidad que se conoce como semántica, es decir, aquella que abarca también el contenido, el significado y la interpretación de la información que es intercambiada; y la interoperabilidad técnica, es decir, el proceso técnico de intercambio de información entre sistemas.

Ouksel y Sheth (1999) categorizan la interoperabilidad en cuatro clases:

- Interoperabilidad entre sistemas: la compatibilidad entre hardware y sistemas operativos.
- Interoperabilidad sintáctica: similitud en codificación y representación.
- Interoperabilidad estructural: unificación de modelos de datos, estructuras de datos y esquemas.
- Interoperabilidad semántica: consistencia terminológica y de significado.

Extendiendo el concepto a un contexto general, Miller (2000) amplió la clasificación contemplando además:

- Interoperabilidad política/humana: referente a las decisiones que hacen los recursos ampliamente

disponibles.

- Interoperabilidad inter-comunidad: relativa a compartir información interdisciplinaria a través de fronteras.
- Interoperabilidad legal: relativa a características de disponibilidad de la información, regulaciones de protección de datos y derechos de propiedad intelectual.
- Interoperabilidad internacional: relacionada con la variedad de idiomas (Alemu, 2011).

En el contexto bibliotecario, según Gómez Dueñas (2005) se define como la capacidad que presenta un sistema de información de comunicarse y compartir información efectivamente con otro mediante una conexión libre y transparente (compartir metadatos, documentos y objetos digitales). Y según Martínez Equihua (2007) como la capacidad de un sistema o de un producto de trabajar con otros sistemas o productos sin esfuerzo especial por parte del cliente. Desde este punto de vista computacional, la interoperabilidad permite generar un enlace entre sistemas de trabajo para las diferentes tecnologías de información promoviendo una sana convivencia y operatividad.

La interoperabilidad es el motor para el intercambio de

información y se consigue cuando las máquinas entienden el significado de los datos distribuidos y son capaces de procesarlos de manera correcta (Subirats, 2013). La interoperabilidad entre repositorios es el elemento clave de la iniciativa OAI. La iniciativa OAI impulsó, en 1999 en la Convención de Santa Fe (Estados Unidos), la creación del protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Van de Sompel y Lagoze, 2008). En dicha convención se buscaba un servicio universal de autoarchivo de *e-prints* para los autores (Fox, 2000), y la clave era la interoperabilidad entre los sistemas, para poder entenderse entre ellos y compartir así los recursos que almacenan. La interoperabilidad es la capacidad de intercambiar datos entre dos sistemas sin la intervención de un tercero, como se ha definido anteriormente. De este modo la información compartida puede ser utilizada sin requerir una comunicación previa. Esta capacidad entre los archivos que cumplen con el protocolo OAI-PMH ha sido un factor clave en el éxito del movimiento de acceso abierto, pues un recurso de información aislado no es recuperable, no se comunica y por lo tanto no tiene visibilidad ni puede ser citado.

OAI surge como respuesta al problema de la dispersión de los documentos en múltiples depósitos, integrando en una única herramienta y haciendo posible la consulta a través de una única interfaz. OAI-PMH es un protocolo que proporciona una

infraestructura para la interoperabilidad de una aplicación independiente basada en la recolección de metadatos. Lo que permite la transferencia de documentos de distintos archivos distribuidos por la red (Van de Sompel y Lagoze, 2008). Hernández et al., (2007) citando a Suber (2007) lo define así:

La Iniciativa de Archivos Abiertos (OAI) define un protocolo (OAI-PMH) para recoger metadatos de ficheros de datos que residen en archivos separados. Cuando el protocolo es utilizado por servicios de datos como motores de búsqueda, éstos pueden procesar los datos de archivos separados como si residiesen en un solo archivo. En términos técnicos, el protocolo de recolección de metadatos soporta la interoperabilidad.

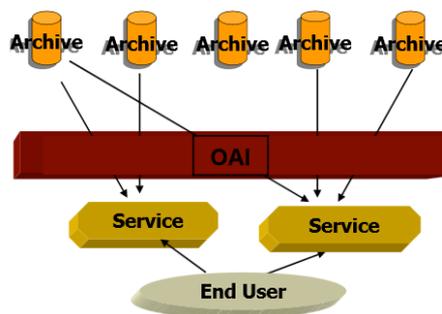
La arquitectura del protocolo define dos participantes en la infraestructura (Van de Sompel y Lagoze, 2008):

- *Data Providers*: los proveedores de datos administran los sistemas que soportan el OAI-PMH como medio para exponer metadatos. Almacenan y preservan los documentos y exponen los metadatos para que sean cosechados por los servicios recolectores.
- *Service Providers*: proveedores de servicios, recolectores o cosechadores de metadatos (*metadata harvesters*). Son herramientas que recolectan metadatos mediante el OAI-PMH como base para construir

servicios de valor añadido. Recogen los metadatos de diferentes repositorios para organizarlos en función de los criterios por los que puedan ser interrogados por los usuarios. Y acceden y muestran al usuario documentos a texto completo alojados en depósitos distribuidos que cumplan el protocolo OAI-PMH.

La ilustración siguiente representa gráficamente la arquitectura del protocolo OAI-PMH, en la que se identifican los proveedores de datos o archivos independientes, y los cosechadores o servicios que mediante el OAI proporcionan la información a los usuarios finales.

Ilustración 7: Gráfico de la arquitectura de consulta de un harvester a múltiples proveedores de datos



Fuente: Peset Mancebo, Fernanda y Ortín Pérez, Alfredo. “Tendencias internacionales en el acceso libre a la documentación científica digital: OAI-Open Archives Initiative”, En: Derecho, historia y universidades: estudios dedicados a Mariano Peset. Oc 978-84-370-6864-0. v ll.978-84-370-6866-4. Universidad de Valencia, 2007pp. 441-452 <http://eprints.rclis.org/6470/1/ortin_y_peset.pdf>; basado en ReDIF Barrueco, José Manuel. Experiencias cooperativas en archivos abiertos en Ciencias Sociales. IV Semana de la ciencia, Madrid, 2004.

El servicio recolector (cosechador) interroga varios de los archivos que exponen metadatos (proveedor de metadatos) mediante el protocolo OAI-PMH. El recolector consulta los datos de algunos de los archivos, no todos (Peset y Ortín, 2005,) para obtener respuestas de sus servidores. El servicio recolector puede solicitar al proveedor de datos que le envíe metadatos según determinados criterios como por ejemplo, la fecha de creación de un documento. En respuesta, el proveedor de datos (archivo en el que se almacena el documento solicitado) devuelve un conjunto de registros en formato XML, incluyendo identificadores (URL por ejemplo) de los objetos descritos en cada registro (Barrueco y Subirats, 2003).

De modo que, mediante este protocolo, los recolectores de metadatos se convierten en herramientas distribuidoras de información a través de una única interfaz de consulta. El cosechamiento de metadatos permite la recuperación de documentos dispersos mientras que el propio objeto digital se encuentra alojado en su repositorio o archivo de origen.

Las características del OAI-PMH que han favorecido su implantación en la mayoría de los repositorios existentes a nivel mundial son (Barrueco y Subirats, 2003; Alonso et al., 2008):

- La arquitectura del modelo recolecta metadatos en lugar de emplear el modelo de búsqueda distribuida como el

protocolo Z39.50 o los sistemas de sindicación de contenidos tipo RSS.

- Interoperabilidad con metadatos y no con los objetos digitales almacenados en archivos distribuidos.
- Uso de metadatos Dublin Core. Los metadatos de descripción de documentos para transmitir vía OAI-PMH deben codificarse en DC sin calificar para minimizar los problemas derivados de las conversiones entre múltiples formatos.
- Es independiente de cualquier aplicación. Generalmente las aplicaciones que se emplean utilizan software libre. Los programas más extendidos son e-prints, Dspace y Fedora.
- Se basa en los estándares normalizados HTTP y XML para la codificación de los metadatos y la transferencia de los datos.

Gracias a estas características han aparecido servicios, siendo los más numerosos los recolectores de los metadatos que residen en múltiples archivos. A continuación listamos los más conocidos por su cantidad de datos o por ser adecuados al objeto de este trabajo, que pueden albergar documentos relacionados con los autores valencianos (Alonso et al., 2008; Barrueco y Subirats 2003).

- OAster recursos académicos digitales. Desarrollado en 2002 por la Biblioteca de la Universidad de Michigan, y financiado por la Fundación Andrew W. Mellon. Es un catálogo colectivo de recursos web que da acceso a una de las mayores colecciones de documentos. En 2009 la gestión pasó a la OCLC. Actualmente cuenta con más de treinta millones de registros aportados por más de 1.500 organizaciones de todo el mundo. Entre las instituciones españolas que contribuyen total o parcialmente con sus repositorios de documentos digitales están los de la Universidad Politécnica de Cataluña, TDX (Tesis Doctorales en Red), RECERCAT, Archivo Digital de la Universidad Politécnica de Madrid, EPrints Complutense, la recientemente creada Biblioteca Virtual del Patrimonio Bibliográfico, Dialnet OAI Artículos y Tesis, Universidad Autónoma de Barcelona, Universidad Autónoma de Madrid, Universidad de Alicante, Universidad de Valencia (OCLC, 2014).

- DRIVER: Digital Repository Infrastructure Vision for European Research. Da acceso a la red de repositorios digitales de acceso libre, que son cosechados de 295 repositorios de 38 países. Alberga publicaciones científicas (artículos de revista, tesis, libros, conferencias, informes) de varias disciplinas académicas. Surge en el marco de las infraestructuras del 7º Programa

Marco de la Comisión Europea, cuyo propósito es proporcionar una infraestructura paneuropea de preservación y difusión de la investigación, para acercar un valioso activo a la industria, y ayudar a acercar la investigación y la educación (DRIVER, 2014).

- Recolecta: recolector de ciencia abierta, es una infraestructura que agrupa todos los repositorios científicos nacionales de acceso abierto. Nace en 2007 de la colaboración entre el Ministerio, FECYT y REBIUN de la CRUE. Sus principales objetivos son impulsar y coordinar la infraestructura, garantizar que los repositorios incluidos sean interoperables según los estándares de la comunidad mundial. Promover, apoyar y facilitar la adopción del acceso abierto por todos los investigadores de las universidades y centros de I+D españoles. Dotar de una mayor visibilidad tanto nacional como internacional a los resultados de la investigación que se realiza en España (Barrueco et al., 2014).
- HISPANA: directorio y recolector de recursos digitales. Hispana tiene cuatro vertientes. Es por un lado un agregador de contenidos, un recolector OAI-PMH, que reúne las colecciones digitales de los repositorios de archivos, bibliotecas y museos españoles, conforme a la Iniciativa de Archivos Abiertos que promueve la Unión

Europea (Martínez Conde, 2012). Entre estas colecciones destacan los repositorios institucionales de las universidades españolas, las bibliotecas digitales de las comunidades autónomas, las bibliotecas de reales academias y el repositorio de museos CER.es de colecciones de museos en red. Ofrece acceso a conjuntos crecientes de todo tipo de materiales (manuscritos, libros impresos, fotografías, mapas) del patrimonio bibliográfico español. Por otro lado es un repositorio OAI-PMH, es también un directorio que recoge los proyectos e iniciativas de digitalización existentes en España, y es un agregador de Europeana: las bibliotecas digitales de las comunidades autónomas, y otras de carácter local, aportan, a través de Hispana, sus contenidos al proyecto *Europeana Linked Open Data* en consonancia con los principios de la web semántica (Martínez Conde, 2012).

Otros aspectos de interoperabilidad más avanzados y más relacionados con la web semántica que veremos en el siguiente capítulo parten también de la Open Archives Initiative. Hasta ahora hemos revisado la difusión de la información científica de forma interoperable, pero no deja de tener sus raíces en la concepción clásica de documento que reúne en un objeto digital

varios tipos de información diferentes (texto, imágenes, tablas). OAI-ORE (Object Reuse and Exchange) trata de la reutilización e intercambio de objetos digitales de forma individualizada, granularizada, definiendo un conjunto de normas para la descripción e intercambio de agregaciones de recursos web entre repositorios. Su objetivo es tratar los recursos que forman una unidad lógica de información. Las colecciones que forman los archivos digitales están compuestas por agregaciones de materiales interrelacionados, denominadas también objetos digitales complejos. Esta agregación de recursos se almacena en ficheros diferentes y en muchas ocasiones están ubicados en distintos servidores, y mediante estas normas es posible tratar estas agregaciones de materiales como un recurso unitario.

OAI-ORE nace en 2006, bajo el amparo de la Fundación Andrew W. Mellon, con el propósito de aumentar la interoperabilidad, pero no únicamente a través de metadatos sino de objetos digitales complejos, que son el conjunto de unidades resultantes de cualquier investigación, es decir, los *datasets* o conjuntos de datos, tales como tablas, textos, software, imágenes, albergados en uno o varios repositorios, pero relacionados de forma consistente como uno solo (Borillo y Gambau, 2008).

La especificación ORE define una estructura que identifica a la agregación y describe las relaciones entre los componentes de ésta, de forma que los usuarios, máquinas y personas, puedan

entender el significado y significante del objeto, y de cómo conecta con otros recursos (Kaplan, 2011). La representación del recurso, en la que se describen sus propiedades y los valores de las mismas, puede ser definida mediante tres formatos, en RDF/XML (*Resource Description Framework*) / (*Extensible Markup Language*) y RDFa siguiendo los principios LD. Cada una de estas unidades que componen un objeto digital tiene una URI asociada para ser referenciado (Kaplan, 2011).

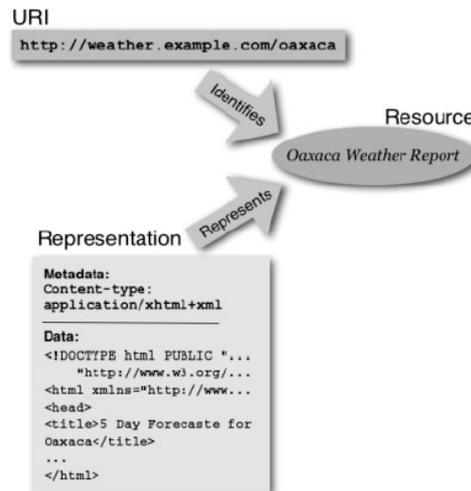
También emplean los estándares RSS y Atom, que son formatos XML de redifusión de información. Se trata de dos especificaciones para la descripción de contenidos sindicados, es decir, que permiten la distribución de contenidos como texto, imagen, audio, video, y de sus metadatos descriptivos. En este sentido existe similitud con OAI-ORE porque trabaja con la agregación y con cada uno de los elementos agregados (Orduña-Malea, 2009).

Tal como Orduña-Malea (2009) expresa, OAI-ORE representa el dominio de una agregación a través de un grafo que está formado por nodos (los recursos) y arcos (las relaciones). Define la existencia de un nodo principal llamado “mapa de recurso” que describe al nodo que representa a la agregación en sí misma. Este mapa de recurso se identifica con una URI (URI-R) que da acceso a través de una HTTP, a un fichero donde, en un determinado esquema de metadatos, se proporcione a la persona/máquina todos los recursos y relaciones existentes en el

grafo. Es decir, donde se explicita el dominio de la agregación de forma estructurada.

La estructura del mapa del recurso está formada por el propio recurso, la representación y la URI como muestra la siguiente imagen.

Ilustración 8: Relación entre identificador, recurso y representación.



Fuente: Borillo Domenech, Ricardo y Gumbau Mezquita, José Pascual. (2008) *Arquitecturas orientadas a servicios para la preservación de objetos digitales*.

Como vemos, existe una similitud entre los proyectos OAI-ORE y OAI-PMH con los principios LD del proyecto de la web semántica en cuanto a la idea de transferencia de los contenidos y a los elementos que los constituyen y definen. Por un lado por la importancia que tiene el uso de URI en sus estructuras de definición. Aunque OAI-PMH y ORE las emplean con propósito

de identificación de objetos digitales, los principios LD las utilizan también con el objetivo de desreferenciar (Haslhofer, 2009). Lo que posibilita que mediante el protocolo de transferencia HTTP se pueda acceder a la definición de la relación que la URI define entre sujeto y objeto en la estructura de descripción de un objeto digital que el lenguaje RDF establece, como se verá en el siguiente capítulo. También OAI-PMH utiliza transacciones HTTP, para la transferencia de contenidos web, en la que está definida sintaxis y semántica, al igual que uno de los principios básico de los datos enlazados (Heath y Bizer, 2011).

El objetivo en los dos proyectos OAI, PMH y ORE, es en definitiva la transferencia de contenidos de una forma estable y normalizada. Para alcanzarlo el elemento clave es la interoperabilidad entre sistemas de gestión y almacenamiento de datos ya que si existe una incompatibilidad entre las normas técnicas que describen los datos se imposibilitaría esta transferencia. Por tanto, la interoperabilidad está encaminada al intercambio estandarizado de documentos de diferentes formatos, alojados en fuentes distintas, especialmente en repositorios y bibliotecas. En este sentido la idea de la web semántica y su definición de datos enlazados es muy próxima, y aunque se desarrolla con otras tecnologías, la interoperabilidad es un elemento clave. Además, para posibilitar la recuperación

de la información a intercambiar, tiene que estar descrita en un lenguaje que sea compartido por los sistemas que interoperan entre sí, razón por la que Vestigium había de escoger entre acercarse a los postulados del OAI, todavía en experimentación y más cercanos a la comunidad bibliotecaria, o los de la web semántica, que como esta tesis demuestra, tienen una implantación con más futuro para todo tipo de contenidos web. De hecho, en un entorno web, los elementos que describen la información son los metadatos, que constituyen una forma segura y estandarizada de comunicar y entender la información, razón por la que en el siguiente apartado se definen profusamente.

Los esquemas de metadatos para bibliotecas

Los metadatos constituyen un componente importante en la descripción de los contenidos que preserva cualquier biblioteca, archivo y repositorio digital (Alemu, 2011; Chan y Zeng, 2006; NISO, 2004). Como hemos dicho anteriormente, en el proceso de catalogación, los metadatos son el elemento clave en la descripción bibliográfica. Suponen la evolución de ésta hacia el mundo digital, por cuya razón resulta tan decisivo el conocimiento acumulado por la profesión vinculada a bibliotecas, archivos y museos sobre el tratamiento de la información. No obstante, existen diferencias ya que recogen, por ejemplo, aspectos de propiedad intelectual o de preservación que no son tradicionalmente recogidos en los catálogos de bibliotecas. Además, tras unos cuantos años de asignación de metadatos a modo de registro bibliográfico de los documentos web, las etiquetas han evolucionado hacia otros usos, como puede ser la web semántica que veremos en el siguiente capítulo.

Los metadatos, en suma, son un conjunto de elementos que aportan información sobre los recursos digitales y no digitales en un sistema del entorno web. Según la ISO/IEC 11179 *Information Technology – Metadata Registries* (MDR), metadato se define como un dato descriptivo sobre un objeto. El objeto puede ser de cualquier tipo, como libros, documentos,

bases de datos y datos en sí mismos.

En los años sesenta se acuña este término, y desde entonces son numerosas las definiciones que han dado los autores (Méndez y Senso, 2004). Cada autor define los metadatos de una forma particular, desde Caplan (1995), Xu (1997), Younger (1997) Dempsey y Heery (1998), Tennant (1998), Milstead y Feldman (1999), Wendler (2000), Gorma (2000) a Taylor (2004). Aunque todas las definiciones coinciden en la idea de: información estructurada que describe objetos digitales (Angelozzi y Martín, 2007). Nosotros seleccionamos dos con las que se aclara el concepto:

Para Juárez Santamaría (2007) son un conjunto de elementos que se utilizan para ayudar a la identificación, descripción y localización de los recursos electrónicos por medio de una representación de la descripción bibliográfica de los mismos.

Para Daudinot (2006) los metadatos constituyen instrumentos para describir el contenido semántico de un recurso y están mejor preparados para soportar la recuperación de información que el propio documento, ya que los recursos de información no son capaces de facilitar por sí mismos sus propias relaciones semánticas si no disponen de una caracterización (metadatos) que los describa. Los datos requieren información auxiliar, “datos sobre datos”, que describan la información relativa al recurso. En definitiva, aportan información de contexto.

Como decíamos, los beneficios de la catalogación tradicional se ven reflejados en el ámbito digital con la asignación de metadatos. Los mayores esfuerzos se centran en proporcionar una descripción exhaustiva de los recursos para certificar que seguirán siendo accesibles en el futuro. Entre las ventajas cabe mencionar: asegurar su preservación, mejorar los resultados de las búsquedas de información en la web, el descubrimiento de recursos (objetos y sus objetos relacionados), facilitar la cooperación y el compartido de contenidos a través de un continuo intercambio entre instituciones favoreciendo la interoperabilidad entre espacios de preservación de información académica y científica, que permite ofrecer resultados más completos a las búsquedas realizadas en sus sistemas.

Las características comunes de todos los metadatos, independientemente del tipo que sean, y que definen el diseño de cualquier esquema de metadatos son las siguientes (Duval et al., 2002):

- Modularidad: es un principio que permite a los desarrolladores de esquemas de metadatos crear nuevos conjuntos a partir de combinar los elementos de datos provenientes de otros esquemas diferentes de una manera sintácticamente y semánticamente interoperable.
- Reutilización: los sistemas de metadatos deben permitir añadir calificadores para especificar el significado de los

elementos de acuerdo con el nivel de detalle necesario.

- Extensibilidad: deben permitir extensiones para adaptarse a las necesidades particulares de una determinada aplicación. Algunos elementos son comunes en la mayoría de esquemas de metadatos, pero otros serán específicos de dominios particulares.
- Multilingüismo: deben permitir adaptaciones a distintos idiomas.

Asignar metadatos es incorporar etiquetas textuales al objeto digital, legibles por ordenador y con información adicional para delimitar características del mismo. La información digital puede ser modificada y estas etiquetas ayudan a registrar los cambios producidos a lo largo de la historia del documento. Por ejemplo, la identificación del objeto digital mediante una etiqueta de identificador persistente (URL, DOI), a modo de ISBN en los libros, o por ejemplo, etiquetas de fecha de creación del documento, de autoría, de registro de los cambios de ubicación, las revisiones que ha recibido (Tummarello, 2008). Toda esta tipología de etiquetas ha sido clasificada por diversos autores. A continuación exponemos dos clasificaciones:

Lazinger (2001) y Elba y Caballero (2009) los clasifican en tres categorías de acuerdo con la naturaleza de los datos que describen.

- Metadatos descriptivos: describen información bibliográfica. Son aquellos que describen e identifican los recursos de información, facilitan la búsqueda y recuperación. Los que aplican los estándares Dublin Core, MARC, etiquetas meta de HTML.
- Metadatos estructurales: referidos a formatos y estructuras. Son aquellos que proporcionan la información sobre la estructura interna de los recursos y la relación entre sus elementos o archivos. Contribuyen a facilitar la navegación y la presentación. Los aplican los estándares SGML, XML/RDF y EAD.
- Metadatos administrativos o de conservación: son aquellos que aportan información relativa a la gestión de derechos, uso del contenido digital, información sobre la autenticidad, historia de la custodia del recurso, estructura y características técnicas.

Anne J. Gilliland (2008) establece cinco tipologías:

- Metadatos administrativos: empleados en la gestión y administración de colecciones y recursos de información. Por ejemplo, información de localización, criterios de selección para la digitalización, acceso legal a los documentos, derechos y reproducción, información de adquisición.

- Metadatos descriptivos: para identificar y describir colecciones y recursos de información relacionados. Por ejemplo, para diferenciar entre versiones de documentos, anotaciones sobre los creadores, relaciones de hiperenlace entre recursos relacionados.
- Metadatos de preservación: relacionados con la gestión de la preservación de las colecciones y recursos de información. Información relacionada con las condiciones físicas del documento, documentación de la preservación de las versiones físicas y digitales de los recursos por ejemplo, fecha de actualización y de migración, las versiones de los documentos, documentación relacionada con los cambios en el documento durante la digitalización y la preservación.
- Metadatos técnicos: relacionados con el comportamiento y las funciones del sistema, por ejemplo, información sobre *hardware* y *software*, información técnica de la digitalización, formatos, compresión, autenticación y seguridad de los datos, claves de encriptación, claves y contraseñas.
- Metadatos de uso: relacionados con el nivel y tipo de uso de las colecciones y los recursos, por ejemplo, los registros de circulación, registros de las versiones físicas y digitales, reutilización del contenido.

El uso de metadatos ha conducido a su ampliación, y a la creación de nuevos metadatos para describir las características de los objetos digitales. Los esquemas de metadatos se han ampliado, y actualmente existen numerosos esquemas de descripción y tipologías diversas para adaptarse a las características de los recursos que describen (Vaishnav y Sonwane, 2007). Generalmente los esquemas de metadatos especifican los nombres de los elementos, proveen de contexto a los metadatos definiendo la estructura de relaciones entre los elementos y a veces caracterizando las cosas a las que los elementos hacen referencia (Baker et al., 2003). La situación ideal en un ámbito dado, por ejemplo el sector de las bibliotecas digitales, podría ser adoptar un sólo estándar que proporcione un intercambio de datos y una interoperabilidad eficiente (Hillman, 2010). Pero los esquemas de metadatos permiten el uso conjunto de más de un esquema, lo que produce situaciones en que coexisten varios esquemas y se utilizan en combinación. Por ejemplo, un registro de la base de datos AGRIS (International System for Agricultural Science and Technology) de FAO (Food and Agriculture Organization) de United Nations (UN) que se muestra en la ilustración 9, puede visualizarse en RDF, como observa la ilustración 10, la cual contiene codificación RDF/XML que reúne a un tiempo los esquemas de metadatos DC, FOAF, RDF, como reflejan las cabeceras del código que le continua.

Ilustración 9: Registro de la base de datos AGRIS (FAO)

The screenshot shows a web browser window with the URL <http://agris.fao.org/aos/records/ES8600299>. The page title is "Lupin (lupinus): Its composition and development as a source of proteins and oils, 3: Obtention of protein isolated and oil with low alkaloid content" at AGRIS RECORDS. The page features a table with the following data:

Property	Value
abstract	In this paper the best conditions to obtain proteinic isolates from lupin meal are studied, so as the extraction, refining and debittering of its oil. ... >more> (eng)
creator	[3 anonymous resources]
dateSubmitted	1986
description	6 graphs, 4 numerical tables; 13 ref. Summaries (En, Es)
identifier	ES8600299
isPartOf	<file:///home/fao/transfer/2013-11-24-backup/agris-new/http%3A%2F%2Faims.fao.org%2Fseries%2Fc_c907061a>
issue	4
issued	jul-aug1985
language	esp
source	<http://ring.ciard.net/node/10818>
subject	<http://aims.fao.org/aos/agrovoc/c_10770> <http://aims.fao.org/aos/agrovoc/c_10961> <http://aims.fao.org/aos/agrovoc/c_25005> <http://aims.fao.org/aos/agrovoc/c_37969>
title	Lupin (lupinus): Its composition and development as a source of proteins and oils, 3: Obtention of protein isolated and oil with low alkaloid content (eng)
type	Numerical data Summary
volume	36

The AGRIS records dataset is maintained by FAO. The JAD LOD services are a collaboration project with the MIMOS Bhd. [As RDF/XML](#)

Fuente: AGRIS (FAO) <agris.fao.org>.

Ilustración 10: Visualización en RDF/XML de un registro de la base de datos AGRIS (FAO)

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:j.1="http://purl.org/net/provenance/types#"
  xmlns:j.0="http://purl.org/ontology/bibo/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:j.2="http://spinrdf.org/sp#"
  xmlns:j.3="http://purl.org/dc/terms/"
  xmlns:j.4="http://www.w3.org/2004/03/trix/rdfg-1/"
  xmlns:j.5="http://purl.org/net/provenance/ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <j.0:Article rdf:about="http://agris.fao.org/aos/records/ES8600299">
  <j.3:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_37969"/>
  <j.0:volume> 36</j.0:volume>
  <j.3:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_10770"/>
  <j.3:creator>
  <foaf:Person>
  <rdfs:seeAlso rdf:parseType="Resource">
  </rdfs:seeAlso>
  <foaf:name>Arenas Angulo, P. (Consejo Superior de Investigaciones
  Cientificas, Sevilla (Spain). Instituto de la Grasa y sus
  Derivados)</foaf:name>
  </foaf:Person>
  </j.3:creator>
  <j.3:source rdf:resource="http://ring.ciard.net/node/10818"/>
  <j.3:title xml:lang="eng">Lupin (lupinus): Its composition and development
  as a source of proteins and oils, 3: Obtention of protein isolated and oil
  with low alkaloid content</j.3:title>
```

```

[...]  

<rdf:Description  

  rdf:about="http://agris.fao.org/aos/data/ES8600299?output=xml">  

  <rdfs:label>RDF description of ES8600299</rdfs:label>  

  <foaf:primaryTopic  

  rdf:resource="http://agris.fao.org/aos/records/ES8600299"/>  

</rdf:Description>  

<j.4:Graph rdf:nodeID="A0">  

<j.5:createdBy>  

<j.5:DataCreation>  

<j.5:usedData>  

<j.4:Graph>  

<j.5:createdBy>  

<j.1:QueryExecution>  

<j.5:usedGuideline>  

<j.2:Describe>  

<rdfs:label  

  rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DESCRIBE  

  &lt;http://agris.fao.org/aos/records/ES8600299&gt;</rdfs:label>  

  <rdf:type  

  rdf:resource="http://purl.org/net/provenance/types#SPARQLQuery"/>  

[...]  

<foaf:primaryTopic  

  rdf:resource="http://agris.fao.org/aos/records/ES8600299"/>  

</j.4:Graph>  

</rdf:RDF>

```

Fuente: AGRIS (FAO) <agris.fao.org>.

Como consecuencia del uso de dos o más estándares se han detectado algunos problemas, por ejemplo, relacionados con la semántica: podrían emplearse diferentes etiquetas para conceptos similares, como por ejemplo, “nombre” “autor” “compositor” para referirse a la persona o institución responsable de la creación intelectual del trabajo (Miller, 2000; Alemu, 2011). De modo que, para asegurar una continua y eficaz comunicación de información entre diferentes sistemas de preservación, se han elaborado una serie de soluciones para mantener una efectiva interoperabilidad en estas circunstancias (Alemu, 2012; NISO 2004; Chan and Zeng, 2006).

- El uso de metadatos derivados (*metadata derivation*): consiste en desarrollar un nuevo esquema a partir del que

ya existe. Se trata de adaptar un estándar existente a unas necesidades de descripción específicas, por ejemplo, el esquema MARC, el más común en ámbito bibliotecario, adaptado al ámbito web es MARC XML, MARC-Lite o MODS.

- Perfiles de aplicación de metadatos (*metadata application profiles*), conocido también como *Mixing and Matching Metadata Schemas*: es el conjunto de los elementos de metadatos, las políticas y las guías definidas para una aplicación concreta. Los perfiles de aplicación se define como esquemas que consisten en elementos de datos de uno o más conjuntos, combinados en conjunto e implementados para una aplicación particular (DCMI, 2004).
- Mapeo o conversión de metadatos (*Metadata-cross walks o mapping*): son equivalencias entre los elementos de unos formatos de metadatos y otros. St. Pierre y LaPlant (1998) lo definen como un conjunto de transformaciones aplicadas al contenido de los elementos de un estándar de metadatos que se traducen en el almacenamiento de contenido, convenientemente modificado, en los elementos análogos de un estándar de metadatos de destino.

Estas conversiones, son mapas que establecen las

equivalencias de los elementos, semántica y sintaxis, entre varios esquemas de metadatos (NISO, 2004). Los *crosswalks* son importantes en las colecciones virtuales compuestas por gran variedad de fuentes. Pero el mapeado de esquemas con pocos elementos (baja granularidad) a otro con muchos elementos (gran granularidad) es complicado. Así también, el mapeado de un esquema de gran granularidad a uno de baja entraña dificultades ya que los campos que no tienen equivalencia se pierden. Por ejemplo, en el caso de MARC a Dublin Core. La siguiente tabla presenta un ejemplo de equivalencia de elementos de metadatos de los esquemas DC, EAD y MARC 21.

Tabla 4: Ejemplo de *crosswalk* de metadatos DC, EAD y MARC 21

	Dublin Core	EAD	MARC 21
Elemento Título	Title	<titleproper>	245 00\$a (Título)
Elemento Autor	Creator	<autor>	700 1#\$a (Nombre personal) 720 \$a (Nombres no controlados)
Elemento Fecha de Creación	Date.Created	<unideate>	260 ##\$c (Fecha de publicación, distribución)

Fuente: Library of Congress, 2007
<<http://www.loc.gov/standards/mads/mads-mapping.html>>

- Los registros de metadatos: recogen varios esquemas de metadatos y su utilización consiste en elegir elementos de diferentes esquemas para construir aplicaciones que se adapten a las características de los recursos digitales.
- El uso de tecnologías de la web semántica (Chan y Zeng, 2006; NISO, 2004): es otra solución que respaldan varios autores para alcanzar la interoperabilidad semántica. En este sentido el W3C ha desarrollado el marco de descripción de recursos RDF, un modelo para la descripción de recursos en la web que proporciona un mecanismo de integración de múltiples esquemas de metadatos. Permite combinar elementos de diferentes esquemas en una sola descripción de recurso.

Por último, ante la proliferación de esquemas de metadatos intentaremos aclarar la situación en el sector bibliotecario proponiendo una clasificación para la presente tesis que refunda las de Miller (2011), Hillman (2010) y Gilliland (2008) sobre la tipología de los estándares de datos y las normas de catalogación:

- Estándares de estructuras de datos: son categorías o contenedores de datos que permiten describir un registro de un objeto digital como conjuntos de elementos o esquemas (*element sets; schemes, schemas, o schemata*). Por ejemplo, Dublin Core, MODS, LOM.

- Estándares de contenido de los datos: son guías sobre los formatos y sintaxis de los datos. A lo largo del tiempo se denominaron reglas de catalogación o de buenas prácticas. En ocasiones estas recomendaciones son recogidas en los estándares de estructuras de datos. Por ejemplo, AACR, RDA, ISBD.
- Estándares de valores de datos: son términos, nombres y otros valores como vocabularios controlados y esquemas de codificación (*controlled vocabularies, encoding schemes*). Por ejemplo, la lista de encabezamiento de materias de la LC (*Library of Congress Subject Headings* (LCSH)), Archivo de Autoridad de Nombres de la LC *Library of Congress Name Authority File* (LCNAF), *LC Thesaurus for Graphic Materials* (TGM), *Medical Subject Headings* (MeSH), *Art & Architecture Thesaurus* (AAT), *Getty Thesaurus of Geographic Names* (TGN).
- Formato de datos / estándares técnicos de intercambio: es una estructura de datos de codificación particular procesable por máquina e intercambiable. Obviamente, contienen en ocasiones los estándares de descripción mencionados en el primer grupo, pues mencionan esas estructuras de metadatos. Por ejemplo, MARC, MARCXML, XML, SGML, *Simple Dublin Core XML*

schema, Qualified Dublin Core XML schema).

- Estándares para la presentación de datos, donde de nuevo vemos repetirse algunos ejemplos anteriores: ISBD, *CSS/XSLT for display*, *OPAC display settings*.

Como se ha dicho, el uso de metadatos para describir recursos se emplea en la comunidad bibliotecaria desde hace décadas, por ejemplo, los registros bibliográficos MARC lo implementan desde 1968 en los catálogos digitales de bibliotecas de acceso público (*Open Public Access Catalogues*, OPAC), extendiéndose con la consolidación de la automatización de bibliotecas en los años setenta (Ortiz-Repiso, 1999). En la década de los años noventa se produjo una eclosión de nuevos repositorios distribuidos en Internet y aumentó el número de fuentes digitales y de documentos electrónicos en distintos formatos. Ante este crecimiento de recursos surge la necesidad de adecuar los estándares de catalogación existentes a la nueva variedad de contenidos. Es necesario describir, autenticar y gestionar estos recursos y la tendencia en el mundo bibliotecario siempre ha caminado hacia la estandarización. Los organismos internacionales han ido consensuando una serie de adaptaciones de las normas y las herramientas de descripción bibliográfica existentes hasta incluir los nuevos recursos digitales cuya evolución revisamos.

Machine Readable Cataloguing (MARC)

El Formato MARC (*Machine Readable Cataloguing*) fue desarrollado en los años sesenta por la Network Development and MARC Standards Office de la Library of Congress y la Library and Archives de Canadá, como se ha visto. Debido al interés que la LC tenía en facilitar la edición de su bibliografía nacional, en 1963 publicó un informe *Automation and the Library of Congress* donde recomendaba automatizar la catalogación, para lo que en 1995 creó un equipo de trabajo liderado por Henriette Avram, de formación matemática que procedía de la National Security Agency US (Martín Gavilán, 2008).

En 1966 la LC ponía en marcha la catalogación automatizada con el formato MARC. El objetivo era codificar información bibliográfica y relacionada de documentos impresos, monografías, de forma legible por las computadoras, para el intercambio de registros bibliográficos entre centros catalogadores. Posteriormente fue incorporando campos para publicaciones seriadas, archivos de computadora, mapas, música, registros sonoros, material audiovisual y recursos electrónicos (Angelozzi y Martín, 2007).

Su implantación se produjo en primer lugar en la LC junto con el grupo de bibliotecas americanas, y en 1968 inició la

distribución de registros bibliográficos a otras bibliotecas nacionales. LC cooperaba también con la British Library (BL) y como resultado en 1968 apareció MARC II y en 1970 se ponía en marcha el UKMARC para la producción de la BL. Entre 1973 y 1977 aparecieron numerosas versiones de carácter nacional como USMARC (Martín Gavilán, 2008). Su rápida expansión en los años setenta facilitó la comunicación de registros catalográficos fiables entre las bibliotecas de todo el mundo. Se convirtió en modelo de referencia. Estas normas no dan pautas de almacenamiento interno ni de visualización, sino que definen la estructura para registrar los datos bibliográficos y datos relacionados con el objetivo de facilitar la comunicación de la información entre sistemas. Esta facilidad de adaptación a las normas catalográficas nacionales derivó en el surgimiento de versiones como el CANMARC para Canadá, USMARK (EEUU), el IBERMARC (España), AUSMARC (Austria), INTERMARC (Francia), ANNAMARC (Italia), DMARC (Alemania), MARCAL (América Latina), JAPANMARC (Japón), SAMARC (Sudáfrica) (Martín Gavilán, 2008). La existencia de tantas versiones dificultaba el intercambio internacional de los registros por lo que la IFLA desarrolló en 1977 el UNIMARC, con el objetivo de la aplicación de un único formato universal. Como este objetivo no fue alcanzado, en 1999 USMARK, CANMARC y UKMARC fueron armonizados en MARC 21 y poco a poco las bibliotecas empezaron a migrar

sus catálogos a este formato unificado. Como es el caso del IBERMARC español (Martín Gavilán, 2008; Estivill, 2012).

Convertido en referente de catalogación, ha sido adaptado para la descripción de materiales como música impresa, mapas, publicaciones seriadas, recogidos en cinco grupos según el tipo de contenido que describen: *MARC 21 Format for Bibliographic Data* para datos bibliográficos, *MARC 21 Format for Authority Data* para registros de autoridad, *MARC 21 Format for Holdings Data* para registros de fondos y localizaciones, *MARC 21 Format for Classification Data* registros de clasificación y *MARC 21 Format for Community Information* para información de la comunidad. Por lo que se ha consolidado como estándar de la comunicación de la información bibliográfica normalizada (Martín Gavilán, 2008).

Un registro bibliográfico MARC está compuesto por tres elementos (Martín Gavilán, 2008):

- La estructura: regulada por tres normas internacionales la ISO 2709: *Information and documentation - Format for Information Exchange*, que reemplaza a la anterior ANSI Z39.2 *Information Interchange Format*, la ISO 6630: *Documentation - Bibliographic control characters*, y la ISO 1001: *Information Processing - File structure and labelling of magnetic tapes for information interchange*.

Consta de tres componentes: la cabecera, que contiene los datos de control para el procesamiento automático del registro. El directorio, que es equiparable al índice del registro. Y los campos variables, que son las etiquetas que organizan los datos del registro.

- La designación de contenido: son identificadores de campos de contenido. Son etiquetas, códigos de definición de campos y subcampos de contenido.
- El contenido de los datos: el contenido del registro, es decir, la información bibliográfica está definida por las normas de catalogación nacionales, los sistemas de clasificación y tesauros de materias y otras convenciones utilizadas por la institución que genera el registro, que en España han tenido su concreción en las Reglas de Catalogación, los Encabezamientos de Materia y la Clasificación Decimal Universal.

La siguiente imagen muestra un registro bibliográfico en formato MARC procedente de la BNE. En el que se distingue como encabezamientos principales el título y el autor, y los campos de descripción codificados con los códigos del formato MARC. Por ejemplo, los códigos 001 y 005 son campos de control, el 016 de la agencia bibliografía, el 020 el ISBN, el 245

de mención de título, el 650 la materia, el 700 de asiento secundario-nombre personal.

Ilustración 11: Visualización del registro en formato MARC

```
Versionatura/s | Registro del catálogo
Biomateriales [ Texto impreso ] : repuestos para el cuerpo humano : discurso de la académica electa Excm. Sra.
Dª. María Vallet Regí leído, en el acto de su recepción pública el día 18 de febrero de 2004, y contestación del
académico Excmo. Sr. D. Avelino Corma Canós Cambiar visualización
Vallet Regí, María
000      am 0n a
001      bimo0001933917
005      20071016
008      040414s2004 sp l III l spa
016 7    bimoBNE20040351829|2SpMaBN
017      M 525-2004|bOficina Depósito Legal Madrid
020      84-95662-21-3
035      (OCoLC)433577211
040      SpMaBN|bspa|cSpMaBN|erdc
080 0    615.46
100 1    Vallet Regí, María
245 10   Biomateriales|h[Texto impreso]|b : repuestos para el cuerpo humano : discurso de la académica electa
Excm. Sra. Dª. María Vallet Regí leído, en el acto de su recepción pública el día 18 de febrero de 2004, y
contestación del académico Excmo. Sr. D. Avelino Corma Canós
260      Madrid|bReal Academia de Ingeniería|c2004
300      36 p.|c24 cm
504      Bibliografía: p. 28-29
650 7    Materiales biomédicos|2embne
700 1    Corma Canós, Avelino
956      2
980      |.017. |aM 525-2004|.020. |a84-95662-21-3|.852. 40|aM-
BN|bRNALCALA|9MO04290836|j12/232415|xDG|cRICALA
980      |.017. |aM 525-2004|.020. |a84-95662-21-3|.852. 40|aM-BN|9MO04290541|xAH|jAHMo/65608
980      |.017. |aM 525-2004|.020. |a84-95662-21-3|.852. 40|aM-
BN|bRNALCALA|9MO04292224|jDL/1293866|xDG|cCONSERVACI
```

Fuente: Biblioteca Nacional Española

A continuación, en la ilustración 12 se puede ver la diferencia entre la visualización en Formato MARC 21 de catalogación (ilustración 11) y la visualización del registro para los usuarios que consultan el OPAC, esta última sin códigos alfanuméricos y de más fácil lectura.

Ilustración 12: Visualización del registro en formato de ficha para usuarios

Ver signatura/s	Registro del catálogo
<p>Biomateriales [Texto impreso] : repuestos para el cuerpo humano : discurso de la académica electa Excmo. Sra. D^a. María Vallet Regí leído, en el acto de su recepción pública el día 18 de febrero de 2004, y contestación del académico Excmo. Sr. D. Avelino Corma Canós Cambiar visualización</p> <p>Vallet Regí, María</p> <p>N.º depósito legal M 525-2004 Oficina Depósito Legal Madrid</p> <p>ISBN 84-95662-21-3</p> <p>CDU 615.46</p> <p>Autor personal Vallet Regí, María</p> <p>Título Biomateriales [Texto impreso] : repuestos para el cuerpo humano : discurso de la académica electa Excmo. Sra. D^a. María Vallet Regí leído, en el acto de su recepción pública el día 18 de febrero de 2004, y contestación del académico Excmo. Sr. D. Avelino Corma Canós</p> <p>Publicación Madrid : Real Academia de Ingeniería, 2004</p> <p>Descripción física 36 p. ; 24 cm</p> <p>Bibliografía Bibliografía: p. 28-29</p> <p>Encabez. materia Materiales biomédicos</p> <p>Autor Corma Canós, Avelino</p> <p>Enlace a Dialnet</p>	

Fuente: Biblioteca Nacional Española

En los años noventa la LC adaptó el MARC 21 para datos bibliográficos y el MARC 21 de fondos y localizaciones, a la descripción de materiales digitales mediante la ampliación de un nuevo campo, el 856 de Localización y acceso electrónicos, para describir y localizar recursos electrónicos como Correo electrónico, FTP, Conexión remota Telnet, llamada telefónica y HTTP. La información de descripción de estos recursos se recoge en los siguientes subcampos de la tabla 5.

Tabla 5: Códigos de subcampo del campo 856 - Localización y acceso electrónicos

\$a	Nombre del host (R)	\$q	Tipo de formato electrónico (NR)
\$b	Número de acceso (R)	\$r	Parámetros (NR)
\$c	Información sobre la compresión(R)	\$s	Tamaño del fichero (R)
\$d	Ruta (R)	\$t	Emulación del terminal (R)
\$f	Nombre electrónico (R)	\$u	Id. Uniforme del Recurso (URI)(R)
\$h	Procesador de la petición (NR)	\$v	Horario en que el acceso está disponible (R)
\$i	Instrucción (R)	\$w	Número de control del registro (R)
\$j	Bits por segundo (NR)	\$x	Nota no pública (R)

\$k	Contraseña (NR)	\$y	Texto del enlace (R)
\$l (NR)	Proceso de conexión (<i>Logon</i>)	\$z	Nota pública (R)
\$m (R)	Contacto para facilitar el acceso	\$2	Método de acceso (NR)
\$n (NR)	Nombre de la localización del host	\$3	Especificación de materiales (NR)
\$o	Sistema operativo (NR)	\$6	Enlace (NR)
\$p	Puerto (NR)	\$8 (R)	Enlace entre campos y núm de secuencia

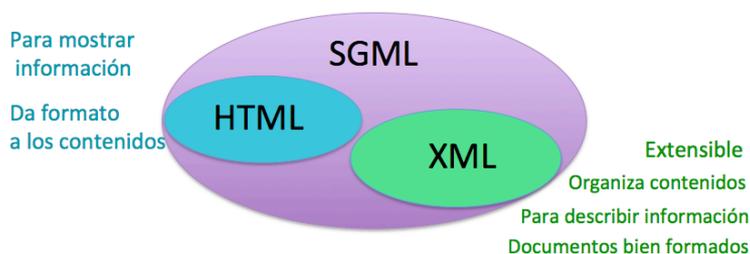
Fuente: Biblioteca Nacional Española

El conjunto de formatos MARC ha marcado toda una época de proliferación de versiones hasta que se consiguió el modelo unificado MARC 21 en 1999, como hemos visto. La adopción de estos formatos para el análisis documental automatizado en gran parte de las bibliotecas y centros catalogadores de la mayoría de los países ha permitido el intercambio internacional. Constituye el primer mecanismo de codificación e intercambio de información automatizado a escala mundial en un sector determinado.

Mientras tanto, en el ámbito informático se producía el desarrollo y la estandarización de los lenguajes de marcas con el mismo propósito. En 1969 el desarrollo de SGM (*Standard Generalized Markup*), en 1980 el primer borrador de SGML (*Standard Generalized Markup Language*), que en 1986 se convirtió en norma ISO. En 1990 HTML, en 1996 XML (*eXtensible Markup Language*) del World Wide Web Consortium, como veremos en el siguiente capítulo dedicado a

la web (Peis y Moya, 2000). La ilustración 13 refleja las principales diferencias entre los lenguajes HTML y XML.

Ilustración 13: Diferencias de los lenguajes de marcas HTML y XML



Fuente: Elaboración propia basada en Peis Eduardo; Moya, Félix de. “Sgml y servicios de información”. En: *El profesional de la información*, 2000, junio, v.9,n.6,pp.4-17.

Ante estos cambios la Oficina de Desarrollo de Redes y Normas MARC de la LC llevó a cabo varias iniciativas para adaptar MARC a los lenguajes de marcas a través de un marco de trabajo que cuenta con numerosos componentes. Así, en 1995 desarrolló dos DTD: Definición de Tipo de Documento (*Document Type Definition*). La DTD es una descripción de la estructura y la sintaxis de un documento XML o SGML. Su función básica es la descripción del formato de datos. Es la manera de decir cuál es la sintaxis correcta del lenguaje. La LC desarrolló las DTD para la conversión de datos del registro MARC a SGML y viceversa sin pérdida de datos. *MARC Bibliographic DTD* para registros bibliográficos, de información

de la comunidad y de fondos; y *MARC Authority DTD* para registros de autoridad y clasificación. Posteriormente la tecnología SGML cambió a XML, y en 2001 las DTDs SGML se convirtieron a DTDs XML (Nogales Flores, 2004; LC, 2014).

A continuación, se muestra un extracto de un ejemplo de marcado con MARC Bibliographic DTD (XML), que recoge todos los campos de descripción del registro MARC en un documento XML con las etiquetas equivalentes a los campos bibliográficos. En este sentido observamos: los elementos raíz son <mrcbfile> (bibliográfica + fondos + información de la comunidad). En la cabecera un elemento hijo del elemento raíz para cada formato MARC:

- *bibliographic* <mrcbldr-bd>,
- *holdings* <mrcbldr-hd>,
- *community information* <mrcbldr-ci>,
- *authority* <mrcaldr-ad>,
- *classification* <mrcaldr-cd>.

En nuestro ejemplo es <mrcbldr-bd> bibliográfico. A continuación se etiquetan en XML las 24 primeras posiciones del formato MARC (00-23) de la cabecera con los elementos de información para el procesamiento del registro, del mismo modo que el registro MARC. A continuación, los campos de control del formato MARC son los códigos 001 al 005, que etiquetado en XML es:

```

<mrcb001>###89048230</mrcb001>
<mrcb003>DLC</mrcb003>
<mrcb005>19911106082810.9</mrcb005>

```

Seguidamente el campo 008 de definición del tipo de material, en el ejemplo es bk para libros, y del mismo modo, el resto de campos del formato MARC se van etiquetando en XML.

La ilustración 14 presenta un extracto de un registro bibliográfico descrito con formato MARC DTD (XML).

Ilustración 14: Extracto de un ejemplo de Marcado con MARC Bibliographic DTD (XML), desde la cabecera hasta el campo de control 008

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE mrcbfile SYSTEM "mrcbxmfile.dtd">
<mrcbfile>
  <mrcb format-type="bd">
    <mrcbldr-bd>
      <mrcbldr-bd-05 value="c"/>
      <mrcbldr-bd-06 value="a"/>
      <mrcbldr-bd-07 value="m"/>
      <mrcbldr-bd-08 value="blank"/>
      <mrcbldr-bd-09 value="blank"/>

      <mrcbldr-bd-17 value="blank"/>
      <mrcbldr-bd-18 value="a"/>
      <mrcbldr-bd-19 value="blank"/>
    </mrcbldr-bd>
    <mrcb-control-fields>
      <mrcb001>###89048230</mrcb001>
      <mrcb003>DLC</mrcb003>
      <mrcb005>19911106082810.9</mrcb005>
      <mrcb008-bk name="FIXED-LENGTH DATA ELEMENTS--BOOKS">
        <mrcb008-bk-00-05 name="Date entered on file" value="891101"/>
        <mrcb008-bk-06 name="Type of date/publication status" value="s"/>
        <mrcb008-bk-07-10 name="Date 1" value="1990"/>
        <mrcb008-bk-11-14 name="Date 2" value=""/>
        <mrcb008-bk-15-17 name="Place of publication, production, or execution" value="mau"/>
        <mrcb008-bk-18-21 name="Illustrations" value="a"/>

```

Fuente: Nogales Flores, Tomás. MARC en XML. *Universidad Carlos III de Madrid*, 2004

En 2002 MARC XML DTD se sustituyó por MARC XML *Schema* que permite la conversión de los registros MARC 21 al lenguaje XML y viceversa sin pérdida de datos (Library of Congress, 2004). La principal diferencia es que las DTD son

archivos de gran tamaño, por lo que el cambio se hizo básicamente para simplificar la estructura y mejorar la integración con aplicaciones XML. Una de las ventajas que proporciona este esquema es que permite la integración de metadatos para la recolección por el protocolo OAI (Nogales, 2004).

A continuación, en la ilustración 15 se muestra un extracto de un registro bibliográfico de MARC XML Schema, en el que las etiquetas siguen las del registro MARC, pero son distintas al MARC XML DTD.

Ilustración 15: Extracto de un registro del esquema MARC XML Schema

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<collection xmlns="http://www.loc.gov/MARC21/slim"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/MARC21/slim MARC21slim.xsd">
  <record type="Bibliographic">
    <leader>01041cam 2200265 a 4500</leader>
    <controlfield tag="001"> 89048230 </controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19911106082810.9</controlfield>
    <controlfield tag="008">891101s1990 maua j 001 0 eng</controlfield>
    <datafield tag="010" ind1=" " ind2=" ">
      <subfield code="a"> 89048230 </subfield>
    </datafield>
    <datafield tag="020" ind1=" " ind2=" ">
      <subfield code="a">0316107514 </subfield>
      <subfield code="c">$12.95</subfield>
    </datafield>
```

Fuente: Nogales Flores, Tomás. *MARC en XML*. Universidad Carlos III de Madrid, 2004

Aunque en XML, MARC continuaba siendo un formato muy sofisticado que codifica la información bibliográfica de forma muy detallada, que en ocasiones los sistemas no requieren. Por lo que también en 2002 LC publicó una versión reducida con los elementos clave de un registro MARC adaptado a XML. Así

publicó *Metadata Object Description Schema* (MODS) (Coyle, 2005).

Existen actualmente numerosos software de gestión de bibliotecas que trabajan con MARC 21, la mayoría de ellos son de un coste de adquisición y mantenimiento elevado por lo que proyectos singulares difícilmente pueden adquirirlos por no contar con el suficiente presupuesto.

Metadata Object Description Schema (MODS)

Como decíamos, MODS es un esquema de metadatos derivado del MARC 21 de registros bibliográficos. Se publicó en 2002, y emplea etiquetas del lenguaje de marcado XML para la descripción bibliográfica del recurso, que pueden ser todo tipo de materiales bibliográficos (libros, capítulos de libros, series, artículos seriados, música). Consiste en un esquema de MARC XML simplificado, con los campos básicos, que permite definir relaciones jerárquicas entre las descripciones y proporciona una codificación detallada de las partes del recurso (volumen, número, capítulo, sección, párrafo, pista), pudiendo llegar hasta la codificación de secciones o párrafos dentro de un determinado texto (Coyle, 2005). El registro MODS se basa en veinte elementos y un conjunto de subelementos o hijos, que son equivalentes a un conjunto de campos del formato MARC 21.

Estos elementos y subelementos se recogen en el sitio web oficial de MODS de la Library of Congress. A diferencia las etiquetas del registro MARC que son numéricas, las etiquetas de MODS son en lenguaje natural y por tanto comprensibles por las personas y mucho más intuitivas (Coyle, 2005). En la siguiente ilustración se puede observar la gran diferencia que existe en la comprensión de los campos entre un formato y el otro.

Ilustración 16: Ejemplo de los campos titulo de un registro MODS y equivalencia con los campos del registro MARC.

title <titleInfo>	
<title> with no <titleInfo> type attribute and	245 \$a with ind1=1
<subTitle>	245 \$b
<partNumber>	245 \$n
<partName>	245 \$p
<nonSort></nonSort>	ind2=number of characters in content; ind2=0 if <nonSort> not present
<title> with <titleInfo> type="abbreviated" and	210 \$a with ind1=1 ind2=blank
<subTitle>	210 \$b
<title> with <titleInfo> type="translated" and	246 \$a with ind2=1
<subTitle>	246 \$b
<partNumber>	246 \$n
<partName>	246 \$p
<title> with <titleInfo> type="alternative" and	246 \$a ind1=3 ind2=blank
<subTitle>	246 \$b
<partNumber>	246 \$n
<partName>	246 \$p
displayLabel="content"	246 \$i

Fuente: Library of Congress, *MODS to MARC 21 mapping*, 2009

Metadata Authority Description Schema (MADS)

En 2004 la Oficina de Desarrollo de Redes de la LC publicó *Metadata Authority Description Schema* (MADS) para la descripción de autoridades: personas, organizaciones, eventos y términos de materias, geográficos. El formato se expresa utilizando el lenguaje XML del W3C, y emplea las etiquetas de este esquema para describir los campos de autoridad del formato MARC de registros de autoridad. Este registro de autoridad es información vinculada a un registro bibliográfico MODS. MADS y MODS pueden utilizarse de forma independiente, pero la definición de elementos y atributos están en coordinación, por lo que mediante una etiqueta XML de hipere enlace <xlink>, se hace la referencia de un registro MADS de autoridad a uno MODS bibliográfico, y viceversa (Library of Congress, 2012).

En la siguiente imagen se puede ver las etiquetas del formato MADS XML, comprensibles fácilmente por emplear lenguaje natural, y su equivalente en el formato MARC 21, con códigos numéricos para la interpretación automática, de similar forma al ejemplo visto anteriormente para el formato MODS.

Ilustración 17: Ejemplo de campos de autor en formato MADS y los campos equivalentes en formato MARC 21

```

<authority><name type="personal"> = MARC 21 Authority field 100
<authority><name type="family"> = MARC 21 Authority field 100 (1st Indicator=3)
<authority><name type="corporate"> = MARC 21 Authority field 110
<authority><name type="conference"> = MARC 21 Authority field 111

```

Fuente: Library of Congress. MADS User Guidelines, 2012.

Ilustración 18: Ejemplo de un registro de autoridad para un autor personal en formato MADS

```

- <mads xsi:schemaLocation="http://www.loc.gov/mads/mads.xsd">
  - <authority>
    - <name>
      <namePart>Smith,John</namePart>
      <namePart type="date">1995-</namePart>
    </name>
  </authority>
  - <variant type="other">
    - <name>
      <namePart>Smith, J</namePart>
    </name>
  </variant>
  - <variant type="other">
    - <name>
      <namePart>Smith, John J</namePart>
    </name>
  </variant>
  <note type="history">Biographical note about John Smith.</note>
  - <affiliation>
    <organization>Lawrence Livermore Laboratory</organization>
    <dateValid>1987</dateValid>
  </affiliation>
</mads>

```

Fuente: Library of Congress *MADS: Metadata Authority Description Schema*, 2013 <<http://www.loc.gov/standards/mads/>>.

Como muestra el ejemplo (ilustración 18), MADS recoge tanto las formas autorizadas como las no autorizadas de los nombres de las autoridades del documento. La forma autorizada de un nombre de un autor, institución, materia, será única, y las no autorizadas serán todas las variantes y sinónimos para referirse a esa autoridad y que remitirán a la forma normalizada. El conjunto de elementos y subelementos para describir un registro de autoridad de un recurso se recogen en el sitio web oficial de MADS de la LC.

De forma paralela al trabajo realizado por la LC en cuanto a la adaptación del formato MARC a los lenguajes de marcado, la OCLC y NCSA (National Center for Supercomputing Applications), en 1995 promovieron la iniciativa Dublin Core Metadata Initiative, para asignar una serie de etiquetas de información a los documentos digitales generados en HTML, con el propósito de facilitar su recuperación. El conjunto inicial de descripción estaba formado por quince elementos, lo que lo hacía mucho más sencillo que el MARC 21 (Senso y De la Rosa, 2003). A continuación veremos su evolución.

Dublin Core Metadata Element Set (Dublin Core)

Dublin Core Metadata Element Set (en adelante DC) fue el resultado, en 1995, de un esfuerzo internacional surgido en la convocatoria de un taller de trabajo en Dublin (Ohio) para mejorar la normativa relacionada con la recuperación de información en el ámbito de la web. Fue propuesto por una serie de expertos del mundo bibliotecario, de la gestión de redes y comunicaciones, informáticos, como una solución sencilla de catalogación para evitar formatos complejos, que requerían formación experta como MARC. Su sencillez permitiría a los creadores de los documentos digitales describir los recursos imitando al control bibliográfico tradicional empleado en bibliotecas (Senso y Rosa, 2003).

DC surge, del mismo modo que el resto de sistemas de metadatos, para afrontar el problema del crecimiento de documentos electrónicos en la web y la imprecisión de los motores de búsqueda, para garantizar su localización (Senso y Rosa, 2003). La principal característica de DC es la sencillez, por lo que era perfecto para erigirse como norma internacional. Es un estándar sencillo, extensible e interoperable, constituido después de varios años como norma ISO 15836-2007. Permite su integración con XML, HTML, XHTML, y RDF/XML, es

decir, que los metadatos que describen el documento pueden integrarse en la estructura del propio recurso o en otro documento a parte conectado a través de un enlace (DCMI, 2012).

La versión original estaba compuesta de un conjunto básico de quince elementos que permite describir de forma general todo tipo de recursos existentes en la red (páginas web, monografías, recursos sonoros y visuales). Es un sistema extensible, y cuenta con una versión ampliada *Qualified Dublin Core* con más de cincuenta propiedades que permite asignar calificadores para dotar de más detalle a los elementos (DCMI, 2000).

A continuación revisaremos los quince descriptores básicos (ninguno de ellos obligatorios y todos repetibles), se agrupan en tres categorías:

- Elementos relacionados con el contenido del recurso: *Title, Subject, Description, Source, Lenguaje, Relation, Coverage.*
- Elementos relacionados con el contenido desde el punto de vista de la propiedad intelectual: *Creator, Publisher, Contributor, Rights.*
- Elementos relacionados con el recurso como instancia o *Instantiation: Date, Type, Format, Identifier.*

El grupo DCMI ha elaborado varias guías de perfiles de aplicación que recogen los términos de metadatos utilizados en aplicaciones particulares, sus relaciones, las reglas de uso y los formatos de datos para diferentes comunidades. Entre estas guías, *DC-Lib Dublin Core Library Application Profile* (DCMI, 2004; Coyle y Baker, 2009) define el conjunto de elementos de metadatos DC para aplicaciones en las bibliotecas. El uso de DC en un entorno bibliotecario facilita el intercambio entre sistemas que utilizan diferentes estándares de metadatos y el cosechamiento de metadatos de diferentes fuentes de datos dentro y fuera del dominio bibliotecario. Permite descubrir recursos de creadores no bibliotecarios y exponer datos MARC a otras comunidades no bibliotecarias a través de la conversión a DC. La aplicación de DC ha tenido gran relevancia en la iniciativa OAI, repositorios instituciones y en proyectos como el *Cooperative Online Resource Catalog* (CORC) (Eito, 2008).

La ilustración 19 presenta la visualización de un registro bibliográfico del recurso Travesía del Ministerio de Educación, Cultura y Deporte, descrito con DC.

Ilustración 19: Visualización de registro en Dublin Core

Registro completo de metadatos	
Campo DC	Valor
dc.contributor.author	Dublin Core Metadata Initiative Usage Board
dc.date.issued	2012-06-14
dc.identifier.uri	http://hdl.handle.net/10421/3322
dc.description.abstract	Este documento es una especificación actualizada de todos los términos de metadatos mantenidos por Dublin Core Metadata Initiative, incluyendo propiedades, esquemas de codificación de vocabularios, sintaxis de codificación de esquemas y clases.
dc.language.iso	en
dc.publisher	Dublin Core Metadata Initiative Usage Board
dc.rights	Acceso abierto
dc.title	DCMI Metadata Terms
dc.type	Text
dc.subject.tbd	Metadatos
dc.subject.tbd	XML
dc.relation.publisherversion	http://dublincore.org/documents/2012/06/14/dcmi-terms/
Aparece en las colecciones: A2.1. Normas, directrices etc.	

Fuente: Travesía: recursos digitales para la cooperación bibliotecaria
<<http://travesia.mcu.es>>

Desde 1998, los quince elementos empezaron su proceso de estandarización. Con la aparición de las buenas prácticas de descripción en el marco de la web semántica, DC se ha adaptado a las nuevas estructuras de descripción. Desde 2008 se han creado unos *namespaces* para los elementos DC para emplear con el lenguaje RDF y participar en la descripción de recursos digitales en el contexto de la web semántica (DCMI, 2012). Los *namespaces* o espacios de nombre permiten definir un prefijo para renombrar el sistema de marcado. Se emplean para combinar en un solo documento u objeto digital metadatos pertenecientes a diferentes lenguajes. Los espacios de nombre añaden un prefijo al nombre de las marcas utilizadas para

diferenciar así los elementos de cada lenguaje (Codina, 2009).

La versión actual *Dublin Core Metadata Terms* (2012) cuenta con un mayor número de términos en el que cada uno tiene una URI asignada para su identificación y está definido como propiedad del lenguaje RDF para formar parte de la web semántica, como veremos en los siguientes capítulos (DCMI, 2012).

Existen otros estándares de metadatos cuyo propósito no es la descripción del documento en el sentido de catalogación, sino para la codificación de los documentos, cuya función es mantener unidos los archivos que componen un objeto digital (Coyle, 2005). Por ejemplo, existen el *Metadata Encoding and Transmission Standard* (METS), un estándar de codificación y transmisión de documentos en formato electrónico de 2002, y el *Text Encoding Initiative* (TEI), que fue una primera iniciativa de codificación de textos de 1987. A continuación se revisarán brevemente, pues a diferencia de los esquemas de metadatos observados, éstos no tienen destacada vinculación con la web semántica y su adaptación al lenguaje RDF.

Metada Encoding and Transmission Standard (METS)

Es un estándar de metadatos de codificación de los componentes que forman los objetos digitales complejos. Fue creado por la Digital Library Federation, y el organismo que lo mantiene actualmente es la MARC Standard Office de la LC. Surgió en 2002 a partir del proyecto Making of America II (MOA2), para gestionar los metadatos para la transmisión e intercambio de documentos electrónicos entre bibliotecas y repositorios OAI (METS, 2011). Es un esquema XML para el envío de ficheros, imágenes y objetos multimedia de una biblioteca digital. Un documento electrónico puede estar formado por varios ficheros como imágenes, archivos de texto y archivos de audio que forman un único objeto digital. METS relaciona los ficheros que componen el objeto digital y su estructura, y permite expresar las relaciones existentes entre estos ficheros para dotarlos de estructura y que las aplicaciones puedan representarlos de forma ordenada. Es decir, que proporciona los metadatos estructurales. Y los combina con los metadatos descriptivos, administrativos y de gestión para un objeto digital complejo (METS, 2011).

Un documento METS consta de siete secciones, compuestas por varios elementos y sus atributos para describir un objeto digital (METS, 2011):

- La cabecera: <mets:Hdr> contiene los metadatos de

descripción con la información del documento METS relativos a su creador, fecha de creación, editor. Es opcional y no repetible.

- Los metadatos descriptivos del objeto digital o de alguna de sus secciones: <mets:dmdSec> contiene la descripción del objeto al que se hace referencia en el documentos METS. Pueden ser tanto internos como externos al documento. Es opcional y repetible.
- Los metadatos administrativos del objeto digital o de alguna de sus secciones: <admSec> relativos a la creación y almacenamiento de los archivos del objeto digital, gestión de derechos, de uso y copias del contenido, pueden ser externos o codificarse dentro del propio documento. Es opcional y repetible.
- La sección de archivo: <fileSec> es el listado de ficheros que componen el objeto digital. Se agrupan en <fileGrp>. Es opcional y no repetible.
- El mapa estructural del documento: <structMap> define la estructura jerárquica del objeto, es decir, las secciones que componen el objeto digital, las relaciones entre estos ficheros y la secuencia de los mismos. Es obligatorio y puede repetirse.
- Los enlaces estructurales: <smLink> son registros de los hiperenlaces entre las secciones del mapa, es decir,

los enlaces entre las partes, capítulos, secciones del objeto digital. Es opcional y no repetible.

- Los comportamientos: `<mets:Behaviour>` hacen referencia a mecanismos que identifican un módulo de código ejecutable en la interfaz, es decir, registra los comportamientos o servicios que pueden asociarse a un objeto digital. Es opcional y repetible.

La gestión de una biblioteca de objetos digitales requiere la gestión de varios tipos de metadatos sobre esos objetos. Necesita metadatos estructurales, para identificar las imágenes y los archivos de texto que conforman el objeto digital por ejemplo, un libro. Así como también para la gestión de metadatos descriptivos y administrativos. En este sentido, el esquema METS propone un formato para referenciar en un documento XML todos los archivos que forman un objeto digital, y ofrece un medio flexible para codificar metadatos descriptivos, administrativos y estructurales para un mismo objeto digital, y para expresar las relaciones entre estos tipos de metadatos (METS, 2011). De nuevo se trata de lenguajes de codificación muy sofisticados creados para resolver problemas relacionados con la identificación de los elementos que componen el objeto digital y su estructura, que las tecnologías semánticas, tal y como se implementan en Vestigium, solventan con mayor eficacia, como se verá en el siguiente capítulo.

Text Encoding Initiative (TEI)

Fue una de las primeras iniciativas para normalizar la información electrónica para su intercambio. En 1987 fue desarrollada de forma conjunta por la Association for Computational Linguistics, Association for Literary and Linguistic Computing y Association for Computing and the Humanities para el uso de lenguajes de marcas en la codificación de texto completo de documentos, especialmente como aplicación de SGML. No sólo se basa en la identificación de las etiquetas que facilitan la codificación de los textos sino también en los metadatos básicos para su catalogación y posterior recuperación. Se trata de un formato complejo que en su última versión P5 de 2007 estaba formada por unos 600 elementos. TEI Consortium continúa su mantenimiento y cuenta con numerosos proyectos que lo utilizan (TEI Consortium 2007; Eito, 2008).

En resumen, aunque existen otras estructuras de definición de metadatos, los estándares revisados son los más extendidos internacionalmente en el área de la gestión de la información. Con algunos de ellos y sus complementos o con la combinación de varios es posible representar la descripción bibliográfica de los recursos electrónicos mayoritarios de Internet. Estos estándares se han desarrollado en el marco de la comunidad

bibliotecaria y por tanto su base se encuentra en las normas de descripción bibliográfica tradicionales. El formato MARC 21 continúa siendo utilizado en el ámbito de bibliotecas digitales, que ha demostrado su capacidad de adaptación y su capacidad de conversión a diversos esquemas de metadatos incluyendo ya su uso con RDA (Saorín-Pérez y Pastor-Sánchez, 2012), como muestran las actualizaciones aprobadas en 2012 de los nuevos campos y atributos en MARC para utilizarse con RDA (Library of Congress, 2012). A pesar de ello, la situación actual es un punto de transición, en el que los esfuerzos de normalización se centran en el ámbito digital. La LC lanzó en 2011 la *Bibliographic Framework Transition Initiative* dirigida a iniciar el proceso de transición del formato MARC 21 hacia un nuevo modelo de descripción bibliográfica LD denominado BIBFRAME. Aunque el nuevo modelo se encuentra en una fase inicial está en consonancia con los principios LD, el modelo de datos RDF, y apoya la nueva norma de contenido RDA, enfocada al intercambio de datos bibliográficos en la web de datos. De nuevo, con esta iniciativa se pone de manifiesto la incesable labor que están realizando las bibliotecas nacionales para participar de la web semántica (Library of Congress,2012).

La producción bibliográfica, académica y científica, como hemos visto, se genera, se preserva y se consume en el entorno digital. Por esta razón es necesario acercarse a las tecnologías web, y especialmente a las tecnologías de la web semántica que,

es hacia donde se dirigen los esfuerzos de instituciones culturales y patrimoniales internacionales.

**LA WEB DE LOS DATOS
ENLAZADOS, LA WEB SEMÁNTICA**

Según los datos estadísticos recogidos por el INE, en la encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación (TIC) en los hogares en 2008 el 60,7% de la población española utilizaba Internet diariamente, al menos cinco días por semana. En 2009 el porcentaje de población que utiliza Internet diariamente, aumentó al 65,4%, en 2010 al 69,2%, en 2011, al 71,4% y en 2012, al 72,6%. Estos datos revelan que más de la mitad de la población española utiliza diariamente la web, y que su uso va en aumento.

El precursor de la idea de web de enlaces de hipertexto fue Vannevar Bush (1890-1974), ingeniero y profesor del Massachusetts Institute of Technology (MIT) y asesor, durante la Segunda Guerra Mundial, de Investigación Científica y Desarrollo en la administración americana siendo presidente Roosevelt. En 1945 describió la primera aplicación hipertextual, un proyecto llamado MeMex *Memory Extended* y descrito en *As we may think*, (Cómo podríamos pensar) publicado en la revista *Atlantic Monthly* (Vianello, 2004). La idea presentada era la de un sistema distribuido de información accesible a través de un escritorio (interfaz), en el que los archivos de texto e imagen estuvieran ligados y pudieran ser consultados a través de los enlaces. Así los usuarios podrían visualizar la información a través de distintos recorridos (Bush, 2001). Este proyecto es lo que puede considerarse el origen del hipertexto.

En la década de los años 60 Theodor Nelson acuñó el término *Hypertext* como “non-sequential writing”, una escritura no secuencial. Se trata de un conjunto de bloques de texto conectados entre sí mediante unos enlaces que permiten navegar de un texto a otro (Nelson, 1993).

Paralelamente, Douglas Carl Engelbart, del centro Augmentation Research Centre (ARC) de la Universidad de Stanford, trabajó sobre la idea de Vannevar Bush, desarrollando un software capaz de implementar hipertexto o referencias cruzadas automáticas a otros documentos. El proyecto *oN Line System* (NLS) consistía en un sistema de navegación en línea, una red local que enlazaba los documentos publicados por el instituto, para poder acceder a ellos y trabajar en grupo (Lamarca, 2013; Berners-Lee, 2000). Basándose en estos proyectos, fueron surgiendo numerosos sistemas de hipertexto como *Hypertext Editing System (HES)* en 1967 por IBM y la Brown University, *File Retrieval and Editing System (FRESS)* en 1968, *ZOG* en 1972 de la Carnegie Mellon University, *Knowledge Management System (KMS)* en 1983, *Intermedia* en 1985 en la Brown University, *Note Cards* en 1985 por la Xerox Parc o *Smar Text* en 1988 (Berners-Lee, 2000).

En 1990 Tim Berners-Lee y Robert Cailliau dieron a conocer la web como la conocemos actualmente. Presentaron el proyecto World Wide Web (WWW) en la Organización Europea para la

Investigación Nuclear (CERN) en Ginebra (Suiza), que constituyó la primera propuesta de web. Consistía en un programa cliente (navegador/editor) basado en el lenguaje de etiquetas de hipertexto HTML (*Hyper Text Markup Language*), junto con el protocolo de transferencia de hipertexto HTTP (*Hyper Text Transfer Protocol*), lenguaje que los ordenadores usan para comunicarse por Internet, y el identificador uniforme de recursos digitales URI (*Uniform Resource Identifier*), para la localización de objetos digitales. El navegador/editor World Wide Web era capaz de seguir vínculos en los archivos de los servidores HTTP (Berners-Lee, 2000).

Previamente y hasta ese momento, el estándar empleado era el protocolo FTP (*File Transfer Protocol*), para la transferencia de archivos de un servidor a otro, que fue puesto en marcha en los años setenta por el grupo de trabajo Network Working Group (NWG) en los laboratorios del ARPA (Agencia de Investigación de Proyectos Avanzados) del Departamento de Defensa de los EEUU (Berners-Lee, 2000).

De esta forma, la información existente en Internet se encontraba ya disponible en la web, y los ordenadores podían comunicarse unos con otros y transferir información, desde cualquier ordenador. Esta web, la del hipertexto, es una web dinámica, multidireccional, que permite la exploración rápida de grandes bloques de conocimiento. Proporciona al usuario la

posibilidad de elegir hacia dónde dirigir su lectura, saltar de un contenido a otro, conectando textos distribuidos y asociados mediante enlaces electrónicos. Este cambio, casi de naturaleza filosófica, requería hacer comprender a los usuarios la importancia de aportar datos a la web. La posibilidad que ofrecía la web no sólo era la consulta, sino la también la oportunidad de incluir contenidos que serían consultados por otros usuarios (Berners-Lee, 2000). De este modo la web se ha expandido, consolidando su uso en todo el mundo y por toda clase de usuarios. Ha evolucionado de forma natural hacia la adaptación a las necesidades de comunicación del ser humano, creando una sociedad y una sociabilidad nueva, on line. Surgen nuevos servicios que cuentan con un alto grado de interactividad con los usuarios y aplicaciones orientadas a la comunicación colectiva. En esta última parte tuvo mucho que ver Tim O'Reilly, propietario de la empresa editorial estadounidense de libros de informática, O'Reilly Media, que en el año 2004 acuñó el término web 2.0 para hacer referencia a estos cambios (O'Reilly, 2005). La web definida como social destaca el alto nivel de participación de los usuarios para crear contenidos y establece relaciones para compartirlos. Se crea toda una gama de servicios de colaboración para el intercambio de información entre usuarios: redes sociales, blogs, microblogging, wikis, folcsonomias, selecciones de música, de favoritos, donde la valoración de los contenidos depende de las relaciones entre los

individuos (O'Reilly, 2009). El año 2009 se consideró el año de la eclosión de las redes sociales en España, que continuó teniendo su auge en 2010.

Ahora la web 2.0 es un fenómeno ya consolidado que puede ser permanente en los próximos años hasta que cambien las tendencias. Algunos autores afirman la continuidad de estos servicios de redes sociales, ampliándose, como ya ha ocurrido a dispositivos móviles (Margaix, 2010). Pero la producción de nueva información en el ámbito digital sumada al volumen ya existente, que continúa creciendo a un ritmo vertiginoso debido a que los consumidores de información se constituyen a su vez en productores de contenidos, genera una serie de problemas de acceso y recuperación de la información. Los buscadores generales recuperan información que en ocasiones no es relevante para los usuarios. A esto hay que añadir que los contenidos académicos, científicos y técnicos no los recuperan estos buscadores, pues se encuentra en el denominado Internet invisible o Internet profunda (Codina, 2003; Aguillo, 2000; 2003; Peset et al., 2000), que hace referencia al sector de sitios y de páginas web que no pueden indizar los motores de búsqueda, sea por necesitar una suscripción, sea por que no es visible por causas técnicas. En este Internet se encuentra por ejemplo, la información producida para su uso interno por una institución u organismo, pero también, en muchas ocasiones, la información científica.

La web es un entorno que evoluciona constantemente como hemos visto, y en este sentido se avanza hacia la denominada web 3.0. La principal diferencia entre la web 2.0 y la web 3.0 es que, mientras la web social (2.0) está gestionada por los usuarios, la web 3.0 está enfocada al procesamiento lógico por los sistemas informáticos (Küster y Hernández, 2013). En cuanto a la web 3.0, y sus otras acepciones como web de datos y web semántica, existe bastante controversia en la literatura a la hora de establecer semejanzas o diferencias entre los conceptos. Para Küster y Hernández (2013) la web semántica forma parte de la web 3.0. Sin embargo el W3C, organismo que lleva la iniciativa y desarrolla esta web, emplea las acepciones web de datos y web semántica, para hacer referencia a la web de datos enlazados y las tecnologías semánticas que emplea para hacerla efectiva, que se verán a continuación. El enfoque del W3C se centra en los datos, y gira entorno a la cada vez mayor cantidad de datos existentes que se encuentran aislados en silos, y en cómo conectarlos (W3C, 2013; Álvarez, 2012).

El W3C empieza a plantearse la idea de web semántica a finales de los años 90, aunque hasta 2001 no se presenta un artículo en la revista *Scientific American* en el que se presentan los postulados de la web semántica. La propuesta de Tim Berners-Lee en 2001 pretende transformar la actual web. La web actual se basa en el lenguaje natural, tiene una falta de estructuración de contenidos y una carencia de descripciones normalizadas de

los recursos digitales, lo que se ve reflejado en la ambigüedad de los resultados que recuperan los motores de búsqueda generalistas (Vallez et al., 2010). Como consecuencia, para resolver el actual problema de accesibilidad a la información en Internet, surge la idea de web semántica, cuyo objetivo es identificar la información de forma unívoca y establecer relaciones entre los objetos digitales con miras a facilitar una recuperación eficaz de la información.

Para conseguir este objetivo, se emplean herramientas de descripción de contenidos orientadas a proporcionar un contexto informativo y a describir también el significado de la propia información. Tal y como hemos visto que ocurre en la descripción bibliográfica, es necesario el uso de metadatos. Los metadatos aportan información estructurada sobre la propia información que se publica. Por ejemplo, describen el tipo de recurso, su autoría, la fecha de creación. De esta manera los contenidos son etiquetados semánticamente de forma explícita para conseguir que las máquinas puedan interpretarlos, lo que facilitaría el procesamiento automático de los contenidos de la web. Uno de estos procesos automáticos sería precisamente la recuperación de información. El etiquetado y la asignación de metadatos son por tanto elementos básicos del proyecto de la web semántica (Vallez et al., 2010).

Ya hemos dicho que el elemento clave en el que se centra esta

web son los datos, y particularmente en los LD. Este término hace referencia al conjunto de buenas prácticas para publicar datos estructurados en la web y conectarlos con otros datos relacionados (Bizer et al., 2011). Es decir, exponer la información como datos bien definidos, identificados y enlazados, de forma que puedan ser conectados con otros conjuntos de datos de diferentes fuentes.

Seguir los principios de publicación de LD para definir la información, hace que los datos sean legibles por los programas de software y los buscadores semánticos. Uno de objetivos finales de la web semántica es la creación de un sistema de agentes inteligentes que sean capaces de llevar a cabo inferencias de forma automatizada con la información publicada en la web (Codina, 2006). A continuación revisaremos las pautas y las tecnologías que el W3C definió para describir y enlazar la información en el marco de la web semántica.

Generalidades de la web semántica

La adopción de las buenas prácticas de LD contribuye a la extensión de la actual web a un espacio global en el que se conectan datos relacionados entre sí pertenecientes a diversos dominios (Bizer et al., 2011). Los datos son de tipologías distintas, como por ejemplo, estadísticos, matemáticos, geográficos, químicos, bibliográficos, de autoridades, que se encuentran alojados en bases de datos y servidores de entidades independientes, como se verá más adelante en el proyecto The Data Hub de preservación de conjuntos de datos. Emplear estas prácticas en la publicación de información digital hace los datos disponibles para poder establecer conexiones, de forma que puedan ser compartidos, reutilizados y combinados, con el fin de ampliar la información de la web o recurso inicial de una forma estructurada.

Hans Rosling en una conferencia en TED en 2006 (Rosling, 2006) expuso la existencia de gran cantidad de datos en bases de datos aisladas, en las universidades, en las agencias nacionales de estadística, en instituciones no gubernamentales. Expuso la existencia de Internet, de público que necesita información y de las numerosas aplicaciones existentes. Y planteó la cuestión: “¿por qué no utilizamos los datos que tenemos?”. Y su respuesta fue: “No se utilizan los datos de forma efectiva. Es necesario

liberar los datos y hacerlos buscables, pero también se necesitan *linking data tool design* y *search function*, porque las funciones de búsqueda de las herramientas de datos enlazados pueden encontrar los datos de las bases de datos, desde un formato buscable y exponerlos al mundo”.

Para exponer la información almacenada en estas bases de datos aisladas hay que describirla y establecer relaciones semánticas (enlaces con significado) entre los datos. Crear enlaces con significado entre los diferentes elementos de un recurso es similar a las relaciones que se crean entre los campos de un registro como resultado de la descripción bibliográfica en un OPAC. Las relaciones que se definen en los LD se extienden más allá del catálogo y enlazan con otros conjuntos de datos de recursos externos que se encuentran publicados y dispersos por todo Internet. Para alcanzar este proceso hay que seguir las prácticas que establece el W3C. En 2006 Tim Berners-Lee (Berners-Lee, 2006) definió los cuatro principios de buenas prácticas LD:

1. **Utilizar URI** (Identificador Uniforme de Recursos) **como nombres de los objetos de información**, para identificar los recursos e identificar de forma unívoca y sin ambigüedades los recursos a los que se quiere hacer referencia. Así se podrían evitar las ambigüedades que puede producir el lenguaje natural en varios idiomas para

referirse a un mismo recurso. Una URI es una cadena de caracteres consistente asignada a un recurso digital que lo identifica de forma inequívoca como entidad. Identifica el recurso por nombre y cómo acceder al mismo especificando su localización.

2. **Utilizar URI HTTP** para que las personas puedan buscar la información por nombre. Existen varios tipos de identificadores de recursos digitales, recomiendan la utilización del protocolo HTTP de transferencia de hipertexto que define la sintaxis y la semántica de los elementos de la web para comunicarse. Este protocolo se basa en los elementos cliente-servidor, donde el cliente (navegador del usuario) realiza una petición a un servidor (base de datos) que devuelve un documento que se muestra a través de la interfaz del navegador. El uso del protocolo permite buscar cualquier recurso almacenado en un servidor y presentarlo a través de la web.
3. **Ofrecer información sobre los recursos usando lenguaje RDF** para que sea comprensible para el sistema informático y búsqueda mediante SPARQL Endpoint. Un SPARQL Endpoint es una interfaz general de consulta que recupera las sentencias de lenguaje RDF almacenadas en los repositorios de tripletas de datos, como se verá más adelante en este capítulo. Utilizar el

lenguaje RDF para describir el recurso ofrece al sistema información sobre el objeto digital representado.

4. **Incluir enlaces a otros URI** para que se puedan descubrir más recursos. Se trata de enlazar un recurso que está identificado con su URI HTTP y descrito en RDF, a otros recursos externos a través de sus identificadores URI HTTP, y que también han de estar descritos en RDF. Los recursos enlazados pueden estar alojados en un mismo servidor, o pueden enlazar a objetos distribuidos por toda la red. Esto permite enriquecer el sitio web local con otra información experta procedente de otros dominios en la red, como también favorece evitar la duplicidad.

Estas cuatro pautas contribuyen a publicar datos en la web y formar parte de un mismo espacio global de datos semánticos, creando una extensa red de información relacionada por la que navegar. Para ello, LD se basa en las siguientes tecnologías web:

- HTTP: protocolo para acceder y recuperar recursos en la web.
- RDF: es un lenguaje para la representación de metadatos, un lenguaje de etiquetado con sintaxis XML, para describir la información sobre los recursos web a través de propiedades y valores.
- Modelos de metadatos: como se ha visto en el capítulo

anterior, mantienen su uso en el contexto de la web semántica para describir el contenido de los recursos a través de especificaciones comunes.

- Ontologías: un lenguaje estándar para describir los conceptos y sus relaciones.
 - RDF Schema: es un estándar para describir jerarquías de clases de recursos y las propiedades de los datos.
 - OWL: lenguaje de representación de ontologías para definir clases, propiedades y relaciones.
 - SKOS: modelo para representar la estructura y el contenido de los sistemas de representación del conocimiento.

A continuación veremos estas tecnologías.

Publicación y descripción en el entorno web

HyperText Markup Language (HTML)

La creación de páginas web se desarrolla con el lenguaje de marcas de hipertexto llamado HTML (*HyperText Markup Language*) creado a principios de 1990 como lenguaje para la World Wide Web (WWW). Su origen está en el lenguaje SGML (*Standard Generalized Markup Language*) que empezó a desarrollarse en 1969 por IBM aunque inicialmente se denominó GML (*Generalized Markup Language*). GML es un marco general para la descripción de otros lenguajes de marcado. Este lenguaje surgió como formato para transferir y gestionar los documentos de la empresa. La idea fue ampliada hasta llegar a convertirse en la norma SGML durante una reunión del ANSI (American National Standards Institute) en 1978. Es norma ISO 8879 desde 1986. SGML es un metalenguaje cuyo ámbito de aplicación es general, y con el que se pueden diseñar lenguajes de marcado como HTML (Peis y Moya, 2000).

Los lenguajes de marcas son un conjunto de normas que especifican las convenciones de marcaje para codificar textos. Permiten definir la información de los recursos electrónicos de una forma estructurada. Utilizar un lenguaje de marcado para

codificar un documento da como resultado un archivo de texto plano, en el que el contenido del documento está estructurado mediante las etiquetas (Peis y Moya, 2000).

La gran aportación del lenguaje de marcas HTML viene dada por la transformación del concepto de "documento" por el de "metadocumento", es decir, la transformación de un corpus textual único en un corpus compuesto por múltiples elementos (texto, sonido, imagen fija, imagen en movimiento, enlaces a otros documentos) relacionados entre sí (Martín y Méndez, 1998). El lenguaje de marcas HTML define la estructura del documento mediante la asignación de un conjunto limitado de etiquetas. Estas marcas o etiquetas definen las partes del documento, y todos los elementos que se visualizan, su distribución por la página, el tamaño y el color. Los documentos HTML son archivos de texto y la visualización de la información se realiza mediante un navegador web que interpreta el código. La siguiente imagen es un ejemplo de un documento HTML, estructurado mediante las etiquetas del lenguaje de marcas.

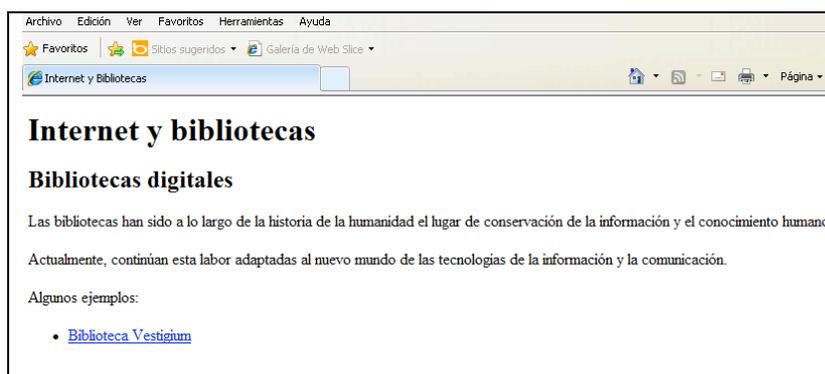
Ilustración 20: Ejemplo de documento HTML

```
<html><head> <title>Internet y Bibliotecas
</title></head><body><H1>Internet y bibliotecas </H1><H2>
Bibliotecas digitales</H2><p> Las bibliotecas han sido a lo largo de la
historia de la humanidad el lugar de conservación de la información y
el conocimiento humano. </p><p> Actualmente, continúan esta labor
adaptadas al nuevo mundo de las tecnologías de la información y la
comunicación. </p><p>Algunos ejemplos:</p><ul><li><A
href="http://www.vestigium.es/biblioteca/index.php"> Biblioteca
Vestigium</li></ul></body></html>
```

Fuente: Elaboración propia

La ilustración 20 ejemplifica una versión simplificada de una página web escrita en HTML. Las etiquetas se duplican: etiqueta de inicio y etiqueta de cierre, la de cierre se identifica con /. El documento se estructura en dos bloques: el encabezamiento (<head>) y el cuerpo del documento (<body>). Dentro del encabezamiento se incluyen las etiquetas de título. A continuación, la etiqueta del cuerpo del documento, en la que se incluyen las etiquetas de título y subtítulo con dos tamaños distintos de letra. También, se incluyen etiquetas para presentar el texto en dos párrafos, para establecer elementos en un listado y para incluir un enlace de un sitio web externo. Finalmente, las correspondientes etiquetas de cierre de los elementos. La visualización de este documento se presenta en la siguiente ilustración.

Ilustración 21: Visualización del documento HTML en la web



Fuente: Elaboración propia

La ilustración 21 muestra cómo se visualiza ese documento mediante un navegador web. La presentación de los datos en el navegador es inmediatamente comprendida por una persona, sin embargo para el programa que ejecuta las instrucciones no es posible entender el contenido informacional (la semántica) de la página web. La visualización del contenido, es decir, el aspecto visual de la información se define en la estructura del documento mediante etiquetas, como las etiquetas con las que se han establecido tres tamaños de letra diferentes en el ejemplo. En el lenguaje HTML la definición del aspecto visual del contenido se define en el mismo documento, y no es posible separar el contenido informativo de la forma de presentación visual. Este modo de descripción no permite combinar elementos de diferentes lenguajes, convirtiéndolo en un lenguaje no extensible. En cuanto a la interoperabilidad para intercambiar datos, es limitada, pues permite un escaso número de metadatos

en la definición del encabezamiento del documento (De la Rosa, 2003).

El W3C desde 1996 centralizó el desarrollo de HTML, y ha propuesto y desarrollado el formato XML. Buscaba una norma para intercambiar información estructurada en la red con independencia de plataformas y aplicaciones. Aunque SGML era una opción ideal, su complejidad y tamaño dificultaban su estandarización. La otra opción era pues, HTML, pero este metalenguaje sólo sirve para formalizar el formato de los documentos y no su contenido. Por lo que se seleccionó y adaptó un subconjunto de la norma SGML, que es lo que ha dado lugar a la actual especificación XML (De la Rosa, 2003).

eXtensible Markup Language (XML)

Tras el éxito de Internet como sistema para el intercambio de documentos, en 1996 el W3C desarrolló el lenguaje de marcas extensible XML (*eXtensible Markup Language*). En la evolución de HTML a XML, destaca un elemento clave: la separación entre estructuración de la presentación y estructuración del contenido. XML es un medio para estructurar el contenido de los documentos digitales, separa el contenido del documento de su forma de presentación. A diferencia del HTML, lo que consigue es relacionar el contenido con el tipo de

presentación. Pero eso no afecta al contenido en sí. XML no es un vocabulario como HTML, las etiquetas XML no significan nada para un navegador. XML es un lenguaje que podemos utilizar para gestionar el contenido de un documento digital (De la Rosa, 2003).

Como se ha dicho, XML permite el intercambio de información estructurada con independencia de las aplicaciones y entre diferentes plataformas web (por ejemplo, una aplicación en Linux que es un sistema operativo de código libre puede comunicar perfectamente con otra aplicación en un sistema operativo propietario como por ejemplo, Windows). Para trabajar con XML es necesario estructurar la información. Del mismo modo que se describen los componentes del documento en la descripción bibliográfica, XML denomina a todos estos componentes estructurales, elementos. Estos elementos tienen atributos que definen sus características y, estos atributos tienen unos valores (De la Rosa, 2003). La forma de definición del documento se basa en el anidamiento de las etiquetas en una estructura estrictamente jerárquica, donde sólo existe un elemento raíz del que derivan los demás elementos, sus atributos y los valores que éstos reciben (W3C, 2010a), como vemos en la ilustración 22.

Ilustración 22: Ejemplo de documento escrito en XML con etiquetas de definición de la estructura pero sin etiquetas de estilo para la visualización

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<libro>
<titulo></titulo>
<capitulo>
<titulo></titulo>
<seccion>
<titulo></titulo>
</seccion>
</capitulo>
</libro>
```

Fuente: W3C, Guía breve de tecnologías XML, 2010b
<<http://www.w3c.es/Divulgacion/GuiasBreves/TecnologiasXML>>

XML es un lenguaje extensible con el que es posible crear nuevas etiquetas para las infinitas variaciones de las relaciones entre partes de documentos. Es flexible porque permite combinar elementos de diferentes vocabularios específicos de ámbitos de aplicación distintos (documentación, medicina, economía y otros) que puedan ser combinados y mezclados según sea necesario (De la Rosa, 2003). Sus características lo convierten en un lenguaje adecuado para emplear en el conjunto de tecnologías que forman la web semántica.

El W3C ha ido desarrollando numerosas extensiones, y estándares relacionados (XQuery, XML Schema), y una familia de lenguajes con los que describir la visualización de los archivos codificados en XML. La visualización de los archivos XML se describe en un fichero de estilo anexo donde se definen

todos los parámetros del formato de presentación como las tipografías, los tamaños de la fuente, los márgenes, los colores y todas las características de estilo. Este lenguaje es el XSL (*eXtensible Stylesheet Language*), del que se derivan tres lenguajes: XSL Transformation (XSLT), XSL Formatting Objects (XSL-FO) y XML Path Language (W3C, 2010b).

Teniendo en cuenta los lenguajes vistos con anterioridad, las pautas de identificación de contenidos y protocolos de transferencia de datos, en el año 2000 Berners-Lee escribió una recomendación de cómo debía ser la arquitectura de la web semántica y las tecnologías empleadas necesarias que recogemos en la ilustración 23.

Ilustración 23: Arquitectura de la web semántica según el W3C de 2000

Metadatos	Ontologías	Namespaces	RDF
XML			
Unicode		URI	
Protocolos de comunicación			

Fuente: Berners-Lee, Tim. Semantic web on xml, 2000, <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>

En la recomendación, el W3C propone como base del intercambio de información en la web, varios protocolos de comunicación de datos entre computadoras, entre los que

destaca HTTP para la web semántica. En la capa inmediatamente superior de la estructura propone los códigos de caracteres Unicode y los identificadores de recursos URI. El siguiente elemento propuesto es el lenguaje XML para estructurar documentos. XML se propone como un estándar universal para el intercambio de información estructurada entre diferentes plataformas, que además es combinable con otras tecnologías existentes en la web. XML es la base en cuanto a sintaxis para el desarrollo de las tecnologías de la web semántica en las que se basan los LD. Estas tecnologías se sitúan en la capa superior de la estructura propuesta como arquitectura de la web semántica y son los metadatos, las ontologías, los *namespaces* y el lenguaje RDF para la definición de datos enlazados, que describiremos a continuación (Berners-Lee, 2000).

Resource Description Framework (RDF)

RDF (marco de descripción de recursos) es un lenguaje de definición de metadatos para describir o catalogar distintos objetos de información distribuidos en la web (W3C, 2014). Fue desarrollado por el W3C en 1997. Es una DTD (definición del tipo de documento) que se expresa en XML para facilitar la interoperabilidad entre diferentes modelos de metadatos

(Méndez, 1999). Para Méndez (1999) es el lenguaje más adecuado para la descripción de contenidos de las bibliotecas digitales por su vinculación con los metadatos, por lo que se adapta perfectamente al objeto de esta tesis, que será ilustrada preferentemente con ejemplos de la comunidad bibliotecaria.

Una de las diferencias más significativas que supone emplear el lenguaje RDF para describir los recursos digitales es la estructuración de la información y el uso de la semántica de los conceptos. XML no es suficiente, pues modela documentos, mientras que RDF puede representar la estructura de relaciones del mundo real. Este lenguaje identifica los elementos del mundo real y cómo se relacionan entre sí en el ámbito de la web, ya que vincula estos elementos a través de relaciones semánticas explícitamente diferenciadas (Heath y Bizer, 2011). Se fundamenta en estas relaciones que se codifican en forma de tripletes constituidas por un sujeto, un predicado y un objeto (Heath y Bizer, 2011; Cyganiak et al., 2014; Pastor, 2011):

- El **sujeto** es el recurso informativo o elemento del mundo real que se quiere representar (por ejemplo, una persona, un libro, un lugar).
- El **objeto** es el valor que tiene el elemento sujeto (por ejemplo, “Miguel” o “El ingenioso hidalgo Don Quijote de la Mancha”). El sujeto y el objeto se describen con URI que identifican el recurso por su

nombre y su localización de forma unívoca y duradera.

- El **predicado** es la relación específica que existe entre sujeto y objeto, una característica o propiedad del sujeto, y puede ser una cadena de texto o también una URI (por ejemplo, tipo “persona” “autor”). Las relaciones se describen mediante los elementos de vocabularios de metadatos (DC, Schema).

La ilustración 24 ejemplifica qué tipo de información contiene cada uno de los elementos de un triplete de lenguaje RDF.

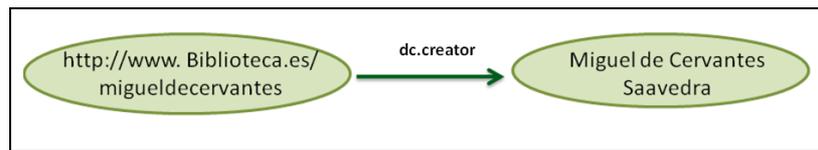
Ilustración 24: Ejemplo de enlaces RDF de un triplete

Sujeto:	http://www. Biblioteca.es/ migueldecervantes
Predicado:	http://purl.org/dc/elements/1.1/creator
Objeto:	Miguel de Cervantes Saavedra

Fuente: Elaboración propia

La representación gráfica del triplete RDF se muestra en la ilustración 25. Es un grafo en el que sujeto y objeto son nodos y el predicado es un enlace que identifica la relación entre los nodos mediante un elemento de un vocabulario de metadatos (Castells, 2003). En este caso, la URI del predicado está identificando la propiedad autor del esquema Dublin Core que vimos en el anterior capítulo.

Ilustración 25: Representación gráfica del triplete RDF de la ilustración 24

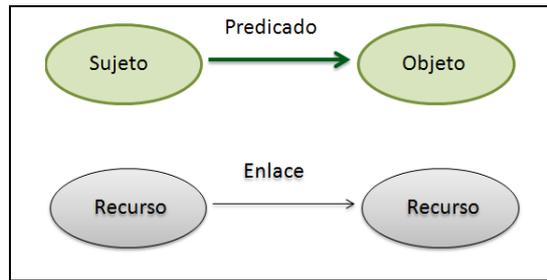


Fuente: Elaboración propia

La interpretación completa del triplete podría ser: la página web de la biblioteca cuya URL es <http://www.biblioteca.es/migueldecervantes> es sobre el autor (elemento *creator* del vocabulario DC) llamado Miguel de Cervantes Saavedra. Con esta forma de descripción se define tanto la estructura (sintaxis) de la información como el significado (semántica), es decir, queda expresado el tipo de recurso y el tipo de relación que mantiene con otros recursos de información (Castells, 2003).

La siguiente imagen diferencia la representación gráfica de dos grafos formados por nodos, el primero representa la estructura de la información descrita con RDF, y el segundo representa la estructura de la información de un documento descrito en HTML.

Ilustración 26: Representación gráfica de un triplete y representación gráfica de un enlace hipertextual



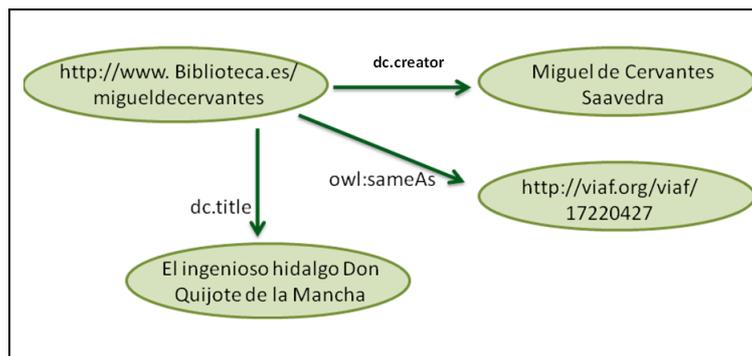
Fuente: Modificado de Cyganiak, Richard; Wood, David; Lanthaler, Markus. *RDF Concepts and abstract Syntax*, 2014 <<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>>

El primer grafo es un triplete RDF formado por un nodo sujeto (el recurso), por un nodo objeto (el valor) y el arco que los enlaza, el predicado (la relación identificada que existe entre los nodos). El segundo nodo es un conjunto de elementos descritos en lenguaje HTML, donde los nodos son del mismo tipo (recursos) y no están diferenciados con más información. El arco es un enlace HTTP a otro documento. Este enlace hipertextual o referencia cruzada automática no especifica el tipo de relación que existe entre los dos recursos, sino que se trata de un vínculo que conduce de un recurso a otro (Castells, 2004). Tal y como muestra el gráfico se observa cómo la relación ha variado sustancialmente. El grafo RDF evidencia que ambos nodos tienen valores identificados que los programas informáticos pueden interpretar; en cambio, en el grafo HTML los nodos no están definidos de forma que los programas puedan

interpretar el tipo de información que contienen. Por tanto son únicamente dos recursos enlazados.

RDF permite con ello crear estructuras agregadas, es decir, la información sobre un recurso puede ser ampliada añadiendo estructuras de tripletes descritas con este mismo lenguaje. Su estructura y composición permite utilizar otros modelos de vocabularios y combinarlos, para definir así las relaciones entre los elementos sujeto y objeto (Pastor, 2011). La siguiente imagen muestra de forma gráfica cómo se amplía información sobre el recurso. Cada nuevo dato se define en RDF y se visualiza como un nuevo nodo enlazado al recurso mediante elementos de distintos vocabularios como DC y OWL.

Ilustración 27: Grafo compuesto por tres tripletes RDF



Fuente: Elaboración propia

El predicado (relación) de la imagen 24 `<http://purl.org/dc/elements/1.1/creator>` indica a través del elemento *creator* del vocabulario DC, que la relación existente

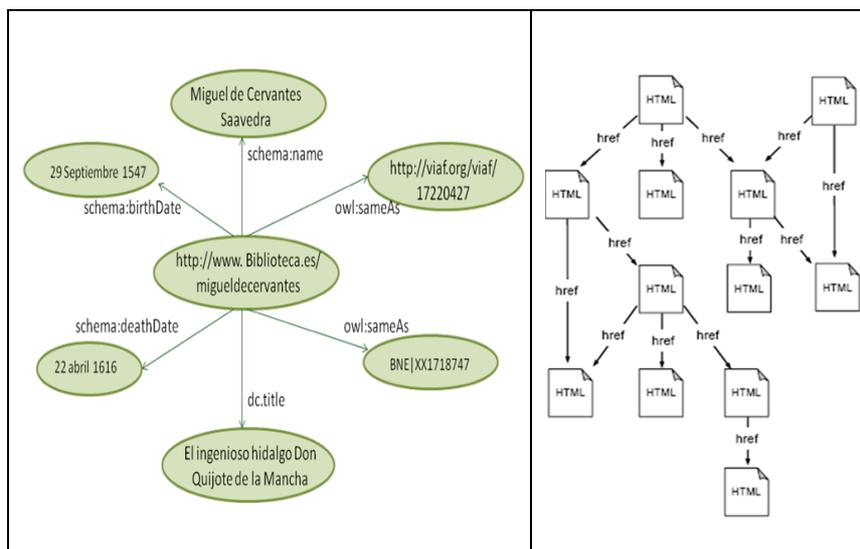
es de “autor o creador”. En el gráfico de la imagen 25 este predicado se representa mediante el prefijo *dc* para abreviar el namespace `<http://purl.org/dc/elements/1.1/>`, y el elemento nombre cualificado (*Qname*):*creator* en lugar de la URI completa. Se trata de una equivalencia que identifica de forma exacta los elementos y los vocabularios sin posibilidad alguna de ambigüedad semántica. Es un mecanismo importante cuando se emplean diferentes vocabularios con elementos muy similares en cuanto a significado (Pastor, 2011).

Por tanto, todos los mecanismos que emplea el lenguaje RDF para describir e identificar elementos informativos se orientan a eliminar posibles ambigüedades semánticas e identificar unívocamente los objetos digitales y sus relaciones. Frente a los escasos mecanismos de control de ambigüedad o vacío semántico que tiene por sí mismo el lenguaje HTML. Esta especificidad que aporta el lenguaje RDF en la descripción de los objetos digitales repercute en la interpretación del contenido que pueden hacer las aplicaciones y buscadores semánticos sobre los recursos. Así pues, una página web escrita con HTML se visualiza a través de un navegador web que interpreta el código de forma comprensible para los usuarios. El software que realiza la presentación visual de los recursos interpreta las etiquetas HTML que describen la estructura y la apariencia del documento, y muestra el contenido a los usuarios a través de la interfaz. Pero este lenguaje no permite al software comprender

el significado de la información que muestra porque no contiene información suficiente sobre el contenido informativo. Solo los usuarios que navegan son capaces de interpretar el significado de la información que se muestra. Como resultado, la información que proporcionan estas páginas web escritas en HTML a los motores de búsqueda no permite hacer búsquedas basadas en el significado de la información sino en el reconocimiento de cadenas de caracteres (Caro, 2012).

Este proceso de caracterización y desambiguación que proporciona RDF marca la diferencia esencial de la web de LD con la web tal y como se conoce hasta el momento (Agenjo y Hernández, 2010). A continuación se detalla gráficamente cómo se visualizan ambas webs. La ilustración 28 presenta dos imágenes. La imagen de la izquierda representa el grafo de la web semántica, donde los recursos digitales están descritos con lenguaje RDF. Los nodos y las relaciones contienen metadatos que describen el tipo de información que contiene el recurso, de una forma comprensible para los software. La imagen derecha representa el grafo de una página web escrita en lenguaje HTML donde las relaciones entre los documentos son únicamente enlaces que redireccionan, sin aportar más información sobre el contenido o el tipo de documento.

Ilustración 28: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo)



Fuente: Modificación basada en Castells, Pablo. La web semántica, 2003

Este avance es clave en el ámbito de la recopilación, difusión y recuperación de la información, y por tanto ofrece una ventaja competitiva a los productos de información que son compatibles con la web semántica, tal y como lo es Vestigium. La web que describe su información con lenguaje RDF, permite al programa o buscador semántico interpretar semánticamente el contenido de un recurso. Este lenguaje describe la información de forma que el programa es capaz de identificar el tipo de recurso, sus valores y la relación que existe entre los elementos. De este modo los buscadores pueden identificar el tipo de información del recurso digital, el significado de los datos que lo componen

y las relaciones que existen entre ellos (Heath y Bizer, 2011; Byrne y Goddard, 2010). La estructura del lenguaje RDF describe el recurso informativo diferenciando el propio objeto, aportando en el predicado la información que permite identificar el tipo de objeto del que se trata, y el valor que tiene. Por esta razón, los vocabularios y las ontologías que se mencionaron en el capítulo anterior cobran un peso clave proporcionando y fijando el significado de los recursos en contextos muy diferentes, tal y como vamos a ver a continuación.

Vocabularios y ontologías para describir datos

Hemos visto que la definición de los recursos mediante tripletes RDF proporciona una descripción abstracta de la información del recurso que resulta comprensible para los programas. Esto tiene diversas aplicaciones al favorecer la recuperación de la información de forma automatizada mejorando las prestaciones de los motores de búsqueda, optimizando la catalogación en bibliotecas digitales al especificar las relaciones de contenido, y al facilitar el intercambio de información legible por ordenador a través de la web (Agenjo y Hernández, 2011). RDF proporciona la estructura de definición, pero además es necesario definir un dominio específico (*namespace*) en el que describir, clasificar y expresar las relaciones de los objetos del mundo real con otros elementos. Esta función de organización y de descripción la desarrollan las herramientas de control del vocabulario y de organización del conocimiento.

Los lenguajes documentales han sido importantes herramientas de control en los procesos de indización y de recuperación de información que se emplean para estandarizar el lenguaje natural y solucionar los problemas de ambigüedad. Problemas que las personas son capaces de deducir a partir del contexto. Para un procesamiento automático es necesaria una información adicional que no está explícita en los textos, y que las

herramientas de control y las estructuras de conocimiento aportan (Caro, 2012). Las primeras manifestaciones de lenguajes documentales datan de finales del siglo XIX con la aparición de las clasificaciones bibliográficas, pero el concepto moderno se consolidó en el siglo XX con los encabezamientos de materia de Charles A. Cutter en *Rules for a Dictionary Catalog* (1904) (Gil Urdiciaín, 1996). El lenguaje documental es “un lenguaje convencional que utiliza una unidad de información para describir el contenido de los documentos con miras a su almacenamiento y a la recuperación de la información” (Guinchat y Menou, 1983).

Blanca Gil Urdiciaín (1996) define lenguaje documental como “todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios. Es un lenguaje no natural, que utiliza signos de éste, que adquieren un valor semántico por medio de su normalización y de las reglas morfosintácticas que lo articulan”.

El uso de herramientas de control terminológico ayuda a evitar las imprecisiones del lenguaje natural (sinonimia, polisemia, homonimia, antonimia) ordenando los conceptos mediante términos que se relacionan entre sí de forma sintáctica (Moreiro, 2013). Existe una extensa variedad: los índices, los esquemas de

conceptos, las listas de encabezamientos de materia, las clasificaciones, las nomenclaturas geográficas, las taxonomías, los tesauros. Cada tipo de lenguaje documental cuenta con unas características diferentes que se emplean para funciones diversas. Existen diversas formas de clasificar los lenguajes documentales a partir de diversas perspectivas (García Gutiérrez, 1999):

- Según la función de control ejercida sobre el vocabulario:
 - Vocabularios libres (listas de descriptores).
 - Vocabularios controlados (clasificaciones, tesauros).
- Según el momento de la composición:
 - Precoordinados: combinan los términos en el momento de la descripción (clasificaciones, listas de encabezamientos de materias).
 - Postcoordinados: combinan los términos en el momento de la recuperación (listas de descriptores libres, tesauros).
- Según el tipo de estructura:

- Jerárquica: adopta forma de árbol y las relaciones se establecen mediante criterios hiper e hiponímicos (clase/especie) y partitivos (todo/parte) (clasificaciones).
- Asociativos: el lenguaje se estructura horizontalmente y desaparece el criterio de dependencia terminológica.
- Según extensión temática:
 - Lenguajes enciclopédicos: abarcan todas las áreas del conocimiento.
 - Lenguajes especializados: se centran en un sector del conocimiento.

Los vocabularios de control tienen tres objetivos principales (NISO, 2005). En primer lugar controlan los sinónimos, pues en el lenguaje natural un concepto puede ser representado por varios términos con el mismo o similar significado. El vocabulario controlado asegura que cada concepto se represente por un único término preferente, es decir, designa las formas autorizadas para representar conceptos, personas, lugares, instituciones, en un registro bibliográfico, y establece un listado con los sinónimos y variantes como términos no preferentes. En segundo lugar, estos lenguajes eliminan la ambigüedad ya que en el lenguaje natural un término puede tener más de un

significado, con el uso de vocabularios controlados se asegura que cada término tenga un solo significado. Y por último, establecen las relaciones entre términos, identificando las que puedan existir entre los términos.

En la gestión convencional de la información se han empleado mayoritariamente clasificaciones y encabezamientos de materia, es decir, lenguajes precoordinados en los que el hombre construía el concepto. La evolución de las tecnologías web junto a otras necesidades ha llevado a aplicar herramientas conocidas de organización del conocimiento como clasificaciones y tesauros y a ampliar el uso de otras, como las ontologías, que contextualizan los conceptos dentro de un campo específico del conocimiento y establecen relaciones semánticas de equivalencia, jerarquía y asociación entre los términos. Los sistemas de organización del conocimiento evolucionan y se amplían adquiriendo mayores niveles de abstracción, por lo que se ha dado un paso más en la progresión de los términos a los conceptos y en sus relaciones, como se puede observar en taxonomías, tesauros y ontologías (Moreiro, 2013).

Las taxonomías son estructuras clasificatorias que agrupan los conceptos respondiendo a algún criterio que relaciona los elementos entre sí y los separa de los que semánticamente no están relacionados. Los categoriza jerárquicamente estableciendo niveles de generalización y especificación de los

elementos con un grado mayor de abstracción que las clasificaciones (Currás 2005). Emilia Currás (2005) las define como: “Una lista organizada de palabras y frases, o sistemas de notación que se usa para iniciar un proceso de indización y recuperación de información. Se organiza en niveles controlados”, “Un esquema de navegación ordenado jerárquicamente”, así como “Estructuras conceptuales adecuadas para su uso en las webs semánticas”.

Por su parte, los tesauros son vocabularios de términos relacionados que cubren un dominio específico del conocimiento. Consignan las formas autorizadas y las alternativas de los conceptos y establecen las relaciones jerárquicas, asociativas, de equivalencia y semánticas entre los términos. Cuentan con un grado mayor de abstracción que las taxonomías (Currás 2005). Los tesauros han sido utilizados en contextos muy variados desde sus orígenes en el mundo analógico. Para la ISO (1983) son “vocabularios controlados de términos que tienen entre sí relaciones semánticas para facilitar su indización y su organización”. Slype et al., (1991) añaden que representan sin ambigüedades el contenido de los documentos y las consultas “dentro de un sistema documental determinado”.

Finalmente, con el desarrollo tecnológico de nuevas herramientas informáticas en las últimas décadas, se originan las ontologías, principalmente para ámbitos como la información y

la comunicación. Estos cambios han producido el uso de nuevos términos en áreas de ciencias de la información, como el término ontología, que se utiliza como sinónimo de lenguaje documental. Tienen una importante proximidad con los sistemas clasificatorios, particularmente por su estructura basada en la clasificación jerarquizada de las taxonomías (Sánchez-Cuadrado et al., 2007). Pero las ontologías constituyen un sistema de términos y relaciones que se expresan mediante un conjunto de axiomas y hechos con un lenguaje lógico que puede ser entendido por un ordenador. Se emplean para la representación del conocimiento formal en la web y proporcionan un mayor grado de abstracción.

Vamos a ver detenidamente los instrumentos que pueden ser aplicables al proyecto de biblioteca semántica que propone este trabajo.

Fundamentos de las ontologías

El concepto de ontología tiene su origen en la filosofía clásica, aunque se ha adoptado en el área de la ingeniería del software y en inteligencia artificial. Sus definiciones han variado en el tiempo según el área de conocimiento en la que se emplea, y actualmente se aplica en el contexto de la web semántica (Pedraza-Jiménez et al., 2007). Para Sánchez-Cuadro, et al.

(2007) las definiciones más citadas en este contexto son varias: Thomas R. Gruber en 1993 afirmó que “una ontología constituye una especificación explícita y formal de una conceptualización compartida”; Rudi Studer et al. en 1998 y diez años después Vilches-Blázquez et al. dicen que son “un modelo abstracto de la realidad en el que todos sus componentes deben estar definidos explícitamente, debe ser entendible por los ordenadores y debe capturar el conocimiento consensuado por un grupo o comunidad de expertos”. Para Guarino (1998) son el “producto de ingeniería consistente en un vocabulario específico usado para describir una realidad más un conjunto de asunciones relacionadas con el significado del vocabulario”. La definición más amplia y más aclaratoria es la de Pastor-Sánchez (2009), que expresa: “una ontología web realiza descripciones de objetos definiendo clases, propiedades, relaciones y axiomas. [...] desarrollan estructuras complejas para simplificar tareas de comunicación entre personas y máquinas con la intermediación de aplicaciones informáticas, que realizan procesos de intercambio de datos y simplifican la interoperabilidad de diferentes representaciones de datos con herramientas de unificación, traducción y mapeado”.

Por tanto, el conjunto de definiciones evidencia que se trata de una herramienta para la organización del conocimiento en un dominio específico, que se expresa mediante lógica formal con

una capacidad de descripción muy precisa que se implementa mediante programas informáticos.

Es decir, las ontologías clasifican en categorías los elementos del mundo real definiéndolo mediante lenguaje matemático (lógica descriptiva). Una ontología está formada por una taxonomía de conceptos y por un conjunto de axiomas o reglas de inferencia. Los axiomas son especificaciones de base lógica-matemática que describen los objetos del mundo real y cómo éstos se relacionan (Codina y Pedraza-Jiménez, 2011). Es decir, los axiomas son enunciados que se presuponen verdaderos sin necesidad de demostración, describen los objetos y sus relaciones en lenguaje matemático. Un ejemplo de axioma puede ser el siguiente: si el concepto A es mayor que el concepto B, y B es mayor que el concepto C, entonces A también es mayor que C. Este tipo de enunciados constituyen reglas o afirmaciones verdaderas para las aplicaciones informáticas y les permiten manejar los elementos y sus relaciones deduciendo conclusiones lógicas (Rodríguez Perojo y Ronda León, 2011).

A continuación, en la ilustración 29 se muestra un ejemplo de axioma clase del lenguaje OWL. En este ejemplo se define la clase Ópera Tradicional Italiana como subclase de Ópera. Y los tipos de operas con varios valores: Ópera Seria y Ópera Buffa (Bechhofer et al., 2004).

Ilustración 29: Ejemplo de axioma clase en lenguaje OWL

```
<owl:Class rdf:ID="TraditionalItalianOpera">
  <rdfs:subClassOf rdf:resource="#Opera"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasOperaType"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:oneOf rdf:parseType="Collection">
            <owl:Thing rdf:about="#OperaSeria"/>
            <owl:Thing rdf:about="#OperaBuffa"/>
          </owl:oneOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Fuente: Bechhofer et al., OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004 <<http://www.w3.org/TR/owl-ref/#ClassAxioms>>

Los componentes de las ontologías han sido ampliamente estudiados (Pedraza-Jiménez et al., 2007; Pastor-Sánchez, 2011; González-Crespo et al., 2012). Pueden estructurarse de la siguiente manera:

- Clases: son los conceptos que se organizan en taxonomías, es decir clasificados. Son entidades del mundo real y pueden ser objetos o conceptos, por ejemplo, un libro o una teoría científica. Además, una clase puede tener subclases que por lo general heredan las propiedades de la clase a la que pertenecen. Normalmente la relación de subordinación más empleada en estas estructuras jerárquicas establece

tipologías de esa clase. Por ejemplo, la clase libro puede dividirse en las subclases obra de referencia, manual, libro de texto.

- Instancias: es un ejemplar particular de cada una de las clases, es decir, un ejemplar de libro o una teoría científica específica.
- Atributos: son las propiedades que describen las características de los conceptos y permiten asignarles valores concretos que consisten en pares del tipo atributo/valor. Describen las características de las entidades que forman las clases. Por ejemplo, en la ontología RDF, una propiedad de un recurso de información es “tipo de recurso” que se define mediante la etiqueta `rdf:type` (Brickley y Guha, 2014). En Vestigium la propiedad “tipo de recurso” tendría un valor de “persona”.
- Relaciones: son un tipo de interacción, un enlace entre conceptos de un mismo dominio, de forma que los conceptos forman estructuras y las relaciones establecen el valor entre los conceptos. Los tipos de relaciones existentes son:
 - Clase-instancia: esta relación asocia un ejemplar (instancia) con la clase a la que pertenece. Es una relación de tipo “es un”. Por ejemplo, “El Mito de

Vesalio-es un-libro”. De modo que El Mito de Vesalio es una instancia de la clase libro.

- Instancia-propiedad: las instancias (ejemplares) toman valores asociados a sus características o atributos. Esta relación es de tipo *tiene el valor*, por ejemplo, la instancia “obra de referencia-tiene el valor-145” para el atributo (propiedad) “extensión”.
- Clase-propiedad: la clase tiene propiedades que comparten todos sus atributos por pertenecer a ese conjunto. Se denominan restricciones que excluyen a otros atributos de pertenecer a esa clase. La clase “libro” tiene la propiedad de “tiene capítulos” que excluye de esta clase a las revistas.
- Clase-subclase: las clases pueden tener subclases y la relación de subordinación existente entre ellas es de tipo “es un”. Por ejemplo, una novela es una subclase de libro.
- Funciones: es un tipo de relación que se establece a través del cálculo de una función de varios elementos de la ontología.
- Axiomas: son expresiones verdaderas, condiciones impuestas, que siempre deben cumplir los elementos de

la ontología, y que sirven para inferir nuevo conocimiento de forma automatizada.

Por ejemplo, “si A y B pertenecen a la clase C, entonces A no puede ser subclase de B”. El proceso de generar nuevo conocimiento de forma automática se basa en extraer conclusiones a partir de la información en forma de axiomas, sentencias descritas en lógica de primer orden (LPO), o cálculo de predicados de primer orden (CPPO), que establecen que determinadas clases cumplen ciertas condiciones (Pastor Sánchez, 2009).

A continuación, en la ilustración 30, se muestra un ejemplo de clases y subclases de la ontología OWL y la representación gráfica de los conceptos y las relaciones jerárquicas de éstos (Pedraza-Jiménez et al., 2007).

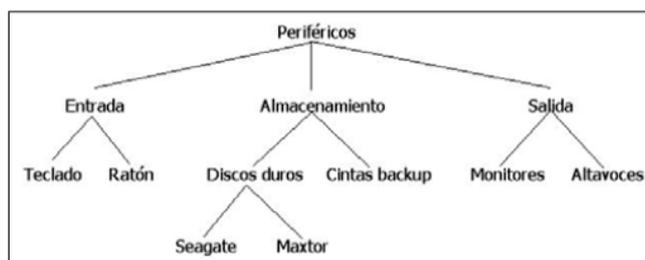
Ilustración 30: Ejemplo de una ontología OWL

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
...
<owl:Class rdf:ID="perifericos">
<rdfs:comment>
Los periféricos de ordenador están conectados a la CPU pero no forman parte de
ella
</rdfs:comment>
</owl:Class>
<owl:Class rdf:ID="entrada">
<rdfs:comment>
Los periféricos de entrada son una subclase de periféricos de ordenador.
</rdfs:comment>
<rdfs:subClassOf rdf:resource="#perifericos" />
</owl:Class>
<owl:Class rdf:ID="teclados">
<rdfs:comment>
Los teclados son una subclase de los periféricos de entrada.
</rdfs:comment>
<rdfs:subClassOf rdf:resource="#entrada" />
<rdfs:subClassOf rdf:resource="#perifericos" />
</owl:Class>
...
</rdf:RDF>
```

Fuente: Pedraza-Jiménez, Rafael; Codina, Lluís; Rovira, Cristòfol. "Web semántica y ontologías en el procesamiento de la información documental". En: El profesional de la información, 2007, noviembre-diciembre, v. 16, n. 6, pp. 569-578 <<http://eprints.rclis.org/14298/1/webSemanticaOntologias2007.pdf>>

La representación gráfica de esta ontología es una estructura jerárquica que adopta forma de pirámide o árbol como se muestra en la ilustración 31.

Ilustración 31: Representación gráfica de la ontología owl de la ilustración 30



Fuente: Pedraza-Jiménez, Rafael; Codina, Lluís; Rovira, Cristòfol. "Web semántica y ontologías en el procesamiento de la información documental". En: El profesional de la información, 2007, noviembre-diciembre, v. 16, n. 6, pp. 569-578 <<http://eprints.rclis.org/14298/1/webSemanticaOntologias2007.pdf>>

Además de las relaciones mencionadas, hay otras relaciones de sinonimia, hiponimia, antonimia y meronimia entre conceptos (clases), que son bien conocidas en el ámbito del control del vocabulario.

Las ontologías, por tanto, ofrecen un modelo de datos para representar la estructura y el contenido de los vocabularios controlados. Para representarlas, para concretar estos niveles abstractos, existen varios lenguajes. Difieren entre ellas por los niveles de representación del conocimiento y las limitaciones de expresividad para el diseño de algoritmos de razonamiento automático (Heath y Bizer, 2011). Los lenguajes de definición de ontologías que han alcanzado una mayor difusión y uso en el entorno de la web semántica y que cuentan con distintos grados de expresividad son:

- *RDF Schema* (RDFS)
- *Web Ontologic Language* (OWL)
- *Simple Knowledge Organization System* (SKOS)

RDFS y OWL proporcionan una base para crear los vocabularios y describir las entidades del mundo real y sus relaciones, mientras que SKOS es un modelo para expresar sistemas de organización del conocimiento (Caro, 2012), como veremos a continuación.

Lenguajes de definición de ontologías

RDF Schema (RDFS)

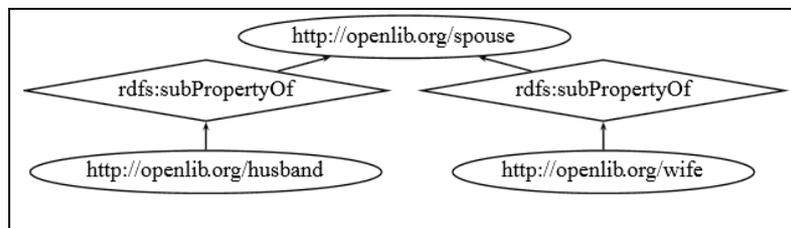
RDF *Schema* constituye un mecanismo cuya principal función es describir las ontologías simples. Amplía el lenguaje RDF para proporcionar un vocabulario que indique de forma jerárquica los recursos, las clases de recursos y las subclases, sus atributos, así como las restricciones de las propiedades y las relaciones entre ellos (Heath y Bizer, 2011; Pastor-Sánchez, 2011). Sin embargo, no permite declarar axiomas para inferir conocimiento (Peis et al., 2003). En el apartado anterior se han visto los elementos de las ontologías en general, en particular para RDFS son los siguientes:

- Las clases (rdfs:Class) constituyen el tipo de recurso y se describen mediante las propiedades RDF. Pueden representar cualquier tipo de recurso, desde una persona hasta un tipo de documento. Por ejemplo, “clase: persona”
- Las propiedades (rdf:Property) describen el tipo de relación entre el sujeto y un objeto. Es el predicado. Por ejemplo, “propiedad: nombre”

- Las restricciones de propiedades definen como se aplican las propiedades y qué valores pueden tomar. Por ejemplo, “valor: Juan”

RDFS desarrolla clases y propiedades para los recursos. Con RDFS se proporciona un vocabulario a RDF que permite describir los recursos mediante clases y subclases y propiedades (Krichel, 2002). Por ejemplo, a continuación el gráfico (ilustración 32) muestra la definición de dos propiedades (*husband* y *wife*) para la clase *spouse*.

Ilustración 32: Gráfico jerárquico con la definición de propiedades de la clase *spouse*



Fuente: Krichel, Thomas *The Semantic web and an introduction to RDF.*, 2002
 <<http://hdl.handle.net/10760/4214>>

El lenguaje RDFS se describe al completo en el documento *RDF Vocabulary Description Language 1.0: RDF Schema* (Brickley y Guhna, 2003) y en *RDF Schema specification* (Brickley y Guhna, 2014). Es un lenguaje sencillo que se complementa con un lenguaje más genérico que veremos a continuación, el OWL (*Web Ontologic Language*) (Pastor-Sánchez, 2011).

Web Ontology Language (OWL)

Web Ontology Language (en adelante OWL) es el otro lenguaje de representación de ontologías, basado en lógica descriptiva, con el que es posible inferir deducciones lógicas en función del nivel de expresividad y complejidad. Es un estándar propuesto por el W3C, para que los programas informáticos puedan interpretar relaciones entre documentos, que permite definir clases, propiedades y relaciones. Su función es dotar de significado las tripletas de RDF especificando a qué tripletas se les asigna un significado, y cuál es (Codina et al., 2009). Está compuesto por tres subconjuntos de lenguajes que se caracterizan por tener distintos niveles de expresividad para representar el conocimiento: OWL-Lite, OWL LD y OWL Full. Se asemeja a RDFS, pero es mucho más preciso y permite crear ontologías a un nivel mucho más detallado (Smith et al., 2004; Heath y Bizer, 2011).

Está diseñado para ser utilizado en aplicaciones que procesan automáticamente el contenido de la información en lugar de únicamente representar información para las personas. OWL complementa y amplía RDF y RDFS, ya que permite la expresión de relaciones semánticas más complejas entre diferentes clases RDFS, y mayor precisión en las restricciones de clases y propiedades específicas (Smith et al., 2004).

Un ejemplo de esta mayor precisión de descripción frente a RDFS podría ser, definir recursos para determinar si todos los miembros de una clase tendrán una propiedad en particular, o si puede ser que sólo algunos la tengan.

En el apartado anterior se han visto los elementos de las ontologías en general, en particular, los elementos del lenguaje OWL son las clases, sus atributos, las instancias, y las propiedades. Las clases son los conceptos, es decir, los objetos de información. Se expresan mediante clases generales y subclases específicas. Por ejemplo, clase “persona” y subclase “escritor”. Los atributos de los objetos de información son sus características, como por ejemplo, “lugar de nacimiento”. Las instancias de clases son los valores que alcanzan, por ejemplo, “Miguel de Cervantes”. Las propiedades de los objetos de información son las relaciones.

El lenguaje de descripción de ontologías OWL utiliza el marco RDF para representar las ontologías (Pedraza et al., 2007), como se puede ver en el siguiente ejemplo de una ontología sobre vino. En este ejemplo se representa la clase “vino”, una subclase del recurso “vino” es “comida y líquido potable”, y el atributo “idioma” proporciona una variante del nombre. Este atributo está duplicado, y tiene dos valores “wine” y “vin” (Smith et al., 2004).

Ilustración 33: Ejemplo de un extracto de lenguaje owl para representar una clase y una subclase de una ontología sobre vino

```
<owl:Class rdf:ID="Wine">
<rdfs:subClassOf rdf:resource="&food;PotableLiquid"/>
<rdfs:label xml:lang="en">wine</rdfs:label>
<rdfs:label xml:lang="fr">vin</rdfs:label>
</owl:Class>
```

Fuente: Smith, Michael K.; Welty, Chris; McGuinness, Deborah L. *OWL Web Ontology Language Guide*, 2004 <<http://www.w3.org/TR/owl-guide/>>

La especificación del lenguaje OWL se recoge en tres documentos fundamentales de 2004 del W3C entre otros: *OWL Web Ontology Language Semantics and Abstract Syntax*, *OWL Web Ontology Language Guide* y *OWL Web Ontology Language Reference*.

El lenguaje OWL Full también se emplea para representar sistemas de organización del conocimiento, como un tesoro o una taxonomía con la estructura y las relaciones semánticas de contenido, entonces se denomina SKOS (*Simple Knowledge Organization System*). Es un modelo de descripción que utiliza OWL Full y está basado en RDF para identificar los conceptos que forman el sistema de organización del conocimiento mediante URI (Pastor-Sánchez et al., 2012), como veremos a continuación.

Simple Knowledge Organization System (SKOS)

Simple Knowledge Organization System (SKOS en adelante) es un modelo de datos para definir e implementar sistemas para representar el conocimiento de forma estructurada. Es decir, permite migrar los sistemas tradicionales de organización del conocimiento tales como tesauros, esquemas de clasificación, glosarios, listas de encabezamiento, taxonomías, y cualquier vocabulario controlado, al formalismo lógico de los lenguajes de ontologías de la web semántica. De esta forma se hace posible su reutilización e interoperabilidad (Isaac y Summers, 2009; Sánchez-Cuadrado, et al., 2012; Pastor-Sánchez et al., 2012). Desde 2009 es una recomendación del W3C, aunque sus orígenes se encuentran en el grupo de trabajo SWAD-Europe hacia el año 2002. Se difundió públicamente mediante un borrador en noviembre de 2005 (Sánchez-Cuadrado et al., 2012; Pastor-Sánchez et al., 2012).

Como hemos comentado, consiste en una ontología OWL para expresar un amplio abanico de esquemas conceptuales o tesauros y admite relaciones de términos más generales y más específicos así como etiquetas aceptadas y alternativas (Agenjo y Hernández, 2011). El lenguaje de formalización o codificación de la información es, de nuevo, RDF, por lo que observamos que estructura los datos en forma de tripletas. Su base, como vimos, es OWL, pero puede ser combinado y extendido con

otros lenguajes formales de representación del conocimiento. Por ejemplo, admite la incorporación de los vocabularios de metadatos Dublin Core o FOAF (Sánchez-Cuadrado et al., 2012) para completar informaciones.

Los elementos de un vocabulario determinado se representan con conceptos, y mediante el uso de etiquetas se identifican las relaciones semánticas entre los conceptos, que pueden ser jerárquicas (simples o transitivas) y asociativas. Las etiquetas son las expresiones que se utilizan para referirse a los conceptos en lenguaje natural formal, igual que hacen los tesauros. Cada etiqueta expresa un estado, y se utilizan para asociar un concepto con una cadena de caracteres RDF (Pastor-Sánchez et al., 2012). En el ejemplo siguiente (ilustración 34) se puede ver el concepto mamíferos y mediante la etiqueta `skos:broader` se define la relación jerárquica de concepto más amplio o genérico.

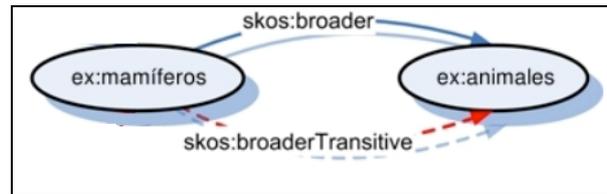
Ilustración 34: Relación jerárquica entre dos conceptos de un sistema de organización del conocimiento en SKOS

```
ex:mamíferos rdf:type skos:Concept;  
skos:prefLabel "mamíferos"@es;  
skos:broader ex:animales.
```

Fuente: Manual SKOS <<http://skos.um.es/TR/skos-primer/>>

La ilustración 35 muestra de forma gráfica la relación jerárquica entre los dos conceptos expresados en la ilustración 34.

Ilustración 35: Representación gráfica de la etiqueta de relación jerárquica entre dos conceptos de un sistema de organización del conocimiento



Fuente: Pastor-Sánchez, Juan Antonio. SKOS Simple Knowledge Organization System: Qué es skos? <<http://skos.um.es/acerca/index.php>>

Veamos en la siguiente tabla cómo en lenguaje SKOS se solucionan las relaciones de sinonimia, es decir, cómo se identifican los descriptores escogidos. Al igual que en un tesoro, se asignan a cada recurso las etiquetas preferentes para identificar el término descriptor, y etiquetas alternativas para los términos no preferentes: “skos:prefLabel”, etiqueta empleada para términos preferentes (descriptores en la terminología de tesoro); “skos:altLabel”, empleada para términos no preferentes.

Tabla 6: Ejemplo de etiquetas SKOS para representar el término preferente y el término no preferente de un concepto

```
ex:animals rdf:type skos:Concept;
skos:prefLabel "animals"@en;
skos:altLabel "creatures"@en;
skos:prefLabel "animaux"@fr;
skos:altLabel "créatures"@fr.
```

Fuente: Isaac, Antoine; Summers, Ed. *SKOS Simple Knowledge Organization System Primer*, 2009 <<http://www.w3.org/TR/skos-primer/>>

La etiqueta alternativa, asignada a varios términos proporciona diversas entradas a un mismo concepto en un vocabulario. Sería equivalente al término relacionado en un tesoro y se emplea para controlar sinónimos, abreviaturas y acrónimos (Isaac y Summers, 2009).

La definición completa del lenguaje se recoge, entre otros, en dos documentos del W3C de 2009: *SKOS Simple Knowledge Organization System Primer* y *SKOS Core Vocabulary Specification*.

En definitiva, con los lenguajes de representación de ontologías se consigue adaptar las herramientas de representación del conocimiento empleadas tradicionalmente en el ámbito bibliotecario a la representación de información en la web semántica. En la siguiente tabla se resumen las tecnologías web, los lenguajes de representación de metadatos y las ontologías que hemos revisado y que se emplean en el ecosistema de los datos enlazados para el intercambio de datos estructurados.

Tabla 7: Tecnologías de la web semántica

Protocolo de identificación y recuperación de recursos en la web.	HTTP/URI
Lenguaje de marcado estándar para el intercambio de información estructurada.	XML
Lenguaje para la descripción de metadatos (descripción semántica de los recursos).	RDF
Lenguajes para la representación de Ontologías.	OWL / RDFs
Modelo para representar sistemas de organización del conocimiento.	SKOS
Descripción del contenido de los recursos mediante especificaciones comunes.	Metadatos
Asignar valores a los elementos de metadatos que describen el recurso de información.	Vocabularios de valores

Fuente: Elaboración propia a partir de Berners-Lee, Tim. Semantic web on xml, 2000 <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>> y de Pastor-Sánchez, Juan Antonio. Tecnologías de la web semántica. Barcelona: editorial UOC, 2011

Como hemos visto, en primer lugar se encuentran los protocolos de transferencia de información y los de identificación de recursos digitales. Seguidamente el lenguaje XML, como se ha dicho, es la base para el desarrollo de diferentes lenguajes de la web semántica. A continuación, el lenguaje RDF es la base y establece la estructura para la descripción semántica de la

información. A partir de entonces son necesarios los lenguajes RDFs y OWL para la descripción de ontologías y reglas de inferencia. Por último, SKOS sirve como modelo de representación de sistemas de organización del conocimiento.

Los elementos resultantes para la descripción de los recursos de información son los vocabularios de metadatos y, los vocabularios de valores para asignar valores a los campos de metadatos que describen el objeto digital, como veremos a continuación. En ellos convergen las tecnologías web junto con las herramientas de descripción y de control procedentes de la gestión bibliotecaria que vimos en apartados anteriores.

Vocabularios de valores

El desarrollo de modelos de metadatos, vocabularios controlados y sistemas de organización del conocimiento (también conocidos como KOS: *knowledge organization systems*) se han empleado históricamente en el sector de la gestión de información para indexar y mejorar la recuperación. Ahora, aplicados en el área de los datos enlazados en la web, mejoran la descripción de los recursos digitales y han evolucionado hacia la idea de compartir conocimiento de forma automática, pues los sistemas tradicionales requerían un alto

componente de intervención humana (Sánchez-Cuadrado et al., 2012).

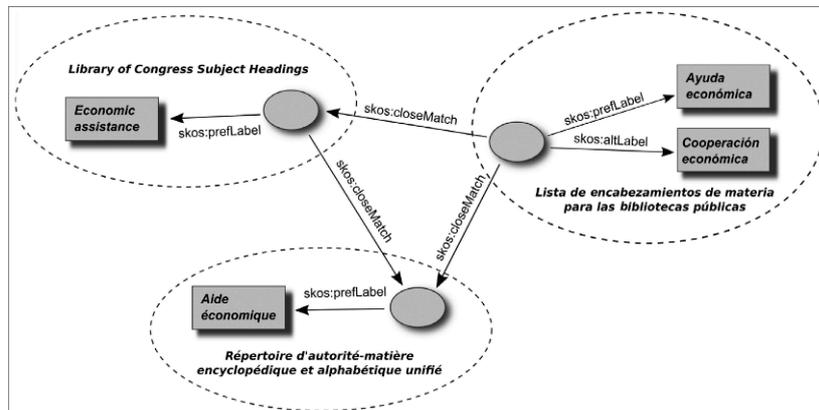
Es incuestionable que utilizando vocabularios para asignar valores a los campos de metadatos se consigue eliminar ambigüedades derivadas de la sinonimia del lenguaje natural, y al mismo tiempo, se contextualiza semánticamente cada concepto inscribiéndolo en un campo del conocimiento en el que se vinculará con otros conceptos mediante relaciones jerárquicas o asociativas (Caro, 2012).

Los conjuntos de elementos de metadatos y los vocabularios de valores son imprescindibles para la correcta interpretación de los datos por parte de los programas informáticos y de los motores de búsqueda semánticos. Aportan el valor a los campos de descripción del objeto digital. Hemos visto que se publican como vocabularios RDF, ontologías OWL o SKOS (en adelante Vocabularios RDF que incluye los tres conceptos). Como ejemplos de vocabularios controlados formalizados con SKOS cabe citar Agrovoc, Eurovoc o LCHS (Sánchez-Cuadrado et al., 2012), como veremos más adelante. Estas herramientas KOS establecen una relación semántica entre los elementos informativos del objeto digital, de forma equivalente a cómo se relacionan los elementos del registro bibliográfico mediante el uso de tesauros, índices de materia y listas de autoridades. Se emplean de forma paralela varios vocabularios según las características particulares de los objetos que se describen, ya

que se considera una buena práctica la reutilización de términos de los vocabularios ya definidos en RDF (Bizer et al., 2009).

Por tanto, los vocabularios de valores se utilizan para asignar valores a los elementos de metadatos. Es decir, que sirven para llenar de contenido los metadatos que describen el recurso de información. Así, un vocabulario es una lista de los valores permitidos para un determinado elemento. Por ejemplo, para asignar el valor del elemento de metadato “materia” se utiliza un vocabulario como la lista de encabezamientos de materia. En la siguiente imagen se pueden ver tres ejemplos de vocabularios de valores representados mediante SKOS, las listas de encabezamientos de materia LCSH, LEM y RAMEAU, que se describen más adelante, utilizados en la fuente de LOD *The Data Hub* (Pastor-Sánchez et al., 2012).

Ilustración 36: Aplicación de los vocabularios LCSH, LEM y RAMEAU



Fuente: Pastor-Sánchez, Juan-Antonio; Martínez-Méndez, Francisco-Javier; Rodríguez-Muñoz, José-Vicente. “Aplicación de SKOS para la interoperabilidad de

vocabularios controlados en el entorno de linked open data ”. El profesional de la información, 2012, mayo-junio, v. 21, n. 3, pp. 245-253 <<http://hdl.handle.net/10760/17066>>

Además, para que los vocabularios sean útiles en la vinculación de datos deben seguir los estándares tecnológicos que promueven las buenas prácticas de LD, en este caso los vocabularios deben asignar URI a cada valor de metadato en lugar, o además, del valor literal, lo que garantiza la identificación unívoca (Isaac et al., 2011). En el triplete RDF constituyen el objeto y asignan el valor al sujeto. Como se representa en la siguiente tabla sobre la estructura de triplete RDF, el elemento de metadato es el predicado, en este caso es un elemento de metadato DC y el objeto es la URI del vocabulario de valor VIAF.

Tabla 8: Ejemplo de URI de valor de metadato de vocabulario de valor

Sujeto	Autor
Predicado RDF	Dcterms:creator
Objeto	http://viaf.org/viaf/55499576

Fuente: Elaboración propia basado en DCMI Terms <<http://dublincore.org>> y Vial <<http://viaf.org/>>

Las instituciones patrimoniales y culturales han percibido la importancia de adaptar las herramientas para la descripción de los recursos en un entorno digital a los lenguajes de la web

semántica, para lo que han desarrollado conjuntos de elementos de metadatos y vocabularios de valores. De la misma manera que el W3C dio cuenta del valor de estas instituciones, y sus contenidos, para la web. Por ello en 2010 se formó el Grupo Incubador de Datos Vinculados de Bibliotecas (LLD-XS) cuya misión era analizar la situación de los modelos y esquemas de metadatos y los estándares y protocolos de interoperabilidad que se deberían usar para la publicación y uso de LOD con datos de bibliotecas (Peset et al., 2011). Tras unos casos de estudio de las iniciativas llevadas a cabo por la comunidad bibliotecaria, en 2011 LLD-XS presentó su Informe Final en el que recoge un inventario de los recursos existentes hasta ese momento: *Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets* (Isaac et al., 2011).

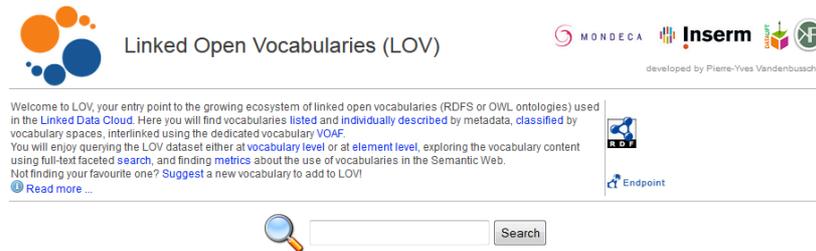
El informe final *Library Linked Data Incubator Group Final Report* “examina cómo utilizar los principios de datos enlazados para que los activos de información que las bibliotecas crean y gestionan sean más visibles y reutilizables más allá de su contexto bibliotecario original, diseminándose en toda la extensión de la web” (Agenjo, 2012). Resume las ventajas de datos enlazados de bibliotecas para diversos actores (investigadores, bibliotecarios y organizaciones), y focaliza como principal ventaja el establecimiento de conexiones entre conjuntos de datos para compartir, extender y reutilizarlos fácilmente en colaboración, no sólo con otras bibliotecas, sino

también con otras comunidades (Baker et al., 2011). Resume la discusión de temas relacionados con los datos bibliotecarios tradicionales, las iniciativas existentes de datos enlazados de bibliotecas y los derechos legales de los datos de las bibliotecas, así como una serie de recomendaciones para los siguientes pasos a dar. El informe compendia también las conclusiones de un estudio sobre las tecnologías actuales de datos vinculados y el inventario de los recursos disponibles en LD de bibliotecas (Baker et al., 2011; Agenjo y Hernández, 2011). Algunos autores como Saorín-Pérez (2012), destacan que el informe del Incubator Group del W3C es crucial para entender la oportunidad que supone para las bibliotecas, y para todas las instituciones culturales y patrimoniales en general, al recoger por primera vez los casos de uso que muestran la implementación exitosa de las tecnologías de web semántica en diferentes instituciones culturales.

Con posterioridad, el proyecto Linked Open Vocabularies (LOV) de la Open Knowledge Foundation (OKFN) es un registro colectivo que recoge un listado con más de 300 vocabularios y ontologías. Su objetivo es facilitar el acceso a lo que ya existe para fomentar su visibilidad, su reutilización, el refinamiento y la ampliación de los mismos. Este conjunto de vocabularios controlados y ontologías están representados en RDF, OWL y SKOS (Méndez y Greenberg, 2012), por lo que resulta de interés al objeto de este trabajo, de forma que

revisamos su listado (disponible en <http://lov.okfn.org/dataset/lov/>) para seleccionar aquellos que podrían ser utilizados en Vestigium. La ilustración siguiente muestra la visualización de la interfaz de consulta del listado LOD de la OKFN.

Ilustración 37: Interfaz de búsqueda del registro de vocabularios Linked Open Vocabularies de la OKFN

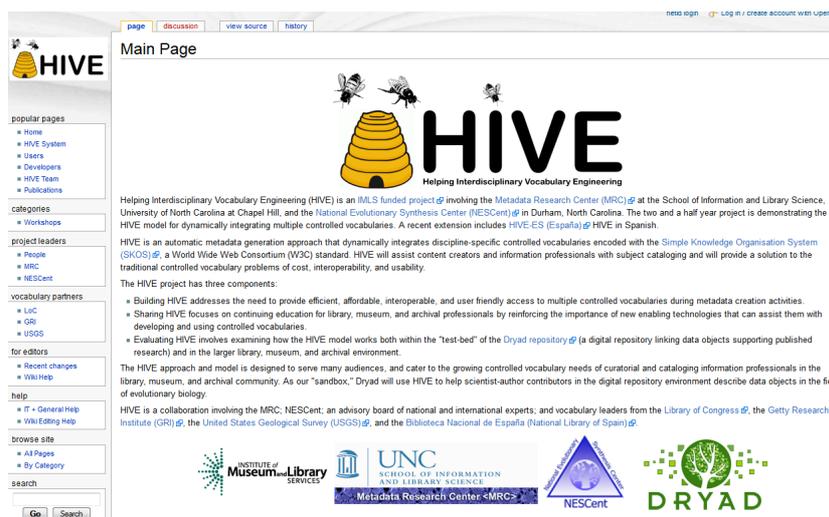


Fuente: OKFN. Linked Open Vocabularies, <http://lov.okfn.org/dataset/lov/>

HIVE (*Helping interdisciplinary vocabularies engineering*), es un proyecto iniciado con el apoyo del Institute of Museum and Library Services (IMLS) de Estados Unidos. HIVE presenta un modelo para la utilización de vocabularios enlazados abiertos para crear metadatos de materia de forma dinámica en el momento de la indización, extrayéndolos a partir de múltiples vocabularios. Este proceso posibilita una selección de los conceptos más adecuados para representar el contenido de un recurso, que no se limita a un único vocabulario (Méndez y Greenberg, 2012).

HIVEES (España), es una extensión del proyecto para los vocabularios en español, llevada a cabo por el grupo de investigación Tecnodoc de la Universidad Carlos III de Madrid, la BNE y el centro de investigación de metadatos de la University of North Carolina at Chapel Hill (SILS MRC). Los vocabularios integrados en HIVE ES están creados como LD, utilizando SKOS como lenguaje de codificación para representar sus clases y propiedades. A fecha de 2012 lo integraban tres vocabularios: la sección española de AGROVOC-FAO, la Lista de Encabezamientos de Materia (LEM) y los Encabezamientos de materia de la BNE (EmBNE) (Méndez y Greenberg, 2012).

Ilustración 38: Interfaz del proyecto HIVE



Fuente: Proyecto HIVE, 2012

<https://www.nescent.org/sites/hive/Main_Page>

La ilustración 38 presenta la interfaz del proyecto HIVE. Es una aplicación tecnológica para almacenar conceptos provenientes de distintos vocabularios y hacerlos accesibles a través de la web. Los vocabularios pueden importarse a HIVE usando el formato SKOS RDF/XML. Requiere instalación y cuenta con una interfaz que permite a los usuarios navegar y buscar en los distintos vocabularios incluidos en el sistema (Sistema HIVE, 2012).

A continuación se presenta la selección de aquellos vocabularios controlados publicados como RDF o SKOS que son susceptibles de ser utilizados en Vestigium porque se adaptan al tipo de contenido que preserva.

1. VOCABULARIOS DE VALORES

1.1 SISTEMAS DE CLASIFICACIÓN (Estivill, 2012):

- *Dewey Decimal Classification (DDC)*

Es el sistema de clasificación del conocimiento desarrollado por Melvin Dewey, de la LC hacia 1876, que ha sido adaptado a LD. La OCLC mantiene un servicio web <<http://dewey.info>> en el que ha asignado una URI a cada concepto DDC. La clasificación está codificada en SKOS y representada en RDF. Las más de

38.000 clases de Dewey están expresadas con el número completo, la información del título y las relaciones semánticas expresadas en SKOS. Cuenta con un SPARQL Endpoint de consulta <<http://dewey.info/sparql.php>>, además de una interfaz de consulta para usuarios para que no tengan que lidiar con el lenguaje RDF. Como se muestra en la ilustración siguiente, por ejemplo, la clase “641” en DDC tiene una URI asignada que redirige a una web en la que se recogen todas las variantes lingüísticas (OCLC, 2014).

Ilustración 39: Ejemplo de la clase 641 con las variantes lingüísticas consultado desde la interfaz para usuarios

Dewey Decimal Classification		
sv	640 Hem & hushåll 641 Mat & dryck	http://dewey.info/scheme/e22/
pt	640 Gerenciamento da casa e família 641 Alimentos e bebidas	http://dewey.info/scheme/e22/
ru	640 Управление домашним хозяйством и семейной жизнью 641 Продукты питания и напитки	http://dewey.info/scheme/e22/
zh	640 家庭 & 家政 641 食品 & 饮品	http://dewey.info/scheme/e22/
ar	640 الإقتصاد المنزلي 641 الأغذية والمشروبات	http://dewey.info/scheme/e22/

Fuente: OCLC. Dewey Decimal Classification/Linked Data <<http://dewey.info>>

- *Universal Decimal Classification (CDU)*

La clasificación decimal universal es un esquema de clasificación multilingüe para todos los campos del

conocimiento. Fue adaptado de la DDC por Paul Otlet y Henry Lafontaine en 1889, y desde 1992 ha sido gestionada por el Consorcio de la CDU (UDCC, Universal Decimal Classification Consortium). En España es norma UNE (UNE 50001:2004) y lo utilizan obligadamente las bibliotecas del sistema público español de bibliotecas para clasificar sus fondos (Slavic et al., 2009). Ha sido adaptada a LD mediante SKOS. El UDCC mantiene una herramienta web de consulta de la DCU en 49 idiomas con unos 70.000 registros (UDCC, 2013).

1.2 ENCABEZAMIENTOS DE MATERIA

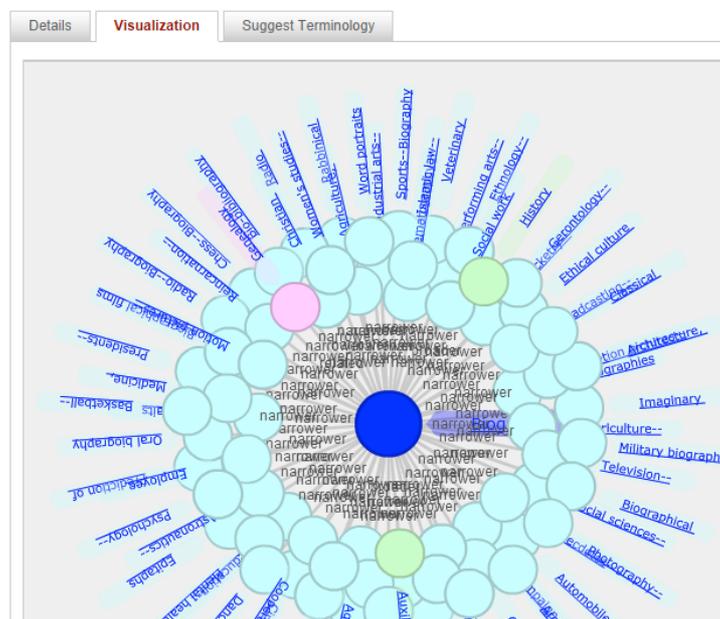
- Library of Congress Subject Headings, LCSH

Es la lista de encabezamientos de materia de la LC. Se ha convertido a lo largo del siglo XX en uno de los vocabularios controlados con mayor autoridad en el mundo. Recoge ordenadas alfabéticamente las materias o temas con las que da acceso a numerosos documentos tanto impresos como electrónicos. Se consulta desde la interfaz de la LCSH. Ha sido adaptado a los LD de modo que los conceptos tienen una URI de identificación unívoca y, las variantes y los términos genéricos y específicos del concepto se expresan mediante tripletas, que pueden visualizarse desde la LC. Por ejemplo, en la ilustración siguiente se observa la visualización del concepto Biografía.

Ilustración 40: Visualización del concepto Biography de la LCSH como LD

Biography

From [Library of Congress Subject Headings](#)



Fuente: Library of Congress

<<http://id.loc.gov/authorities/subjects/sh85014152>>

- *Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié* (RAMEAU)

RAMEAU es el vocabulario de encabezamientos de materia utilizado por la French National Library (BnF), las bibliotecas universitarias y numerosas bibliotecas públicas y de investigación de instituciones privadas. Ha sido desarrollado a partir del LCSH.

- Lista de Encabezamientos de Materia (LEM)

Fue creada para las bibliotecas públicas españolas por la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación Cultura y Deporte. Esta lista de encabezamientos de materia es la que seleccionamos para Vestigium, por tanto, se verá con más detalle en el siguiente capítulo.

1.3 TESAUROS

- *Medical Subject Headings Thesaurus* (MeSH)

Es la lista de encabezamientos de materia publicada y mantenida por la National Library of Medicine (NLM) para la clasificación de información biomédica y de salud. Se emplea en la base de datos Pubmed-Medline desde 1997 para indexar los artículos de revistas biomédicas que custodia. Está disponible on line, cuenta con una interfaz de usuario para la consulta <<http://www.nlm.nih.gov/mesh/MBrowser.html>> y para descargar en diversos formatos legibles por ordenador (NLM, 2013). La información y las publicaciones relacionadas se recogen en el sitio web MeSH de la NLM.

- EuroVoc

EuroVoc es un tesoro multilingüe y multidisciplinario que abarca la terminología de los ámbitos de actividad de la Unión

Europea, con especial hincapié en las labores parlamentarias. EuroVoc está disponible en 23 lenguas oficiales de la Unión Europea (alemán, búlgaro, checo, croata, danés, eslovaco, esloveno, español, estonio, finés, francés, griego, húngaro, inglés, italiano, letón, lituano, maltés, neerlandés, polaco, portugués, rumano y sueco), además de serbio. Está basado en ontologías y tecnologías de web semántica acordes con las recomendaciones del W3C. Está mantenido por la Oficina de Publicaciones de la Unión Europea, y lo utilizan el Parlamento Europeo, parlamentos nacionales y regionales, y administraciones de los países miembros de la EU (Unión Europea, 2014).

- UNESCO *Thesaurus*

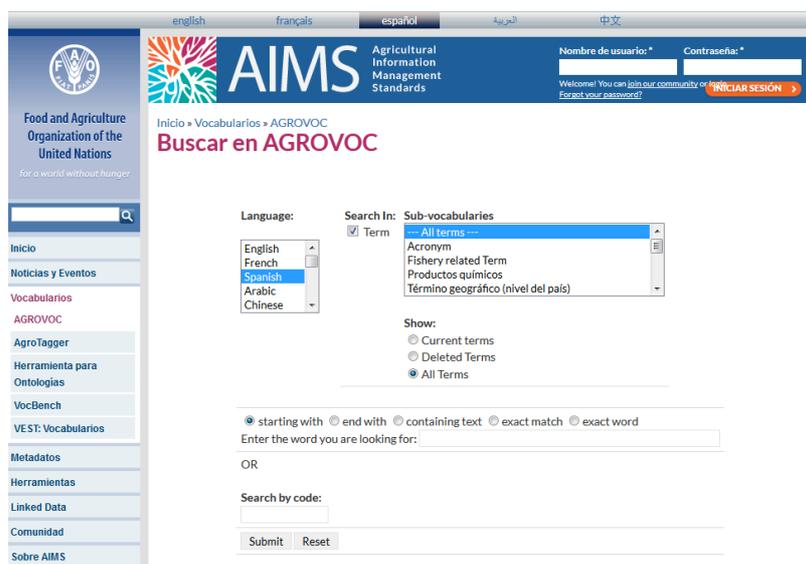
Es una lista controlada y estructurada de términos para el análisis temático y la búsqueda de documentos y publicaciones en los campos de la educación, la cultura, las ciencias naturales, las ciencias sociales y humanas, comunicación e información. Continuamente ampliada y actualizada, su terminología multidisciplinaria refleja la evolución de los programas y actividades de la organización. El Tesoro contiene 7.000 términos en inglés y en ruso, y 8.600 en francés y en español. El tesoro se consulta en línea <<http://databases.unesco.org/thesp/>> y también está disponible en CD-ROM (UNESCO, 2010).

- AGROVOC

Es un tesoro estructurado y multilingüe publicado por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). Cubre todas las áreas de interés de la FAO, especialmente la alimentación, la agricultura, la pesca, la silvicultura, el medio ambiente. Fue desarrollado en 1980 para estandarizar el proceso de indización de la base de datos bibliográfica AGRIS de FAO. Con el auge de las tecnologías semánticas, fue pionero en adoptarlas pues en 2004 fue adaptado como modelo para la organización del conocimiento. En la actualidad se compone de más de 32.000 conceptos en varios idiomas (inglés, checo, francés, alemán, español, italiano, polaco, portugués, ruso, eslovaco, tailandés, turco). En su primera versión semántica se basaba en la ontología OWL por lo que es semántico y está basado en el lenguaje RDF. En 2009, el W3C recomienda SKOS para expresar tesauros en la web, por lo que OWL fue reemplazado por SKOS. AGROVOC contiene un conjunto de datos LOD alineado con otros trece sistemas de organización del conocimiento (vocabularios, tesauros y ontologías), multilingües y todos ellos relacionados con la agricultura (Caracciolo et al., 2013; 2012).

Se puede descargar como LOD *dataset*, consultar desde la interfaz de usuarios o consultar a través del SPARQL Endpoint, que está implementado para el procesamiento automático. La imagen siguiente muestra la interfaz de consulta.

Ilustración 41: Interfaz de consulta del tesoro AGROVOC de agricultura de FAO de UN.



Fuente: FAO. AIMS. AGROVOC.

<<http://aims.fao.org/es/standards/agrovoc/functionalities/search>>

La edición de los datos de AGROVOC se gestiona a través de la herramienta VocBench (VB). Esta herramienta ha sido desarrollada por el equipo AIMS (*Agricultural Information Management Standards*) de FAO, y tiene el propósito de permitir una edición colaborativa basada en web para tesauros multilingües y recursos RDF-SKOS (Caracciolo et al., 2013; 2012).

En LD se utilizan los vocabularios de dos formas, como acabamos de ver, vocabularios que describen el rango de valores que puede tener una propiedad, y como veremos a continuación, los esquemas que proporcionan las propiedades que puede tener un objeto de información (Méndez y Greenberg, 2012).

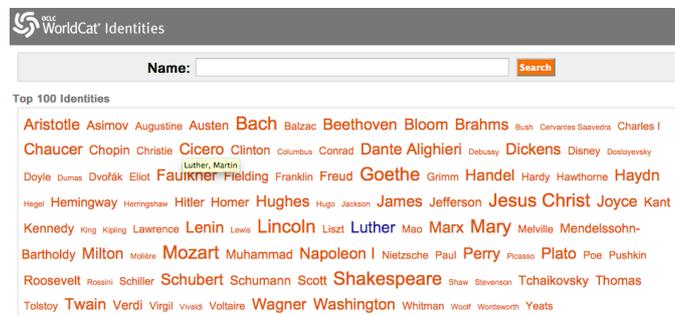
1.4 AUTORIDADES

- *Virtual International Authority File (VIAF)*

Fichero de Autoridades Virtual Internacional es un proyecto conjunto de varias bibliotecas nacionales y organizaciones. Fue iniciado en 2003 por la Library of Congress, la Biblioteca Nacional de Alemania (Deutsche National Bibliothek), la Biblioteca Nacional de Francia (Bibliothèque Nationale de France - BNF) e implementado y alojado posteriormente por la OCLC (OCLC, 2014). El peso específico de las bibliotecas durante décadas para el control de autoridades confiere a este fichero un valor inimaginable. Especialmente ahora que la extensión de contenidos en Internet se ha disparado, las cuestiones de desambiguación de los nombres y la fijación de una o varias formas es esencial no sólo para la comunidad bibliotecaria. Por tanto, el hecho de publicar VIAF como dataset en Data Hub asume una importancia capital pues su utilización puede ser tan transversal como la DBpedia.

Es un proyecto que quiere disminuir el coste e incrementar la utilidad de los ficheros de autoridad de las bibliotecas, mediante la comparación y la correspondencia entre los ficheros de autoridades de las bibliotecas nacionales participantes. El objetivo es enlazar los registros de autoridad de muchas bibliotecas de forma que queden vinculados los diferentes nombres existentes en diferentes idiomas para una misma persona u organización. El resultado es un servicio global disponible en Internet libremente que recoge todas las variantes de cada autoridad. Una experiencia previa de la OCLC fue WorldCat Identities, un software de comparación de nombres, en el que se recogía desde WorldCat, el catálogo colectivo más grande del mundo, todas las variantes de un mismo autor y se explotaban los datos con cronologías, nubes de tags (OCLC, 2014). La ilustración siguiente presenta la interfaz de WorldCat Identities y la nube de tags.

Ilustración 42: Interfaz de consulta de autoridades WorldCat Identities



Fuente: OCLC. WorldCat Identities <<https://www.worldcat.org/identities/>>

Su aplicación se verá en el siguiente capítulo, pues está seleccionado para utilizar en Vestigium.

- *Library of Congress Name Authority File (LC/NAF)*

Proporciona datos para los nombres de personas, organizaciones, eventos, lugares y títulos. El formato es MADS/RDF.

- Geonames

Es una base de datos de nombres geográficos, que podría emplearse en Vestigium, por lo que se verá en profundidad en el siguiente capítulo.

2. CONJUNTOS DE METADATOS

2.1 METADATOS DE AUTORIDADES

- *Metadata Authority Description Schema in RDF (MADS/RDF)*

Como vimos en el capítulo anterior, la Library of Congress adaptó el formato bibliográfico a los lenguajes de marcas. La LC se ha sumado a la iniciativa de la WS y constituyó un servicio LC Linked Data Service Authorities and Vocabularies que han desarrollado formatos para expresar los conjuntos de elementos

de metadatos en lenguajes ontológicos. Por ejemplo, MADS/RDF es el esquema de metadatos MADS para la descripción de autoridades para asignar valores de nombres personales, de organizaciones y geográficos. Está desarrollada como una ontología SKOS de control de autoridades basada en RDF y provee de significado la descripción semántica de las autoridades. El namespace que utiliza es `<xmlns:mads=http://www.loc.gov/mads/rdf/v1#>` (Isaac et al., 2011).

- *Friend of a Friend Project (FOAF)*

Es un lenguaje para describir información sobre las personas, grupos, documentos, proyectos y las relaciones entre ellos. FOAF Project (*Friend of a Friend*) es un proyecto iniciado en el año 2000 dedicado a enlazar personas e información en la web. El lenguaje describe información sobre las personas, las relaciones, las redes sociales y las de colaboración entre los individuos. Se basa en RDF y está escrito en XML, lo que le proporciona los beneficios de poder ser combinado con otros lenguajes RDF. Describe los objetos (personas, documentos) mediante propiedades (*name, familyName, age, birthday, status, mbox, phone, publications, fundedBy, title, topic*), y los tipos de objetos a través de clases (*agent, document, group, image, organization, person, project*). Identifican a las personas de forma unívoca en la web, asignando propiedades como URL y

dirección de correo. El namespace es `<xmlns:foaf=http://xmlns/foaf/0.1/>` (FOAF Project, 2000). Este lenguaje podría ser interesante para describir vínculos entre investigadores en Vestigum.

2.2 METADATOS GENERALISTAS

- *Dublin Core Metadata Elements Terms*

Como se ha visto en el capítulo anterior, DC es un conjunto de elementos de metadatos para la descripción de documentos electrónicos. Revisamos la versión inicial (DCM Elements), y la versión actual (DCM Terms) conjunto de elementos de metadatos basados en RDF para describir los datos como LD se verá en profundidad en el siguiente capítulo ya que se utiliza en Vestigium.

- *Vocabulary of Interlinked Datasets (VOID)*

Es un vocabulario RDF Schema para expresar metadatos con los que describir conjuntos de datos enlazados (*linked datasets*) y las relaciones entre estos paquetes de datos. El elemento fundamental son los linked datasets, entendido como un conjunto de tripletas RDF que son publicadas o agregadas por un proveedor de datos (una biblioteca por ejemplo). Este paquete de datos se concibe como una colección de datos significativos y relacionados entre sí. Estos datasets son

publicados, mantenidos y agregados por su proveedor y se encuentran disponibles en la web como RDF. Es posible acceder a estos datos a través de URI HTTP desreferenciables y a través de un punto de consulta SPARQL Endpoint (Alexander et al., 2010; 2011). VOID está concebido como un puente entre los editores de datos RDF -quienes realizan tareas de gestión, catalogación y almacenamiento de datos- y los consumidores de datos RDF, quienes realizan tareas de consulta de datos, descubrimiento e indexación de datasets.

Estos conjuntos o paquetes de datos se enlazan o vinculan con otros paquetes de datos a través de *linksets*. Un linkset es una colección de enlaces RDF entre dos datasets. Por ejemplo, es posible que en un dataset se encuentren todos los sujetos de la tripleta y en otro dataset todos los objetos. VOID describe con metadatos tanto los linked datasets como los linksets. Cubre cuatro tipos de metadatos: metadatos generales como el modelo DC, metadatos de acceso para describir cómo acceder a datos RDF, metadatos estructurales para describir la estructura de los datasets, y la descripción de los enlaces entre conjuntos de datasets para entender cómo se relacionan múltiples conjuntos de datos (Alexander et al., 2010; 2011).

La descripción completa del vocabulario y sus elementos se encuentra en Alexander, Keith; Cyganiak, Richard; Hausenblas, Michael; Zhao, Jun. *Describing Linked Datasets with the VOID*

Vocabulary, 2011 <<http://www.w3.org/TR/void/>>, y en Cyganiak, Richard; Zhao, Jun; Alexander, Keith; Hausenblas, Michael; *Vocabulary of Interlinked Datasets (VoID) DERI Vocabularies*, 2011 <<http://vocab.deri.ie/void/>>.

2.3 METADATOS PARA LA DESCRIPCIÓN DE DATOS ESPECÍFICOS

- Bibliographic Ontology Specification (BIBO)

BIBO proporciona los conceptos principales y las propiedades para describir citas y referencias bibliográficas. Describe los elementos bibliográficos en RDF. Utilizar RDF permite a BIBO combinarse con otros vocabularios de descripción de metadatos como DC y FOAF. Entre los objetos que describe se encuentran artículos, libros, capítulos de libros, comunicaciones, presentaciones de conferencias, casos legales, cartas, manuscritos, tesis. El namespace de la ontología es <<http://purl.org/ontology/bibo/>> (D'Arcus y Giasson, 2009). Un ejemplo de la descripción bibliográfica de un libro con la ontología BIBO combinada con etiquetas de metadatos de descripción DC es el que aparece en la siguiente ilustración.

Ilustración 43: Ejemplo de la ontología BIBO aplicada en la descripción de un libro

```
<urn:isbn:23983498> a bibo:Book ;  
  dc:creator <http://examples.net/contributors/2>  
  dc:title "Book Title"@en ;  
  dc:date "2000" ;  
  dc:publisher <http://ex.net/agents/1> .
```

Fuente: D'Arcus, Bruce; Giasson, Frédérick. Bibliographic Ontology Specification: examples, 2009 <<http://bibliontology.com/examples>>

La siguiente imagen recoge la descripción de un artículo de revista que combina etiquetas de descripción de BIBO, como por ejemplo, los campos volumen y número de la revista, y página de inicio y página final del artículo que se está describiendo, junto con etiquetas de descripción DC para los campos título del artículo y creador.

Ilustración 44: Ejemplo de la ontología BIBO aplicada en la descripción de un artículo de revista

```
<info:doi/10.1134/S0003683806040089> a bibo:Article ;  
  dc:title "Effect of argillaceous minerals on the growth of phosphate-mobilizing bacteria"  
  dc:date "2006-01-01" ;  
  dc:isPartOf <urn:issn:23346587> ;  
  bibo:volume "42" ;  
  bibo:issue "4" ;  
  bibo:pageStart "388" ;  
  bibo:pageEnd "391" ;  
  dc:creator <http://examples.net/contributors/2> ;  
  dc:creator <http://examples.net/contributors/1> ;  
  bibo:authorList ( <http://examples.net/contributors/2> <http://examples.net/contributors/1> ) .
```

Fuente: BIBO examples, 2012 <<http://bibliontology.com/examples>>

- *Data Cube Vocabulary* (qb)

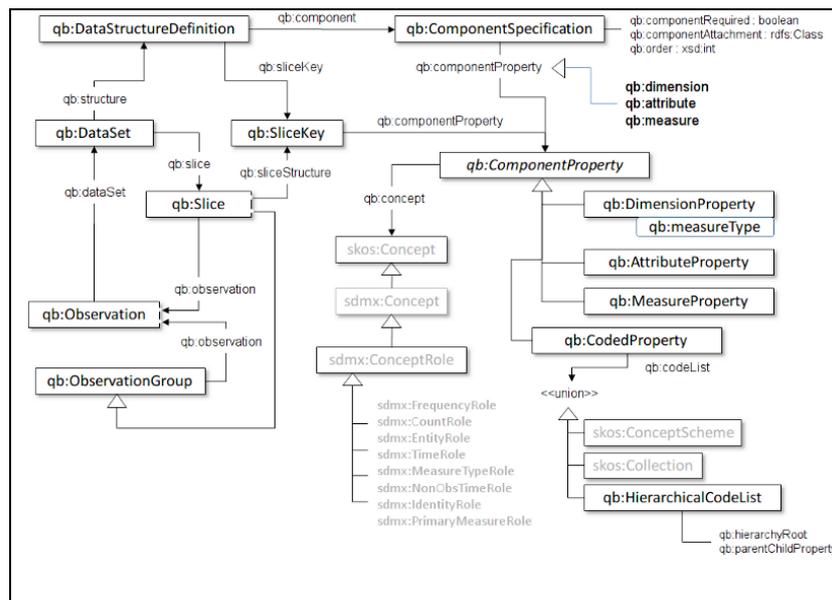
El vocabulario *Data Cube Vocabulary* (qb) se basa en el lenguaje RDF para publicar en la web datos multidimensionales como por ejemplo, estadísticas, y enlazar con conjuntos de datos y conceptos relacionados. Los datos estadísticos se emplean en muchas operaciones en diversos campos. Se emplean por ejemplo, para hacer las políticas de predicción, para la planificación de procesos, en el desarrollo de aplicaciones y para visualizaciones de datos en la web. Esto hace necesario que se publiquen en un formato que permita enlazarlos y combinarlos con información relacionada (Cyganiak y Reynolds, 2014). En el caso del presente trabajo, Vestigium cuenta con tablas estadísticas procedentes de datos bibliométricos que podrían ser descritos con *Data Cube*.

Data Cube define la estructura de uno o más conjuntos de datos a través de las propiedades dimensiones, atributos y medidas, utilizadas en el dataset (Cyganiak y Reynolds, 2014). *Data Cube* es el núcleo del vocabulario, y utiliza otros vocabularios como SKOS, DC Terms y VoiD para describir el acceso a datos, y FOAF para describir personas. También soporta extensiones de vocabularios modulares que se están desarrollando para apoyar la publicación de otros aspectos relacionados con los datos estadísticos como los flujos de datos, los metadatos para

entrevistas y los *static reference metadata* (Cyganiak y Reynolds, 2014).

El siguiente esquema presenta de forma gráfica los conceptos clave para la descripción de un conjunto de datos estadísticos y sus relaciones.

Ilustración 45: Grafico de los conceptos clave de descripción de un dataset estadístico descrito con el vocabulario qb



Fuente: Cyganiak, Richard; Reynolds, Dave. *The RDF Data Cube Vocabulary*, 2014 <<http://www.w3.org/TR/vocab-data-cube/>>

Como se puede observar en esta ilustración, los conceptos clave son el dataset y la definición de la estructura de los datos, los componentes que especifican la estructura del dataset y las propiedades, que son dimensión, atributo y medida. Cada uno de

estos conceptos clave se definen con una serie de etiquetas que combinan el vocabulario *Data Cube* con otros como SKOS, FOAF, VOID, DC Terms, OWL, RDFS (Cyganiak y Reynolds, 2014). Por ejemplo, en la siguiente imagen se describe la estructura de definición de datos con los componentes dimensión, atributos y medida, y *slice* por región.

Ilustración 46: Extracto de definición de la estructura de los datos del dataset con el vocabulario qb

```
eg:dsd-le3 a qb:DataStructureDefinition;
qb:component
  # The dimensions
  [ qb:dimension eg:refArea;          qb:order 1 ],
  [ qb:dimension eg:refPeriod;       qb:order 2; qb:componentAttachment qb:Slice ],
  [ qb:dimension sdmx-dimension:sex; qb:order 3; qb:componentAttachment qb:Slice ];

# The measure(s)
qb:component [ qb:measure eg:lifeExpectancy];

# The attributes
qb:component [ qb:attribute sdmx-attribute:unitMeasure;
               qb:componentRequired "true"^^xsd:boolean;
               qb:componentAttachment qb:DataSet; ];

# slices
qb:sliceKey eg:sliceByRegion ;
.

eg:sliceByRegion a qb:SliceKey;
rdfs:label "slice by region"@en;
rdfs:comment "Slice by grouping regions together, fixing sex and time values"@en;
qb:componentProperty eg:refPeriod, sdmx-dimension:sex ;
```

Fuente: Cyganiak, Richard; Reynolds, Dave. *The RDF Data Cube Vocabulary*, 2014 <<http://www.w3.org/TR/vocab-data-cube/>>

En esta ilustración vemos que utiliza el concepto *slice* para hacer referencia a una parte o subconjunto de observaciones dentro de un dataset. Por ejemplo, se utiliza para todas las observaciones sobre un indicador determinado dentro del conjunto de datos, ya que cuando se trabaja con datos

estadísticos suele ser común seleccionar y combinar datos de indicadores determinados, así como excluir otros indicadores.

El namespace es `<http://purl.org/linked-data/cube#>`. Más información y los componentes del vocabulario se recogen en la recomendación *Data Cube Vocabulary* del W3C (Cyganiak y Reynolds, 2014).

- *Citation Typing Ontology* (CiTO)

Es una ontología para describir la naturaleza de las referencias bibliográficas citadas en artículos de investigación y trabajos académicos, y para publicar esta información en la web semántica. La ontología describe la citación en términos de relación entre publicaciones que citan y publicaciones citadas. Tiene en cuenta información sobre las frecuencias globales de citación de cada trabajo citado y la naturaleza de los trabajos citados, entre otros elementos (Shotton, 2010). Los dos conceptos clave son:

- Publicación que cita (*citingwork*): cito:cites
- Publicación citada (*citedwork*): cito:isCitedBy

CiTO se basa en RDF para publicar información sobre la cita y la obra citada en formato legible por máquina, se escribe en el lenguaje ontológico OWL y su namespace es

<<http://purl.org/net/cito/>> (Shotton, 2010). En la siguiente tabla se define un ejemplo de relación entre un trabajo citante y un trabajo citado con estructura de lenguaje RDF.

Tabla 9: Ejemplo de relación entre un trabajo citante y uno citado descrito con la ontología CiTO estructurado en RDF

Sujeto	Predicado RDF	Objeto
< http://example1.com/citingwork >	cito:cites	< http://example2.com/citedwork >
< http://example2.com/citedwork >	cito:isCitedBy	< http://example1.com/citingwork >

Fuente: Shotton, David; Peroni, Silvio. *CiTO, the Citation Typing Ontology*, 2014 <<http://purl.org/spar/cito>>

Esta ontología es la que más se adapta para describir estudios bibliométricos, pues CiTO provee la forma de describir y codificar en formato legible por máquina (RDF) los datos de citación de los trabajos científicos y la red de citas de un conjunto de publicaciones. Por esta razón puede ser un buen vocabulario para incorporar en la biblioteca Vestigium que preserva este tipo de contenidos.

CiTO permite también la caracterización de las citas bibliográficas, es decir, la identificación de citas positivas, negativas o neutras a través de la asignación de etiquetas sobre las citas y sobre los autores (Shotton, 2010), tal y como se observa en la tabla siguiente.

Tabla 10: Ejemplo de algunas entidades y atributos del lenguaje CiTO para la caracterización de las citas bibliográficas

	Positive	Negative	Neutral
<i>cito:cites</i>	<i>cito:confirms</i>	<i>cito:corrects</i>	<i>cito:discusses</i>
<i>cito:citesAsAuthority</i>	<i>cito:credits</i>	<i>cito:critiques</i>	<i>cito:reviews</i>

Fuente: Shotton, David; Peroni, Silvio. *CiTO, the Citation Typing Ontology*, 2014 <<http://purl.org/spar/cito>>

Así como también pueden ser útiles las frecuencias globales de citación, que se emplean como métrica para evaluar la calidad de los trabajos publicados mediante etiquetas de definición de entidades (*cito:globalCitationFrequency*, *cito:globalCountSource*, *cito:inTextCitationFrequency*) (Shotton, 2010).

La definición del lenguaje CiTO y todas sus propiedades se recogen en *CiTO, the Citation Typing Ontology* (Shotton y Peroni 2014).

- Schema.org

Schema.org es una colección de esquemas de metadatos estructurados que se utilizan para describir los recursos existentes en sitios web. Se define con mayor detalle en el capítulo siguiente porque se emplea en la descripción de contenidos de Vestigium.

Aplicar estos conjuntos de metadatos y vocabularios de valores en la descripción de los recursos digitales, y seguir los principios LD que emplean los protocolos para la transferencia (protocolo HTTP) para la descripción (RDF) y la identificación de los recursos de forma unívoca (URI), así como los lenguajes de definición de ontologías (RDF *Schema*, OWL y SKOS), permite a las instituciones y organismos publicar sus datasets como LD y formar parte de la web semántica. La ilustración siguiente es un gráfico recoge a modo de resumen estas tecnologías.

Ilustración 47: Mapa de los elementos de la web semántica



Fuente: elaboración propia

La aplicación de todas estas herramientas asegura que se están representando los objetos digitales y los activos de información

que las bibliotecas generan (datos bibliográficos y de autoridades) de una forma normalizada en el entorno web. Con ello mejoran sustancialmente la visibilidad de los recursos, fomentan su reutilización y facilitan la recuperación de información. Insistimos en que esto supone un cambio cualitativo que no podría haberse producido si no se hubieran incorporado estas herramientas en la comunidad bibliotecaria. Este agregado de prácticas de creación y distribución de datos permiten compartir y enlazar los conjuntos de datos entre bibliotecas y otras comunidades de una forma automática.

Fuentes linked data y su consulta

Publicación en LOD

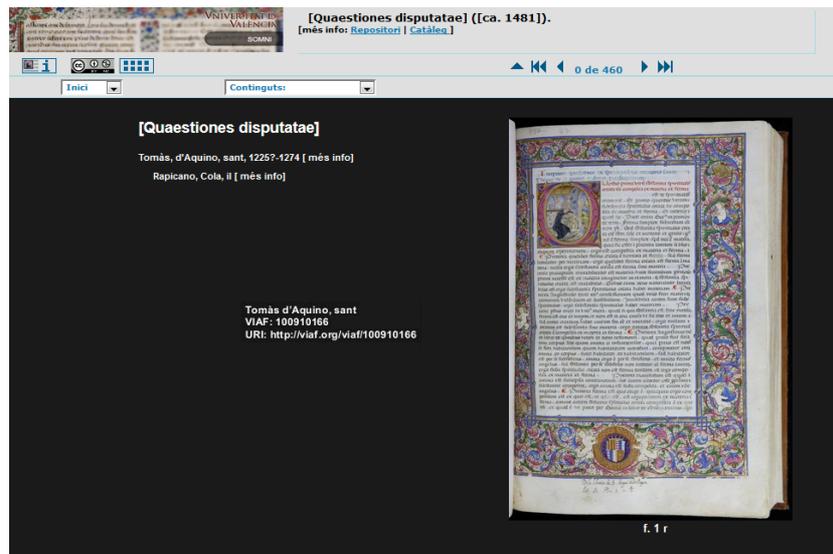
Como hemos estado viendo las pautas para la publicación de LD favorecen el modelo tradicional empleado por las bibliotecas para colaborar e intercambiar información. Los estándares que hemos recogido son beneficiosos para compartir y enlazar fácilmente los datos de bibliotecas. Así, describen los recursos de forma estandarizada en colaboración con otras instituciones y se enlazan con datos publicados por otras entidades.

En el ecosistema de datos enlazados, los recursos se describen a partir de afirmaciones individuales que establecen distintas organizaciones, y la suma de muchas afirmaciones sobre un recurso concreto e identificado de forma unívoca, forma parte de un grafo global de datos. En definitiva, aplicar las buenas prácticas de LD pretende mejorar la información digital a través de una extensa red de datos conectados entre sí por la que navegar de forma continuada y ampliar así la información. Esta red de conexiones de datos puede incrementar el valor de la información que contiene una biblioteca digital o una fuente de información con más información de contexto que se refiere a la que se está describiendo (Baker et al., 2011).

Por tanto, las bibliotecas pueden construir su oferta de información con los fondos propios y con la información generada por otros. La integración de información ajena en las web está propiciada por la interoperabilidad de la tecnología que se comentaba en el capítulo primero. Hasta ahora hemos incidido en las tecnologías que permiten producir los datos para la web semántica, y ahora vamos a ver las fuentes que existen de naturaleza semántica. Para consumir la información que se encuentra en las fuentes semánticas los datos han de estar disponibles libremente.

La siguiente ilustración muestra un ejemplo de cómo se pueden consumir datos de fuentes como VIAF o DBpedia con bastante sencillez, de manera que alimente el repositorio patrimonial de la institución. En este ejemplo la aplicación obtiene datos para enriquecer el repositorio de la Universidad de Valencia (Barrueco y García Testal, 2013).

Il·lustració 48: Interfaz de la aplicació LOD para la consulta de la colecció SOMNI del repositori de la Universitat de València.



Fuente: Barrueco, José Manuel; García Testal, Cristina. “Aplicación de LOD para enriquecer la colección patrimonial de la Universitat de València Somni”. En: *XIII Jornadas Españolas de Documentación: Fesabid 2013*. Toledo 24 y 25 de mayo, 2013 <<http://www.slideshare.net/fesabid/aplicacin-de-lod-para-enriquecer-la-coleccin-patrimonial-de-la-universitat-de-valncia-somni-jos-manuel-barrueco-cruz-cristina-garca-testal>>

La iniciativa Datos Abiertos Enlazados (*Linked Open Data*, en adelante LOD) en las instituciones culturales y bibliotecas digitales viene ampliándose desde 2011. En general como hemos comentado suelen ser fuentes que publican grandes cantidades de datos semánticos. Por ejemplo, ya han migrado sus contenidos a LD organismos como la British Library con la publicación de la *British National Bibliography* (Bibliografía

Nacional Británica), la Biblioteca Nacional de Hungría, la Deutsche National Bibliothek (Biblioteca Nacional de Alemania), la Bibliotheque Nationale de France (Biblioteca Nacional de Francia), y en el ámbito nacional la Biblioteca Nacional Española (BNE) ha publicado conforme a los principios LD la información de los catálogos bibliográficos y de autoridades. Otros proyectos imprescindibles para la comunidad bibliotecaria son Europeana y DBpedia. DBpedia <<http://dbpedia.org>> es un proyecto que genera información semántica según los principios LD a partir de la información disponible en Wikipedia, <<http://es.wikipedia.org/>>, la enciclopedia digital colaborativa y libre. Otro proyecto es la base de datos AGRIS de la Organización de las Naciones Unidas para la Alimentación y la Agricultura FAO. Entre estos proyectos existen algunos fundamentales para el objeto de esta tesis, razón por la que serán descritos más adelante como fuentes semánticas susceptibles de ser consumidas.

Publicar la información bibliográfica siguiendo los principios LD, de forma que estén disponibles para ser enlazados o consumidos en la web no implica que los datos sean de acceso abierto (Hernández-Pérez y García Moreno, 2013). Por otro lado, los LOD amplían las características técnicas de los LD e incorporan el concepto *open* vinculado a las condiciones de accesibilidad y reutilización (Peset et al., 2011). Tal y como se ha visto en el primer capítulo para las publicaciones científicas,

se aplica aquí a los datos enlazados. Por LOD se entiende datos enlazados publicados bajo una licencia abierta. Esto significa una disposición libre y gratuita de los datos, sin restricciones de copyright para los usuarios y las instituciones, de forma que se puedan explotar y combinar para nuevos desarrollos y aplicaciones. Esta nueva dimensión conecta los datos que publican con el movimiento *Open Data*, otra de cuyas grandes bases es el gobierno electrónico y transparente.

Los conjuntos de datos publicados como LOD son accesibles y están a disposición del público para que otras fuentes y recursos puedan enlazarlos. Los principios LOD sugeridos por Tim Berners-Lee (2006) y extendidos por Bizer et al., (2009) establecen un sistema de puntuación de cinco estrellas para medir las características *open* que cumplen los LD:

- * Disponible en la web, en cualquier formato, pero con licencia Open, para ser Open Data.
- ** Disponible como datos estructurados legibles por máquina (por ejemplo, formato Excel en lugar de una tabla o una imagen escaneada).
- *** Datos disponibles en un formato no propietario (por ejemplo, CSV en lugar de Excel).
- **** Datos publicados utilizando estándares de datos abiertos del W3C.

***** Enlazar los datos con datos de otras fuentes para proporcionar un contexto.

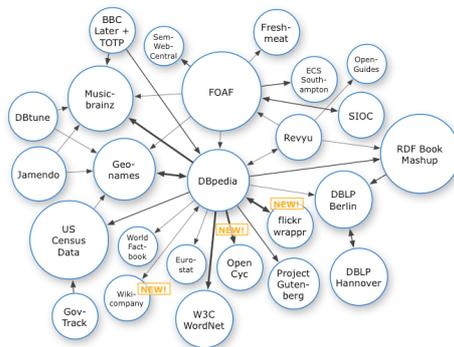
Fuentes disponibles

La utilización de LOD está creciendo, de tal forma que cada vez son más las disciplinas comprometidas con buenas prácticas de LD y con las tecnologías para dar a conocer y enlazar conjuntos de datos, posibilitando un acceso ilimitado a los mismos, compartiéndolos, integrándolos y reutilizándolos (Méndez y Greeberg, 2012). En los últimos tiempos esta tendencia ha tenido un desarrollo asombroso, habiendo ya una gran cantidad de datos semánticos publicados y disponibles para enlazar con otras instituciones y bibliotecas, a nivel internacional. Muchos de estos paquetes de datos se registran en el proyecto The Data Hub o Central de Datos de la Open Knowledge Foundation, y se describen de forma colaborativa. En los últimos años los conjuntos de datasets publicados como LOD han ido aumentando de forma continuada. Un paquete de datos o dataset es un conjunto de tripletas RDF consultable o descargable, sin distinción específica entre datos y metadatos, y cada dataset puede contener una cantidad de tripletas variable (Baker et al., 2011).

Los conjuntos de datos LOD recogidos en Data Hub, se representan gráficamente en el Diagrama de la nube de los Datos Abiertos Enlazados (LOD Cloud Diagram), que recoge su crecimiento desde 2007 hasta 2014. El éxito de esta representación radica en que ofrece información sobre la cantidad de datos de cada uno de los grandes conjuntos y las relaciones entre sí. Por otra parte, en la versión coloreada del diagrama se expresa la pertenencia a las categorías de *Media*, *Geographic*, *Publications*, *User-generated content*, *Government*, *Cross domain*, *Life Sciences*. Al clicar en un nodo enlaza con la página de metadatos del dataset.

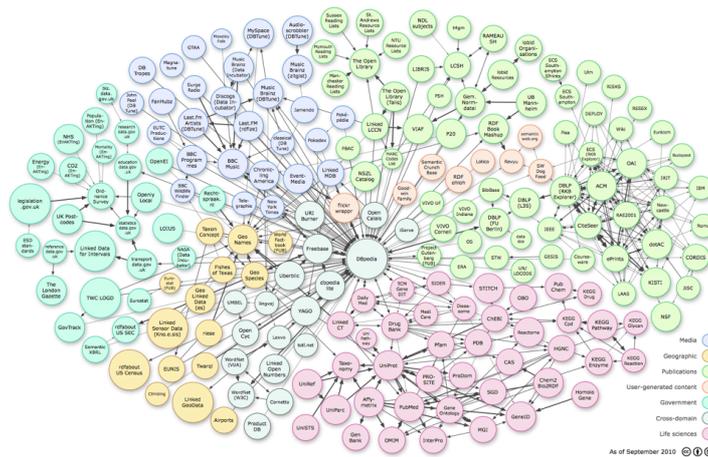
A continuación se muestra en las ilustraciones 49, 50, 51, 52, 53 y 54 la evolución anual de la web semántica a través del diagrama de los datos LOD disponibles desde 2007 hasta 2014.

Ilustración 49: Primer diagrama de LOD Cloud a fecha de 2007



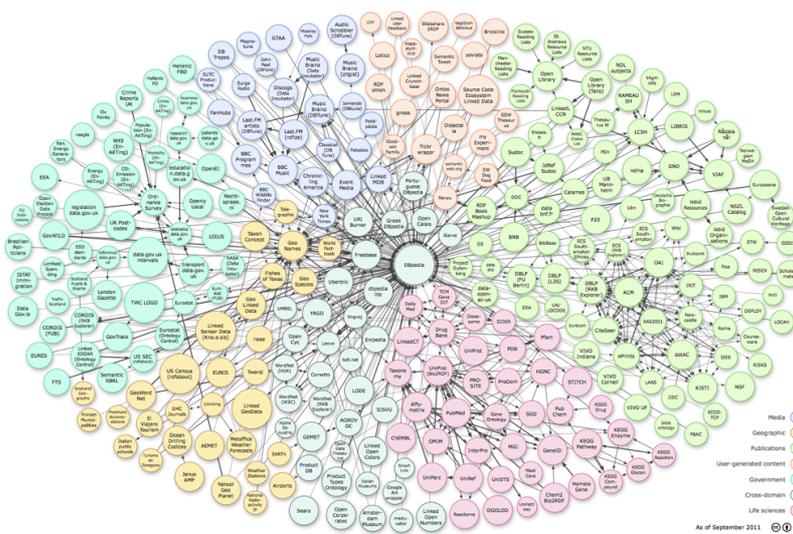
Fuente: Cyganiak, Richard; Jentzsch, Anja. *Linking Open Data cloud diagram*, 2011 <<http://lod-cloud.net/>>

Ilustración 52: Diagrama de la nube de LOD en 2010



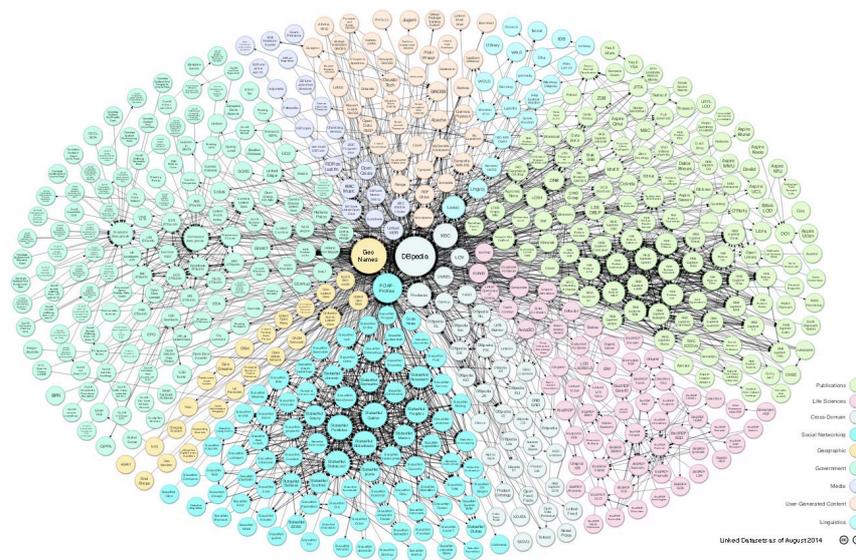
Fuente: Cyganiak, Richard; Jentzsch, Anja. *Linking Open Data cloud diagram*, 2011 <<http://lod-cloud.net/>>

Ilustración 53: Diagrama de la nube de LOD en 2011



Fuente: Cyganiak, Richard; Jentzsch, Anja. *Linking Open Data cloud diagram*, 2011 <<http://lod-cloud.net/>>

Ilustración 54: Diagrama de la nube de LOD en 2014

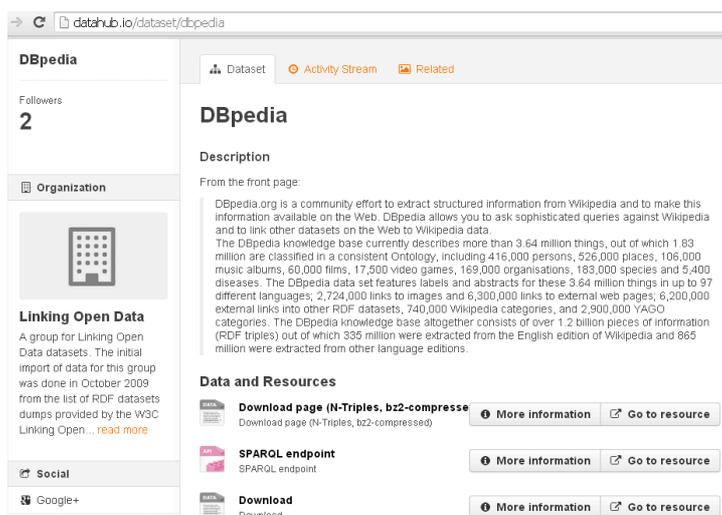


Fuente: Cyganiak, Richard; Jentzsch, Anja. *Linking Open Data cloud diagram*, 2014 <<http://lod-cloud.net/>>

La ilustración 54 representa los conjuntos de datos LOD que se preservan en Data Hub procedentes de otras fuentes. En el centro del diagrama se distingue como nodo central DBpedia, junto con GeoNames y FOAF profiles, que ocupan los lugares centrales y se representan con los nodos más grandes por la gran cantidad de conjuntos de datos LOD, y por el número de enlaces que tienen hacia otras fuentes.

En el repositorio de datos Data Hub, por ejemplo, desde la página sobre DBpedia se puede acceder a la información y al punto de consulta DBpedia SPARQL Endpoint. La siguiente imagen muestra esta página.

Ilustración 55: Página web de información y acceso a los LOD de DBpedia en la interfaz de consulta de Data Hub



Fuente: The Data Hub <<http://datahub.io/es/dataset/dbpedia>>

Como hemos dicho, estos conjuntos de datasets están clasificados por dominios e identificados por colores en el diagrama. Las categorías más relacionadas con el objeto de esta tesis son las de *Publications*, pero también observamos conjuntos de datos bibliográficos que pueden ser de interés en *Life Sciences*.

Por otra parte, la siguiente tabla recoge las áreas temáticas de la nube de LOD, el número de datasets a fecha de 2011, el número de tripletas RDF y el número de enlaces a fuentes externas. En total cuenta con cerca de 295 datasets procedentes de varios dominios, cerca de 30 billones de ítems de datos y unos 500

millones de enlaces entre ellos (Wood et al., 2014). Actualmente estas cantidades deben ser mucho más superiores.

Tabla 11: Clasificación temática de los conjuntos de datos representados en el diagrama de la nube de LOD

Dominio	N° de datasets	N° de tripletas	N° de Enlaces externos
Media	25	1.841.852.061	50.440.705
Geográfico	31	6.145.532.484	35.812.328
Ciencias de la Vida	49	3.036.336.004	191.844.090
Publicaciones (incluyendo datos de bibliotecas y museos)	87	2.950.720.693	139.925.218
Gobierno	49	13.315.009.400	19.343.519
Contenido generado por los usuarios, vocabularios, ontologías, tesauros	20	4.184.635.715	63.183.065

Fuente: Cyganiak, Richard; Jentzsch, Anja. *LODCloud diagram state*, 2011 <<http://lod-cloud.net/state/>>

Este proyecto es el referente en el mundo de la web semántica para conectar con conjuntos de datos LOD y proporcionar un contexto informativo. La tabla anterior permite comparar la diferencia entre producción de datos semánticos (número de tripletas) y consumo o enlaces. La mayor producción es gubernamental, seguida por los datos geográficos; mientras que el mayor número de enlaces los reúne el sector de ciencias de la vida, seguido del de publicaciones. Insistimos en que de cara al objetivo de este trabajo, las categorías resultan un tanto confusas, pero puede verse que existen numerosas fuentes de interés para nuestra biblioteca digital.

Por esta razón, a continuación se recogen algunas de las fuentes más representativas o de interés para la comunidad bibliotecaria que se proponen para enlazar con los contenidos de Vestigium. La clasificación que se utiliza sigue un orden lógico, basado en la naturaleza de los datos desde los más generales a los más específicos. Obviamente, no es el proyecto en sí el que pretendemos describir, sino los contenidos que está recogiendo. No obstante, se dibujarán unos mínimos rasgos de su historia y promotores del proyecto que permitan dimensionar el valor de esos datos.

Fuentes generales:

- DBpedia: es un proyecto que desde 2011 utiliza los datos de Wikipedia y los convierte en LD. Está conectada con la mayoría de las fuentes que refleja LOD Cloud Diagram. A pesar de todas las críticas que ha recibido y recibe Wikipedia, es indudable que los contenidos son utilizados como primera fuente de conocimiento para muchos aspectos. Por otra parte, el hecho de ser libre, multimedia y de responsabilidad compartida produce un uso masivo de sus datos, pero además una flexibilización con respecto a la confianza de la fuente, algo que desde la proliferación de contenidos a partir de la web social se está extendiendo en la cultura social. La siguiente ilustración presenta la interfaz del

proyecto.

Ilustración 56: Interfaz de DBpedia



Fuente: DBpedia <<http://wiki.dbpedia.org/OnlineAccess>>

Este proyecto se realiza en colaboración entre varias universidades: Universidad de Leipzig, Universidad Libre de Berlín y Open Link Software. La información se almacena en lenguaje RDF y la consulta de datasets se realiza mediante SPARQL Endpoint (DBpedia, 2014). Además, para facilitar su consumo, los datos pueden obtenerse íntegramente desde Data Hub, como hemos visto anteriormente.

- Europeaana Linked Open Data: es una biblioteca digital semántica, desarrollada en 2012, que constituye una agregación común de los fondos de numerosas instituciones culturales de todos los países miembro de la Unión Europea, que permite a través de una única interfaz tener acceso a millones de objetos digitales

bibliográficos, archivísticos y de arte, que conforman el acervo cultural europeo. Surge como reflejo del patrimonio europeo después del lanzamiento del *American Memory* de Estados Unidos de América. Actualmente, las instituciones que forman parte de Europeana son proveedoras de contenidos que proporcionan metadatos descriptivos de los objetos digitales y todas sus relaciones, con el autor, con las versiones de la obra, los eventos asociados, las copias, temáticas. Los proveedores forman un grupo controlado por cada país, de manera que no todos los fondos que sean interoperables pueden estar en Europeana, dado que su foco de trabajo son los objetos culturales. La diversidad de los estándares de descripción de cada una de las comunidades que proporciona los datos - bibliotecas, bibliotecas digitales, archivos, museos y el sector audiovisual- pueden ser mapeados por el EDM (*European Data Model*) debido a la flexibilidad y adaptación a las clases y propiedades de cada comunidad. La consulta de los datos se realiza mediante un SPARQL Endpoint (Europeana, 2014). La ilustración siguiente presenta la interfaz de consulta.

Ilustración 57: Interfaz de búsqueda de SPARQL Endpoint de Europeana Linked Open Data



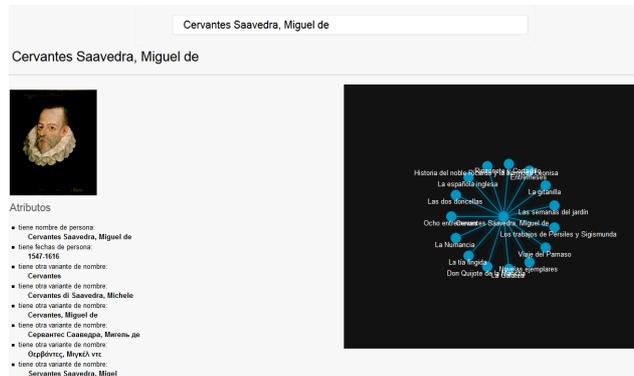
Fuente: Europeana SPARQL Endpoint <<http://labs.europeana.eu/api/linked-open-data/introduction/>>

Fuentes procedentes de bibliotecas: datos bibliográficos y de autoridades.

- Datos.bne.es: es la interfaz de búsqueda de los registros bibliográficos y de autoridad de la Biblioteca Nacional Española que han sido migrados a LD. La BNE inició en 2011 el proceso de migración de los catálogos bibliográficos (descritos en el formato MARC 21) y de autoridades (descrito en FOAF-EAC) a su equivalencia en el lenguaje RDF. También han enlazado con otros conjuntos de datos disponibles en LOD como VIAF, SUDOC (es el catálogo colectivo del Sistema Universitario de Documentación Francés, elaborado por las bibliotecas y los centros de documentación de enseñanza superior y de investigación), GND (German National Library), LIBRIS (es el catálogo colectivo de la

Biblioteca Nacional de Suecia) y DBpedia (Vila-Suero y Gómez-Pérez, 2013). La ilustración siguiente presenta la primera versión de la interfaz de búsqueda de datos.bne.es de 2011.

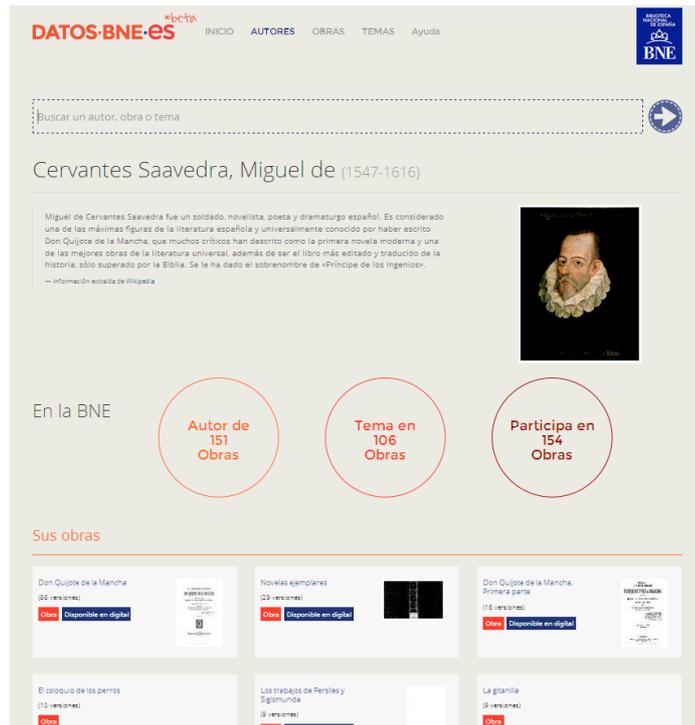
Ilustración 58: Interfaz de búsqueda de Bne.linkeddata.es en 2011



Fuente: Bne.linkeddata.es

La siguiente ilustración muestra la nueva versión de 2014 del portal de consulta que ofrece a los usuarios acceso a los datos del catálogo de la BNE.

Ilustración 59: Interfaz de búsqueda de datos.bne.es en 2014



Fuente: datos.bne.es

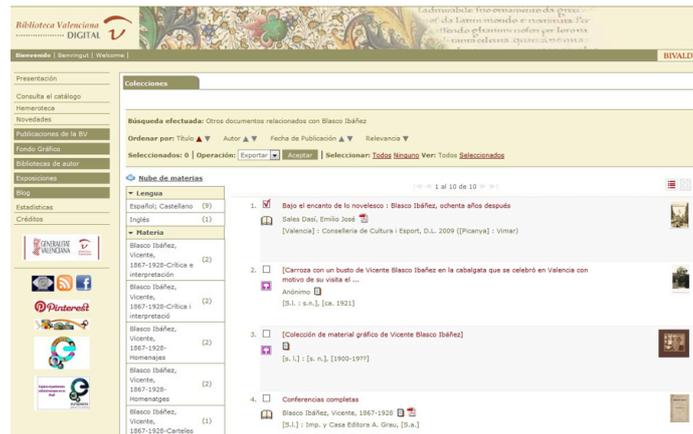
Los resultados de las búsqueda a través de esta interfaz muestran bloques de información de fuentes externas, como la biografía importada de DBpedia, la fotografía importada de Wikimedia y los ejemplares que se encuentran digitalizados en la Biblioteca Digital Hispánica. Esta fuente se verá en el siguiente capítulo, pues se consulta para enlazar con autores de Vestigium (Biblioteca Nacional Española, 2012). El valor de los

datos procedentes de la BNE es indiscutible. Como cabecera del sistema español de bibliotecas, como receptora de la cultura española a través del Depósito Legal y uno de los grandes silos de datos sobre el patrimonio bibliográfico español la producción de sus datasets semánticamente compatibles los pone al nivel de los grandes productores que vimos en la introducción de este apartado, las bibliotecas británica o francesa.

- Biblioteca Valenciana Digital (BIVALDI): en 2002 desde la Biblioteca Valenciana se puso en marcha el proyecto de digitalización de las obras literarias y científicas valencianas más relevantes y significativas, así como aquellas otras que han sido consideradas de un mayor interés para el desarrollo de la investigación científica sobre el patrimonio bibliográfico y cultural de esta comunidad autónoma (Biblioteca Valenciana, 2014). Hace años que es una biblioteca con contenidos abiertos, compatibles con el protocolo OAI-PMH que revisamos en un apartado anterior. A partir de ello, ha evolucionado para insertar sus fondos en la web semántica. Gracias a la utilización de programas de gestión bibliotecaria que cuentan con los protocolos más recientes del entorno bibliotecario, DIGIBIB, las fichas bibliográficas están descritas en RDF y permiten la exportación. Esta fuente

debe ser referente de consulta porque los contenidos que alberga son el acervo cultural valenciano y emplean las tecnologías web semánticas para el intercambio de registros. La siguiente ilustración presenta la interfaz de consulta de BIVALDI.

Ilustración 60: Interfaz de búsqueda de la Biblioteca Digital Valenciana



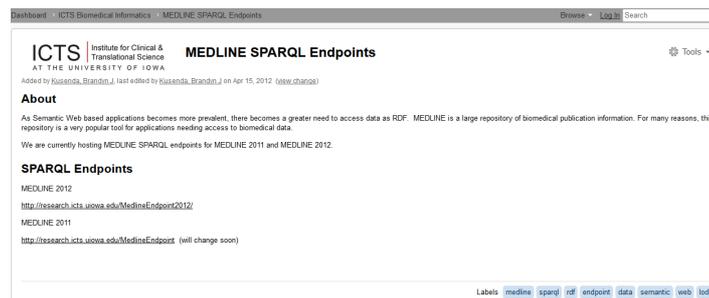
Fuente: Biblioteca Digital Valenciana
 <<http://bv2.gva.es/es/consulta/busqueda.cmd>>

Otras bases de datos de contenidos especializados son:

- Medline RDF: es una representación en RDF de la base de datos bibliográfica Medline. La información se estructura en cientos de datasets que contienen información sobre unos 19 mil artículos, con DOI y enlazados a revistas con Crossref. Utiliza otros

vocabularios de metadatos como RDF, RDFS, OWL, DC, VOID, BIBO, CiTO, FOAF, SKOS (OKFN, 2011). Esta fuente puede resultar útil para los investigadores de las áreas de conocimiento de ciencias de la vida que pueda recoger Vestigium. La ilustración siguiente presenta la interfaz de consulta Medline SPARQL Endpoint.

Ilustración 61: Interfaz de acceso al spaql endpoint de Medline



Fuente: Medline SPARQL Endpoint

<<http://research.icts.uiowa.edu/MedlineEndpoint2012/snorql/>>

- AGRIS: ya comentamos que es el repositorio institucional semántico que mantiene la FAO de UN. Los contenidos que almacena provienen de una red colaborativa con más de 150 instituciones participantes de 65 países distintos. Es una base de datos bibliográfica multilingüe para las ciencias agrícolas, que contiene más de siete millones de registros bibliográficos

normalizados con el tesoro AGROVOC. Utiliza la metodología de LOD y enlaza a datos relacionados de recursos web como DBpedia, World Bank, Nature, FAO Fisheries and FAO Country profiles (Jaques et al., 2012 y Anibaldi et al., 2013). La ilustración siguiente presenta la interfaz de consulta de AGRIS.

Ilustración 62: Interfaz de búsqueda de AGRIS

The screenshot displays the AGRIS search interface. At the top, there is a search bar with the text "Find resources..." and a magnifying glass icon. To the right of the search bar are links for "About" and "Feedback". The main content area shows a search result for the paper "Relation between the aluminum content of the faujasite framework and the isomerization and disproportionation of m-xylene" by Corma, A. The authors listed are Corma, A., Jacobs, P.A., Fornés, V., Pérez-Panente, J., and Martens, J.A. The paper is associated with the ACS symposium series (Journal) and has a start date of 1974. The source is identified as the Information Systems Division, National Agricultural Library. On the right side, there are sections for "Powered by Google" with related articles, "Data from www.nature.com", and "Data from DBpedia". At the bottom, there is a translation tool powered by Google and a footer with logos for CIARD and AIMS, along with the text "AGRIS: International Information System for the Agricultural science and technology".

Fuente: AGRIS <<http://agris.fao.org/agris-search/index.do>>

AGRIS es también una aplicación web de tipo *mash up* que enlaza con datos LOD alojados en fuentes externas. De modo que al realizar una consulta en su interfaz muestra también bloques de datos importados (Jaques et al., 2012 y Anibaldi et al., 2013). En el ejemplo mostrado, para la búsqueda del autor Avelino Corma

obtiene información de la fuente Nature y DBpedia, como se observa en la imagen.

Esta breve selección de fuentes constituye un somero estudio de las fuentes libres para enlazar. Se ha partido de ella como propuesta inicial para enlazar los contenidos de Vestigium con otros conjuntos de datos relacionados. Establecer enlaces de conexión con otras fuentes llevará a Vestigium a formar parte de la nube de LOD. Para alcanzar este último principio de buenas prácticas y participar de la red de contenidos semánticos es necesario realizar los procesos de búsqueda de datos semánticos y enlace con conjuntos de otras fuentes como se verá en el siguiente capítulo.

Consulta de datos enlazados

Cabe insistir que el consumo de datos de la web semántica es un paso cualitativamente diferente respecto de los *mash up* que se conocen desde hace tiempo. Por una parte porque normaliza una única manera de integrar los datos, y por otra porque el uso de tecnologías semánticas supone en sí mismo una mejora con respecto a la recuperación que antes no existía.

Las formas de consulta de contenidos en la web semántica difieren de la búsqueda actual basada en la localización de palabras clave en documentos en lenguaje natural. En la web semántica se interroga sobre un dato o un atributo de una entidad, lo que permite filtrar la información. El resultado de una consulta sobre un dato puede devolver una única tripleta, o un listado de atributos puede devolver datos procedentes de varias combinaciones de tripletas (Morato et al., 2014).

Estas búsquedas a fuentes LOD pueden realizarse desde los puntos de consulta SPARQL Endpoint mediante lenguaje SPARQL, como las que acabamos de ver. SPARQL Endpoint es la interfaz general de consulta que recupera las sentencias RDF almacenadas en los *triplestore*, o repositorios de tripletas, que son sistemas de gestión similares a las bases de datos. Pastor-Sánchez (2012) explica el funcionamiento de este proceso de la siguiente forma: se registra en el sistema un módulo de SPARQL Endpoint y se lanza una consulta que recupera un conjunto de datos, que se manipulan como si fueran contenidos generados por el propio gestor de contenidos (CMS). Tanto la recuperación como la visualización se realizan de forma dinámica y aunque los datos no se incorporan al CMS, el gestor actúa como interfaz de recuperación y visualización.

Existen endpoints generales y específicos de conjuntos de datos de instituciones. En los generales, cuando se define la consulta

hay que identificar la fuente de datos a la que tiene que interrogar, y es posible combinar varias fuentes. Estas interfaces permiten la descarga de los datos recuperados en diversas sintaxis como HTML, XML, RDF/XML, JSON (Formato *JavaScript Object Notation*, para el intercambio de ficheros sin necesidad de XML), Javascript, Ntriples (Harris y Seaborne, 2011).

SPARQL Protocol and RDF Query Language es un lenguaje de interrogación que se utiliza para construir la consulta de tripletas RDF almacenadas. Define las tripletas con las que se interroga al sistema, o en este caso la web, y como resultado recupera las sentencias RDF de un grafo que responden a los criterios de búsqueda definidos. Esto permite recuperar las sentencias RDF que se encuentran distribuidas por la web y reutilizar las que resulten necesarias para otros propósitos. Forma parte de los estándares del W3C desde 2008 (Harris y Seaborne, 2011).

La interrogación de fuentes a través de la consulta con lenguaje SPARQL implica una serie de dificultades (Morato et al., 2014):

- Se trata de un lenguaje complejo, lo que plantea dificultades para formular la consulta ya que hay que conocer la sintaxis del lenguaje de búsqueda y la posibilidad de combinar tripletas y filtros.

- Pueden darse inconsistencias entre los diversos vocabularios al definir un mismo concepto.
- La dificultad de expresar determinadas ideas en una única tripleta.

Debido a las dificultades que plantea la consulta mediante SPARQL, en el siguiente capítulo se consultarán fuentes LOD a través de sus interfaces de consulta y mediante buscadores semánticos.

Las búsquedas de LOD también pueden realizarse utilizando buscadores web semánticos, que pueden navegar entre diferentes fuentes de datos siguiendo los enlaces RDF. Los buscadores web semánticos proporcionan una interfaz amigable de interrogación, que puede presentar formularios, grafos o permitir el uso de lenguaje natural (Morato, 2014). Existen ya algunos buscadores semánticos como WolframAlpha, OpenLink Data Explorer, Sgoogle, Kartoo o DuckDuckGo. Realizan búsquedas y descubren y recuperan fuentes de datos enlazados estructurados expuestos en la web.

Por ejemplo, Síndice es un motor de búsqueda que recopila documentos RDF y los indiza para la recuperación de objetos digitales (Oren et al., 2008). Ofrece una interfaz de búsqueda sencilla basada en palabras clave para que los usuarios humanos puedan encontrar los documentos (Morato et al., 2014). Indiza y

localiza documentos web que contengan sentencias sobre un determinado recurso. Además, permite a los clientes y agentes web semánticos recuperar datos e integrarlos en un corpus de información unificado (Pastor-Sánchez et al., 2013). Constituye una herramienta válida para buscar autores y recursos sobre los autores de Vestigium. La ilustración siguiente presenta la interfaz de búsqueda de Sindice.

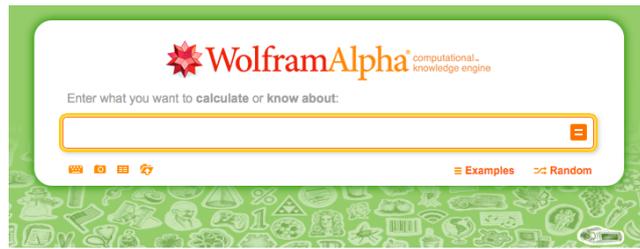
Ilustración 63: Interfaz de búsqueda del índice semántico Sindice



Fuente: Sindice <<http://sindice.com/>>

Otro ejemplo de buscador semántico es WolframAlpha, desarrollado por la compañía Wolfram Research en 2009. Está basado en la plataforma computacional Mathematica desarrollada por esta misma compañía. La ilustración que continúa presenta su interfaz de búsqueda.

Ilustración 64: Interfaz de búsqueda del buscador semántico WolframAlpha



Fuente: WolframAlpha <<http://www.wolframalpha.com/>>

Falcons es un buscador semántico orientado al usuario, con un formulario sencillo como interfaz, de forma similar a los buscadores actuales como Google o Yahoo. La interacción del usuario es a través de una caja de búsqueda mediante palabras clave relacionadas con el ítem o tema. Proporciona opciones de búsqueda por objetos, conceptos y documentos, cada uno de los cuales conduce a diferentes presentaciones de los resultados. Es adecuado para la búsqueda de personas y lugares. La búsqueda por concepto está orientada a la localización por clases y propiedades de las ontologías publicadas en la web. La opción de búsqueda por documentos apunta a documentos RDF que contienen los términos especificados (Bizer et al., 2009). La ilustración siguiente presenta su interfaz de búsqueda.

Ilustración 65: Interfaz del buscador semántico Falcons



Fuente: Falcons

Algunos de estos buscadores como Síndice y WolframAlpha se verán en el siguiente capítulo, pues se han interrogado sobre contenidos relacionados con Vestigium.

Agregar conjuntos de datos (*aggregating LD*)

Agregar conjuntos de datos estructurados de otras fuentes y mostrarlos en la interfaz del recurso propio necesita de aplicaciones (API) de integración o complementos para los gestores de contenido. Cada vez son más las herramientas y complementos para los CMS que generan datos enlazados, aunque son necesarias aplicaciones y módulos complementarios para encontrar y mostrar de forma dinámica los datos en la propia interfaz (Pastor-Sánchez, 2012). Las aplicaciones de integración se basan en algoritmos que son capaces de obtener, a partir de un objeto digital descrito en RDF e identificado mediante una URI, por ejemplo, un autor, la URI

dereferenciable de VIAF, leer las variantes del nombre en FOAF, leer los enlaces externos de las propiedades del objeto digital, entre otras las de identidad (owl:sameAs) en otras fuentes. Además, muestran en HTML la información leída en la interfaz del recurso.

Un ejemplo de esta función aplicada a un recurso de información es AGRIS de FAO, que agrega datos de forma automatizada y muestran información relacionada en su propia interfaz procedente de otras fuentes, como hemos visto en este mismo capítulo.

LA IMPLEMENTACIÓN EN VESTIGIUM

Ya hemos visto la dificultad para obtener resultados relevantes de información científica y académica desde los buscadores generales debido a que no indizan determinadas fuentes de información, además de la producción masiva de contenidos digitales y la falta de descripción normalizada de los mismos. Estos problemas han dado lugar en las últimas décadas al desarrollo de nuevas normas de descripción bibliográfica adaptadas a los contenidos digitales y la puesta en marcha de la iniciativa de descripción semántica propuesta por el W3C, tal y como revisa el capítulo segundo. Emplear las buenas prácticas de descripción de los datos enlazados es una tendencia que va asentándose con solidez y de la que participan, como se ha visto, numerosas y autorizadas instituciones internacionales del ámbito cultural y bibliotecario, así como científico y gubernamental.

Los beneficios son diversos para distintos actores y grupos de interés que se encuentran involucrados en este avance. Así, la publicación de conjuntos de datos de investigación mejora la visibilidad de los datos en sí, pero también de la investigación. Por tanto, los datos se convierten en otro punto de acceso a la investigación y al patrimonio cultural, junto a las vías tradicionales de publicación de los trabajos en revistas o monografías. Este aumento de la visibilidad se ve acrecentado además por la posibilidad de reutilizar los datos para otros propósitos. Los datasets publicados como LOD fomentan la reutilización de los datos para generar nuevas hipótesis e

investigaciones, y sirven también como mecanismo para validar los resultados de la investigación, e incluso para ser explotados de forma interdisciplinar en otros modelos de negocio (Baker et al., 2011).

En este apartado vamos a plantear el diseño de una biblioteca digital de características semánticas, con contenidos vinculados a la historia de la ciencia y las universidades que incluyan básicamente metadatos bibliográficos, su objeto digital y los datos biográficos esenciales para comprender al personaje. Por esta naturaleza histórica consideramos que los contenidos que existen de libre disposición en las fuentes linked data pueden mejorar nuestros fondos. De hecho, en el área de la historia, mediante contenidos derivados de biógrafos o genealogistas, por ejemplo, se pueden enriquecer las descripciones de los recursos bibliográficos de las bibliotecas, añadiendo contexto informativo mediante vínculos a otros datos que habitualmente las bibliotecas no proporcionan. De forma que las biografías pueden enlazar con las colecciones bibliográficas, con datos geográficos, de autoridades, y crear una red de navegación por todos esos recursos (Baker et al., 2011). El valor de la información se acrecienta al proporcionar una información de contexto enlazando con otros datos complementarios de otras fuentes, creando una extensa red estructurada y relacionada semánticamente por la que navegar y acceder a datos contrastados y de calidad.

Además, la web semántica proporciona a las bibliotecas un modelo de cooperación e intercambio de datos bibliográficos que favorece en muchas ocasiones compartir las tareas de catalogación y reducir descripciones redundantes.

Justificación

Las instituciones culturales son los agentes encargados de la preservación y difusión del saber. Su propósito es crear objetos para la memoria colectiva, expandir el conocimiento, acercarlo a los usuarios y conservar el patrimonio. El patrimonio constituye un acervo de valor incalculable para cualquier país y su salvaguarda es una tarea que se aborda desde diferentes puntos de vista y por diferentes instituciones. Todas las entidades que preservan la memoria como pueblo (archivos, bibliotecas y museos) tienen un fin común, la difusión de la cultura. Tradicionalmente el patrimonio se contempló como el conjunto de bienes culturales: monumentos, cuadros, esculturas. Aunque, desde hace años se incluyen otros activos como las manifestaciones intangibles o la ciencia. Su salvaguarda es una labor en la que intervienen distintos aspectos y es tratada por diferentes instituciones culturales (Ayala-Gascón, 2013; Garzón-Farinós, 2011; UNESCO 1972, 1978, 1982, 2003).

Las contribuciones que algunas figuras ilustres han aportado al avance de una civilización son aspectos relevantes que están siendo estudiados. La historia de la ciencia se ha construido en multitud de ocasiones girando alrededor de la vida y la obra de una persona. Un acercamiento de este tipo permite, por un lado, dar a conocer a los próceres y rescatar sus experiencias, como

modo de legarlas al lugar donde ejercieron su profesión (Ayala-Gascón, 2013; Garzón-Farinós, 2011; Aupí y Brines, 1994; López Piñero, 2008, Pascual, 2013).

La producción científica y literaria de estas personalidades constituye una información de valor histórico y científico muy importante. A lo que cabe añadir los estudios biográficos e históricos posteriores que han desarrollado los historiadores de la ciencia y de la universidad sobre estas figuras. Los estudios de personas individuales y grupos han venido desarrollándose desde hace años (Ayala-Gascón, 2013, Garzón-Farinós, 2011). En tal sentido, el estudio de los profesores, ya sean científicos o humanistas, ha sido un ámbito muy cultivado, con técnicas prosopográficas, biográficas y bibliométricas que se han impuesto en las investigaciones sobre historia social de la administración y de las élites, la historia de la ciencia y la historia universitaria (Peset Reig y Peset Reig, 1974; Ayala-Gascón, 2013, Garzón-Farinós, 2011).

Son numerosos los estudios de catedráticos contemporáneos. Distintos autores se habían ocupado de periodos históricos de la universidad española; Mariano y José Luis Peset en *La universidad española* (1974) profundizaron en la historia de las poblaciones universitarias españolas y comenzaron en España los estudios prosopográficos. Hay numerosas aportaciones de historiadores de las universidades, y es en Valencia donde los

estudios están más avanzados. Desde 1730, su rector Francisco Ortí y Figuerola (1730), redactó el primer catálogo de sus profesores (Ayala-Gascón, 2013).

Los trabajos se suceden con las biografías de los catedráticos y opositores hasta el siglo XX, donde cabe destacar a historiadores como María Fernanda Mancebo que cubre el período que va desde la dictadura de Primo de Rivera a la república (Mancebo, 1994). Estudios parciales sobre facultades se encuentran en Blasco Gil (2000) para la facultad de derecho durante la restauración borbónica (Ayala-Gascón, 2013; Garzón-Farinós, 2011).

En la historia de la ciencia, ha sido fundamental la aportación de la obra de José María López Piñero y su escuela, con la publicación sobre personalidades valencianas y españolas (López Piñero y Micó Navarro, 1983), centrados, en muchas ocasiones sobre el ámbito sanitario del que parte su formación. Desde la perspectiva de la historiografía contemporánea ha ido surgiendo una cantidad apreciable de investigaciones que convertían a personajes individuales en el centro del discurso histórico. Para Valencia, nombres como Josep Lluís Barona, Francisco Vera Sempere, Juan Micó Navarro o Josep Bernabeu Mestre dan cuenta del trabajo realizado, que completa la perspectiva individual que se apuntaba con el estudio de las universidades (Ayala-Gascón, 2013; Garzón-Farinós, 2011).

La corriente iniciada por José María López Piñero y M^a Luz Terrada, la bibliometría y la sociología de la ciencia, cuenta con estudios que analizan la actividad científica de los grupos, instituciones y personas han alcanzado un importante desarrollo, pues la aplicación de métodos bibliométricos basados en análisis cuantitativos y estadísticos a la información bibliográfica recogida en las publicaciones científicas permite caracterizar de forma precisa la producción, el grado de colaboración y la difusión del conocimiento científico. Estos estudios se han traducido en publicaciones y trabajos inéditos sobre los individuos, colaboraciones e instituciones (Taton, 1987; Ayala-Gascón, 2013; Garzón-Farinós, 2011).

El objeto de esta tesis es la construcción de Vestigium: Biblioteca Digital de Científicos y Humanistas Valencianos para recoger, aglutinar y preservar en un entorno digital estos estudios relativos a los científicos y humanistas del territorio valenciano que, por su trayectoria académica y profesional, se han convertido en referente en su área de conocimiento. Se tiene especial interés en personalidades que han ejercido durante el siglo XX, pues hubo progresos en todos los ámbitos de la ciencia y cabe destacar la contribución de científicos valencianos ilustres en el desarrollo científico, especialmente en ámbitos de biomedicina, de tecnologías aplicadas y en el desarrollo agroalimentario (Ayala-Gascón, 2013). Actualmente la Comunidad Valenciana es la cuarta en producción científica

en España en el quinquenio 2007-2011 según el Informe FECYT de 2013 (Moya-Anegón, 2013), el Informe COTEC 2013 y el Informe anual de la Generalitat Valenciana de 2011 y de 2012, todos ellos sobre la actividad científica española. Como se refleja en la siguiente tabla la producción valenciana indizada en las bases de datos internacionales de la Web of Science (WOS) sigue su trayectoria ascendente y continuada desde 1990, pasando de un 6,5% del total español en ese año al 12,2% en el año 2009.

Tabla 12: Producción científica (artículos en WOS) Comunidad Valenciana y España 1990-2009

Año	CV	España	%
1990	728	11.210	6,5
1995	1.784	19.085	9,3
2000	2.798	26.507	10,6
2005	4.361	37.095	11,8
2007	4.481	36.897	12,1
2008	4.869	40.019	12,2
2009	4.817	39.345	12,2

Fuente: Generalitat Valenciana. *Investigación Científica y el Desarrollo Tecnológico en la Comunitat Valenciana: Informe anual 2012*, Valencia: Generalitat Valenciana, 2013

En 2009 se publicaron 4.817 artículos en publicaciones internacionales y en el trienio 2007-2009 ha aportado el 12,2% de la producción científica española en el ámbito internacional con un crecimiento que se mantiene estable (Generalitat Valenciana, 2012).

Así pues, se realizó un examen de la competencia a través de búsquedas en Internet para detectar productos similares que pudieran servir de referencia para Vestigium por sus contenidos o por sus tecnologías. Como hemos observado en capítulos anteriores, existen ya bibliotecas digitales con contenidos publicados como LOD. Son representativas la Biblioteca Nacional Española, la British Library, la Biblioteca Nacional de Hungría, la Deutsche National Bibliothek, la Bibliothèque Nationale de France, la Open Library, o Europeana como canal de distribución de contenidos ajenos. En todas ellas se ofrecen datos bibliográficos, aunque la última mencionada Europeana, se centre con mayor atención a los materiales llamados especiales en la comunidad bibliotecaria, como manuscritos, fotografía o partituras.

En general no ofrecen información de contexto a los objetos digitales excepto un caso que merece una mención aparte: la Biblioteca Virtual de polígrafos de la Fundación Ignacio Larramendi. Es un proyecto semejante al nuestro puesto que versa sobre científicos y humanistas españoles y americanos de todos los tiempos. Esta biblioteca publica sus datos como LOD, aunque no recoge específicamente las personalidades valencianas del siglo XX que son el objeto de nuestro trabajo en Vestigium.

Por último, en cuanto a los aspectos de programación cabe decir que el mismo gestor de contenidos que hemos seleccionado para

la biblioteca se usa de manera generalizada en proyectos de información. La existencia de otros recursos de información que emplean tecnologías semánticas y LOD resulta beneficioso para el proyecto al formar parte de la red de LOD y poder compartir datos.

Metodología

El concepto “científicos y humanistas valencianos” es lo que sirve de determinante para la selección de las personalidades. Desde un punto de vista metodológico general, se considera susceptible de ser incluido en el estudio cualquier persona vinculada con el mundo de la ciencia y las humanidades - universidades y otros centros de investigación-, nacida en la Comunidad Valenciana o que haya ejercido en ella la mayor parte de su vida profesional.

La selección de los humanistas y científicos responde a los siguientes criterios:

- Un criterio temporal: se establecen límites cronológicos por los que quedarán incluidos los autores que hayan ejercido durante el siglo XX.
- Un criterio de género: de manera que queden representados de manera uniforme ambos sexos.
- Un criterio de equilibrio geográfico: para que se tomen en consideración las tres provincias.
- Un criterio de publicación en el caso de los investigadores y científicos: los autores con mayor número de publicaciones.
- Un criterio de actividad en el caso de los humanistas: la selección se realizará inspirándose en un criterio amplio de la actividad que realizan. Según la definición de la

Real Academia Española de la lengua, se considera humanista a toda persona instruida en letras humanas. Otros autores definen este término como “todo autor versado o influido de modo preponderante por los autores clásicos” (González, 1989). En este sentido serán susceptibles de selección todos aquellos autores que han trabajado en el área de la historia, las letras, las biografías, la historia de las instituciones académicas y de investigación, así como las personas que han representado cargos académicos como rectores, catedráticos.

Las autoridades se localizan consultando fuentes especializadas ya que para cada área de conocimiento existen unas fuentes específicas. La metodología de trabajo consiste en buscar autores en cada una de las bases de datos de las áreas de conocimiento estableciendo Valencia, Alicante y Castellón como filiación. Después se estudian los resultados caso a caso para ver si cumplen los criterios expuestos anteriormente. Una vez identificados, la información biográfica e icónica de estos autores ya puede ser incluida en Vestigium. Los siguientes pasos de trabajo sobre estos autores se dirigen a la obtención de su producción científica y la gestión de sus derechos para asegurar una difusión de la información dentro de los límites legales que se impongan.

A continuación se establecen unas fuentes particulares a consultar para identificar científicos y otras para localizar humanistas:

1. Para seleccionar científicos e investigadores: las fuentes de información en las que se pueden consultar los autores que tienen mayor producción son:

a. Bases de datos

La Web of Science (WOS), EMBASE de Elsevier, Pubmed-Medline. Bases de datos ICYT, ISOC e IME del Consejo Superior de Investigaciones Científicas. Hay que tener en cuenta la limitación que presentan algunas de las bases de datos para obtener la producción científica del siglo XX, ya que en algunas de ellas sólo se dispone de documentos desde la década de los años 60 o los años 70. Para seleccionar de forma exhaustiva los científicos más productivos habría que consultar también los repertorios impresos previos a la automatización.

A modo de ejemplo de consulta a las bases de datos se expone una búsqueda realizada en WOS:

“WC=Medicine, Research & Experimental AND AD=Valencia”

Limitada cronológicamente desde 1900 hasta 2014

Los resultados que recupera se refinan con la opción de analizar que ofrece el propio sistema por: “Autores”.

Los cinco primeros autores son:

- Aliño Pellicer, Salvador F.
- Castell, José V.
- Llombart-Bosch, Antonio
- Herrero, María J.
- Gomez-Lechon, María José

Este ejemplo de búsqueda se repite con cada una de las capitales de provincia y de las materias por las que clasifica WOS (Anexo 1).

b. Fuentes bibliométricas

Se toman en consideración estudios que evalúan y cuantifican la actividad científica, los estudios de citas, los estudios de impacto, los análisis de coautorías y otros indicadores para medir la producción científica de un autor y su impacto. Permiten identificar y caracterizar la comunidad científica, los grupos que la integran, su relación y los investigadores que la componen.

Por ejemplo, los estudios sobre el profesor de química Eduardo Primo Yúfera (1918-2007):

- Aleixandre-Benavent, R.; Ayala-Gascón, M.; Gandía-Balaguer, A.; Moreno Gálvez, A.; Navarro Moreno, M. A.; Planes Ferrer, M. D. *Eduardo Primo Yúfera. Un adalid de la ciencia. Vida y producción científica*. Valencia: UPV-UCV, 2011.
- Ayala-Gascón, M.; Aleixandre-Benavent, R.; Gandía-Balaguer, A. “Eduardo Primo Yúfera, founder of Revista de Agroquímica y Tecnología de Alimentos and pioneer on Food Science and Technology research in Spain”. *Food Science and Technology Internacional*, 2011, vol.17 n. 6 pp. 549-556. doi: 10.1177/1082013211427619.
- Ayala-Gascón, M.; Aleixandre-Benavent, R.; Gandía-Balaguer, A. “Indicadores de actividad científica en investigadores singulares: perfil bibliométrico de Eduardo Primo Yúfera, expresidente del CSIC”. *Revista Española de Documentación Científica*, 2012, vol. 35 pp. 209-237. doi: 10.3989/redc.2012.2.887.

c. Portales y páginas webs institucionales

Sitios web de organismos académicos y de investigación, provinciales, como la Real Academia de Medicina y Ciencias

afines de la Comunidad Valenciana; y nacionales, como por ejemplo, el CSIC.

d. Fuentes bibliográficas

Estas fuentes proporcionan información biográfica y bibliográfica sobre autores distinguidos. A continuación se recoge una breve muestra de principales obras para todas las épocas de la historia de la ciencia que pueden servir para ilustrar el trabajo con científicos valencianos. Cabe distinguir tres tipologías a grandes rasgos, las obras sobre una persona, las obras sobre una profesión y los diccionarios u obras de referencia:

- Barona, José Luis. *José Chabás Bordeore (1877-1963): Tuberculosis y medicina social en la Valencia del primer tercio del siglo XX*. Valencia: Consell Valencià de Cultura, 2007.
- Bernabeu-Mestre, Josep. *La salut pública que no va poder ser: José Estellés Salarich (1896-1990): una aportació valenciana a la sanitat espanyola contemporànea*. Valencia: Consell Valencià de Cultura, 2007.
- Esquerdo Mániz, José. “Francisco Navarro y Rodrigo (1883-1882): último anatómico clásico de Valencia”. En:

Actas del Simposio 2002 sobre la Historia de la Medicina Valenciana. Valencia: Instituto Médico Valenciano, Ed. Denes, 2002; pp. 121-128.

- Hernández Iranzo, Bernardo José. *José M^a García-Blanco Oyarzabal, el hombre y su obra: una etapa de la bioquímica y de la fisiología valenciana (1941-1968).* Tesis doctoral. Directores: Romero Gómez, Francisco Javier; Barona Vilar, José Luis. València: Universitat de València, 1999.
- López Piñero, José María; Micó Navarro, Juan. *Las publicaciones valencianas de Cajal.* Valencia: Universidad de Valencia, 1983.
- López Piñero, José María; Gómez Reig, Constantino; Navarro Pérez, Jorge. *Los estudios sobre la salud pública en la ciudad de Valencia, 1880-1990: Constantino Gómez Reig.* València: Ajuntament de València 1994.
- López Piñero, José María. *Enrique Ferrer y Viñerta (1830-1891) y la revolución quirúrgica en Valencia.* València: Ajuntament de València, 2000.
- Vera Sempere Francisco José. Los estudios médicos de Justo Ramón Casasús, el padre de Cajal, en la

Universidad de Valencia. *Conferencia Conmemorativa Fundacional del Instituto Médico Valenciano*, Valencia: Editorial Denes, 2002; pp. 11-51.

- López Piñero, José María; Navarro Brotóns, Víctor. *Historia de la ciencia al País Valencià*. Valencia: Edicions Alfons El Magnànim, 1995. ISBN 84-7822-154-9.
- López Piñero, José María. *La actividad científica valenciana de la Ilustración. Vol. 1 y 2*. Valencia: Diputación de Valencia, 1998.
- López Piñero, José María (coord.). *Historia de la Medicina Valenciana*. Valencia: Vicent García, 1992. ISBN 84-85094-68-9.
- López Piñero, José María; García Ballester, Luis; Terrada Ferrandis, María Luz; Balaguer i Perigüell, Emilio; Ballester Añón, Rosa M.; Casas Botellé, Francesc; Maset Campos, Pedro; Ramos García, Elvira. "Bibliografía histórica de la medicina valenciana". *Primer congreso de Historia del País Valenciano: celebrado en Valencia del 14 al 18 de abril de 1971*, Vol. 1, 1973. ISBN 84-600-5660-0, pp. 509-578.

- Navarro Brotóns, Víctor. *Científics valencians*. Valencia: Institució Alfons el Magnànim, 1981. ISBN 84-00-04993-4.
- Peset Reig, José Luis. “Sabios, científicos y técnicos en la ilustración valenciana”. En: Calle, Romà (coord.). *La Real Academia de Bellas Artes de San Carlos en la Valencia Ilustrada*. Valencia: Universitat de València, 2009. ISBN 978-84-370-7207-4, pp. 49-66.
- Peset Reig, José Luis; Hernández Sandoica, Elena. “Instituciones científicas y educativas” En: Laín, Pedro (coord.). *La edad de plata de la cultura española: (1898-1936)* Vol. 2, Madrid: Espasa Calpe, 1993 ISBN 84-239-4999-0, pp. 548-585.
- López Piñero, José María. *Diccionario Histórico de la ciencia moderna en España*. Vol. 1 y 2. Barcelona: Península, 1983.
- Osca Lluch, María Julia; López Ferrer, María Teresa; *Directorio de publicaciones científicas de la Comunidad Valenciana*. Valencia: Instituto de Historia de la Ciencia y Documentación López Piñero, 2000.

2. Para seleccionar humanistas: se plantean tres tipos de fuentes:

a. Fuentes bibliográficas

Dialnet, la Biblioteca Valenciana, las bibliotecas universitarias valencianas, los catálogos de bibliotecas nacionales: Biblioteca Nacional de España, base datos de libros editados en España ISBN del Ministerio de Cultura, o Red de bibliotecas universitarias-REBIUN.

Publicaciones bibliográficas y obras de referencia, tal y cómo se ha visto en el caso anterior. Son de especial relevancia para su consulta, las siguientes obras sobre una persona, obras sobre una profesión y los diccionarios y obras de referencia:

- Agramunt, Francisco. *Diccionario de artistas valencianos del siglo XX*. Valencia: Albatros, 1999.
- Andrés Montón, Joaquín ; [et al.]. *Quién es quién en la comunidad valenciana*. Valencia: Vicent García Editores, 1988. ISBN 84-85094-76-X.
- Associació Professional d'Il·lustradors Valencians (APIV) <<http://www.apiv.com/new/inicio.asp>>.
- Ballester Añón, Rafael (ed. lit.). *Poesía valenciana en castellano 1936-1986*. Valencia: Víctor Orenge, 1987. ISBN 84-86206-32-4.

- Base de datos sobre Teatro español. Valencia: Universidad de Valencia, 1996-. Disponible en <http://parnaseo.uv.es/Bases_teatro.htm>
- Bellveser, Ricardo; García, Manuel y de la Peña, Pedro J. *Clásicos valencianos contemporáneos*. València: Conselleria de Cultura, Educació i Ciència, 1988.
- Bellveser, Ricardo. *Vita Nuova: antología de escritores valencianos en el fin de siglo*. València: Ajuntament, 1993. ISBN 84-86908-99-X.
- Blasco Gil, Yolanda. "Entre la trayectoria universitaria y social: los catedráticos de derecho en Valencia, 1900-1939", en: Pavón Romero, Armando (coord.) *Promoción universitaria en el mundo hispánico, siglos XVI al XX*. México D.F., Universidad Nacional Autónoma de México, Instituto de Investigaciones sobre la Universidad y la Educación, 2012, pp. 191-233.
- Blasco Gil, Yolanda. "Catedráticos de la Facultad de derecho de Valencia, 1900-1939" en: Atienza Navarro, M^a Luisa; et al. (coords.). *Pensamientos jurídicos y palabras dedicados a Rafael Ballarín*, Universitat de València, 2009, pp. 143-163, ISBN 978-84-370-7681-2.
- Blasco Gil, Yolanda; Peset Mancebo, Fernanda. Ferrer Sapena, Antonia. 2013 "Notas sobre la trayectoria universitaria y social del catedrático de historia del arte

Felipe María Garín Ortiz de Taranco". *Cuestiones pedagógicas* 22 pp. 231-258.

- Catalán, Miguel. *Escritores en la Biblioteca: una selección de textos de la literatura valenciana actual*, Valencia: Biblioteca Valenciana, 2002.
- Cerdá Pérez, Manuel (dir.). *Gran enciclopedia de la Comunidad Valenciana*. Valencia: Editorial Prensa Valenciana, 2005. 17 vols (4) ISBN 84-87502-47-4
- *Diccionario histórico de la Comunidad Valenciana*. Valencia: Editorial Prensa Valenciana, 2004. 2 vols. ISBN 84-87502-30-X.
- García Lliberós, María; Parra, Blas. *Los escritores valencianos y su difusión universal (narrativa en castellano siglos XIX y XX)*. Valencia: Ayuntamiento, 2000. ISBN 84-95171-54-6.
- González Santander, Rafael. *Cátedras y catedráticos de Histología: evolución histórica de las cátedras de Histología de las facultades de medicina españolas (1873-1983)*. Alcalá de Henares: Universidad de Alcalá de Henares, Servicio de Publicaciones, 1994

- Gullón, Ricardo. *Diccionario de Literatura española e Hispanoamericana*. Madrid: Alianza Editorial, 1993. 2 vols. ISBN 84-206-5292-X.
- Herráez Serra, Miguel; Catalán, Miguel. *Diccionario de autores literarios de la Comunidad Valenciana*. Valencia: Institución Alfonso el Magnánimo, 1999. ISBN 84-7822-251-0.
- Huguet Chanzà, José; García, Manuel; Aleixandre, José ; et al. *Historia de la fotografía valenciana* Valencia: Levante, El Mercantil Valenciano, 1990. ISBN 8487502083.
- Llorens Serra, Manuel (dir.). *Gran enciclopedia valenciana*. Valencia: Difusora de Cultura Valenciana, 1990. ISBN84-87636-12-8.
- Mas Ivars, Miguel A. (dir.). *Gran Enciclopedia de la Región Valenciana*. Valencia: Gran Enciclopedia Valenciana, 1972. 12 vol.
- Peset Reig, Mariano. “Breve semblanza de los rectores de la Universidad de Valencia” En: Benito Goerlich, Daniel (coord.). *Arena números: colección de fotografía histórica de la Universitat de València*, 2006 ISBN 84-370-6350-7, pp. 159-172.

- Peset Reig, Mariano. “Juristas valencianos en el exilio”
En: Girona Albuixech, Albert y Mancebo, María Fernanda (coord.) *El exilio valenciano en América: obra y memoria*, Valencia: Universitat de València, 1995. ISBN 84-370-2312-2, págs. 157-180.
- Sansano, Gabriel (dir.). *Diccionari de la literatura valenciana actual (1968-2000)*. Institut Alacantí de Cultura Juan Gil-Albert, 2001. ISBN 84-7784-874-2
- Agulló, Víctor. *Los valencianos en Uruguay*. Valencia: Generalitat Valenciana, 2011

b. Archivos y hemerotecas

No hay que olvidar que los archivos son fuentes de especial relevancia para conocer a figuras valencianas con cargos en instituciones públicas, en universidades y sociedades científicas. Especialmente cuando los autores han recibido su formación, o han desarrollado cargos en estas universidades o academias. Se acudirá a ellos en los casos pertinentes. Algunos archivos relevantes son:

- Archivo de la Universidad de Valencia (AUV) y su hemeroteca.
- La Hemeroteca Municipal de Valencia.

- Los archivos de sociedades valencianas representativas como la Real Sociedad Económica de Amigos del País, la Real Sociedad Valencia de Agricultura y Deportes o el Ateneo científico y literario, la Academia Valenciana de Jurisprudencia y Legislación, Real Academia de Bellas Artes de Valencia.
- Archivos personales o privados. Para la consulta de estos archivos y la reproducción de materiales textuales, fotográficos o multimedia se establece un acuerdo de cesión de los documentos, que permita su depósito y difusión a través de Vestigium (Anexo 2).

c. Fuentes orales

Las fuentes orales son válidas como documentos para investigar el pasado de las autoridades a estudiar, pues pueden proporcionar testimonios de familiares, amigos, colaboradores y coautores que los conocieron personalmente. Principalmente permiten ampliar la información conocida sobre humanistas y académicos. Mediante las entrevistas a personas de su entorno es posible conocer relaciones familiares, vínculos con la universidad, con las instituciones académicas, conocer la motivación de su especialidad, la visión como mujer o varón en su época y otros aspectos de su trayectoria que no vienen reflejados en fuentes formales. Se realizarán entrevistas en los casos pertinentes, y se basarán en una metodología definida para

tal caso (Anexo 3). Las entrevistas originales serán conservadas en la biblioteca digital Vestigium como documento sonoro de producción propia, al tiempo que se extraerá información de referencia sobre el propio personaje.

Objetivos y público

El objeto de esta tesis se dirige a preservar el patrimonio científico y humanista valenciano, a mejorar la visibilidad digital de la ciencia valenciana en el mundo y a facilitar el acceso y la difusión de la cultura a través de las biografías y bibliografías en el ámbito digital. Todo ello mediante la aplicación de los avances que se han producido en el mundo web y las tecnologías de la información, en este caso particular las tecnologías semánticas que se han estudiado. El público al que se dirige este recurso incluye tanto a usuarios individuales como académicos, estudiantes y la sociedad en general, así como instituciones culturales, académicas y de investigación.

Los objetivos que persigue aplicando los protocolos de intercambio de información y buenas prácticas de descripción de LD estudiados en los anteriores apartados son:

- Diseñar el sitio web Vestigium aplicando los instrumentos y protocolos estudiados.
- Fijar los tipos de datos interoperables que puedan ser mostrados en Vestigium.
- Identificar el proceso de publicación de contenidos estructurados en Vestigium.

- Ofrecer contenidos propios en RDF para facilitar la reutilización de la información que preserva, mediante enlaces URI.
- Identificar las fuentes de datos LOD que pueden ser adecuadas para conectar con Vestigium.
- Conocer cómo se consumen datos de otras fuentes externas con Vestigium.
- Reutilizar contenidos LOD de otras fuentes externas para aportar un valor añadido a la información y participar de la extensa red de información enlazada.

Diseño e implementación

Para desarrollar el sitio web de la biblioteca digital se escogió un software de gestión de contenidos, Drupal. Los Sistemas de Gestión de Contenidos (*Content Management System*, en adelante CMS) son software que permiten la publicación estructurada de contenidos, separando las funciones de redacción y publicación, y la gestión de los mismos, de los elementos de diseño y presentación visual del sitio web (Saorín-Pérez, 2010).

En los últimos años los CMS han evolucionado notablemente en el entorno digital. Los software de gestión estandarizada se han consolidado como la opción más eficaz para manejar los websites. Los más extendidos se están adaptando a las pautas de la web semántica y cuentan con funciones de etiquetado en RDF. Existen muchas formas de publicar información en la web que aparentan calidad pero que esconden numerosos defectos y no aprovechan la riqueza del medio digital (Saorín-Pérez y Pastor-Sánchez, 2012). En este sentido, la publicación de contenidos con las tecnologías de datos enlazados proporciona numerosas ventajas a la hora de compartir información. Elementos como la visibilidad y la interoperabilidad se incorporan a los sistemas de información cuando las tecnologías semánticas forman parte de ellos, dotando de valor añadido a los

contenidos publicados en la web. La diferencia clave si se implementan tecnologías semánticas en las bibliotecas digitales consiste, como ya dijimos, en que amplían la información del contenido existente a través de los enlaces que proporciona a otros datos externos.

Entre los gestores de código abierto (*open source*) más conocidos -como pueden ser Joomla o Wordpress- Drupal presenta una gran capacidad de creación y desarrollo de servicios de información (Tramullas, 2010), y ocupa el segundo puesto en número de instalaciones a escala mundial (Water & Stone, 2011), por encima de Concrete5, Joomla, Xoops, Alfresco, Ezpublish, Opencms. *Open source* se refiere al software cuya licencia es libre o de código abierto, es decir, que incluye el código fuente y la propiedad de modificarlo, frente al software propietario que sólo incluye el programa ejecutable y emplea la propiedad intelectual o los derechos de autor para reforzar su propiedad (García Escribà, 2011).

Drupal se consolida, tras más de diez años de existencia y siete u ocho de desarrollo formal (Saorín, 2010), como una opción potente para proyectos que prevén crecimiento y con múltiples tipos de información, como es el caso de Vestigium. Constituye un CMS que permite desarrollar casi cualquier tipo de sitio web con funcionalidades muy diversas, mediante la incorporación de elementos específicos. Los módulos amplían las prestaciones

que ofrece su núcleo para adaptarse a las necesidades trazadas (Tramullas, 2013) -como gestión de usuarios, foros de discusión, galerías de imágenes, etiquetado social- por lo que resulta muy apropiado para alcanzar los requisitos necesarios de este proyecto.

Genera URL limpios y permanentes, un elemento importante que facilita la indexación por los motores de búsqueda y asegura su perdurabilidad en el tiempo. Junto con otras características que se recogen en el sitio web del proyecto Drupal.org, el elemento decisivo para su selección ha sido su capacidad de adaptación a las funciones de la web semántica. Es el principal CMS que en su núcleo dispone de herramientas que permiten utilizar tecnologías semánticas. Esta característica es importante ya que nos encontramos en un momento clave en el que la tendencia se dirige hacia una gestión de contenidos semánticos, en la que se requiere producir y consumir datos enlazados entre sistemas de información distribuidos en el universo web.

El CMS funciona en el servidor web en el que se aloja el sitio web con un servidor HTTP Apache, bajo el dominio www.vestigium.es. Desde la plataforma de administración phpMyAdmin se crea la base de datos en MySQL (sistema open source de gestión de bases de datos), y se activa la versión de lenguaje PHP (*Hypertext Processor*) que utiliza este CMS en el servidor para que todo funcione correctamente en un web de

tipo dinámico.

El paquete de software Drupal versión 7 se descarga desde el sitio web del proyecto Drupal.org. Mediante un programa de FTP (*File Transfer Protocol*) se transfirieron los archivos descargados de Drupal (denominado el *core* o núcleo) desde el ordenador hacia la carpeta raíz denominada `public.html` del servicio de alojamiento en el servidor o *hosting* contratado.

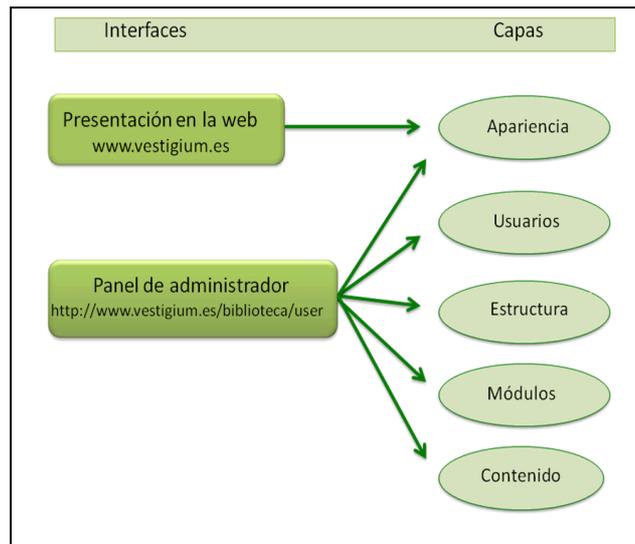
Con un navegador web y desde el dominio contratado se inicia el proceso de instalación de Drupal 7: el perfil de instalación, el idioma, nombre y contraseña de la base de datos que se ha creado y que se encuentra en el servidor, y a continuación se incluyen todos los campos de información del sitio web (nombre del dominio, datos del administrador, contraseñas). De esta forma queda instalado Drupal y se ha creado el sitio web.

Los elementos que se acaban de activar mantienen las conexiones para el funcionamiento del sitio con el sistema operativo, el servidor web, la base de datos y PHP. Sobre estos elementos necesarios se encuentra Drupal, formado por varias capas que se gestionan desde la interfaz de administrador. La capa superior es la interfaz pública, es decir, la presentación del web que ven los usuarios en sus navegadores, accesible desde www.vestigium.es.

En la ilustración siguiente se presentan de forma gráfica los

elementos que componen el sitio web. El sitio web cuenta con dos interfaces, por un lado la interfaz pública que visualiza todo el público desde la web y por el otro, el panel de administración de acceso restringido mediante el usuario y la contraseña que se ha creado.

Ilustración 66: Arquitectura de Drupal: Interfaces y capas



Fuente: Elaboración propia a partir de Tomlinson, Todd. Manual imprescindible de Drupal 7. Madrid: Anaya Multimedia cop., 2012

A la interfaz de administrador se accede desde la dirección <<http://www.vestigium.es/biblioteca/user>>. Desde este acceso se gestionan las cinco capas representadas en el ilustración 66 (Tomlinson, 2012):

1. Acceso y gestión de usuarios: permite la configuración de permisos de acceso, permite controlar los contenidos

que se pueden publicar, quién puede publicarlos, y qué pueden ver determinados tipos de usuarios.

2. Estructura: se organiza mediante los bloques y los menús. Permite categorizar los contenidos, etiquetado, relación entre el contenido, y mediante una taxonomía que se puede definir para organizar la información.
3. Módulos: el núcleo de la aplicación permite la gestión de contenidos, usuarios, menús, bloques y categorías, taxonomías y esquemas de clasificación, elementos suficientes para elaborar sitios con multitud de funciones. Aunque lo interesante del CMS es la capacidad de personalización del sitio que ofrece mediante el uso de módulos específicos que amplían sus prestaciones. Los módulos son pequeñas aplicaciones que expanden las capacidades del gestor. Han sido creadas por la comunidad de usuarios de Drupal y pueden descargarse desde el sitio web del proyecto.
4. Presentación visual al público: para que sea posible cambiar fácilmente el aspecto del sitio web sin modificar los contenidos, la apariencia se personaliza mediante plantillas o temas (*templates*). Permiten gestionar todos los elementos que forman el estilo, el diseño de las páginas del sitio web y la distribución, de una forma independiente al contenido.

5. Contenido: es toda la información que se publica y que será visible en el sitio web. Cada elemento de contenido, por ejemplo, una página o un artículo, se almacena como un nodo y se clasifica por tipos de contenido. Los tipos de contenido han de definirse y señalar a qué tipología pertenecen.

La interfaz de trabajo es el panel de administrador. Para gestionar las cinco capas, se puede acceder a cada una de las opciones del sistema mediante un menú superior. A continuación se detallan los trabajos realizados en Vestigium en cada una de las cinco capas de Drupal (Tomlinson, 2012).

Usuarios

Desde esta capa de Drupal se administran las funciones de seguridad. Mediante un sistema de roles se determinan los tipos de usuarios y los permisos que tienen cada uno de ellos para crear, actualizar y eliminar contenidos. Se entiende por usuarios aquellos que acceden al sitio web y se clasifican asignándoles una serie de roles que rigen los permisos de gestión del contenido. Pueden ser (Tomlinson, 2012):

- Visitantes: son usuarios anónimos, que visitan el sitio web externo sin identificarse ni registrarse. Pueden leer y

buscar contenidos en el sitio web externo.

- Usuarios autenticados: son aquellos que inician sesión con un nombre y una contraseña y pueden aportar información o hacer comentarios.
- Administrador: es el responsable del sistema. Es la cuenta que se crea de forma automática en el proceso de instalación. Es el responsable de la configuración del sitio web. Hay proyectos que tienen un único administrador, pero en este caso creamos un nuevo usuario, también con rol de administrador. De modo que todos los cambios realizados en la configuración y en el contenido, queden registrados en este perfil.

En la imagen que sigue se muestran los tres tipos de roles que pueden tener los usuarios y tres de los numerosos permisos que se gestionan desde la capa de usuarios.

Ilustración 67: Roles de usuarios y permisos

PERMISO	ANONYMOUS USER	AUTHENTICATED USER	ADMINISTRATOR
Advanced help			
View help topics	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
View help popups	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
View help index	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Fuente: Panel de administrador de la Biblioteca Vestigium

Estructura

La estructura del sitio web está directamente relacionada con la organización y la distribución de sus contenidos, con el diseño y la apariencia del web con la finalidad de que sea útil, sencilla de utilizar y visualmente agradable para los usuarios: aspectos de los que se encarga la Arquitectura de la Información (en adelante AI). Esta disciplina se centra en la organización de la información, planificando el diseño del sitio web para que el acceso a los contenidos o recursos de información que alberga sea completamente sencillo para los usuarios.

El origen se le atribuye a Richard Saul Wurman en el año 1975 y la expansión del concepto a Peter Morville y Louis Rosenfeld en el año 1998 con la publicación del libro llamado *Information Architecture for the World Wide Web* por la editorial O'Reilly Media, en el que como misión de la AI destaca: identificar los objetivos del web, las expectativas de los usuarios, determinar los contenidos y facilitar el acceso mediante sistemas de organización (taxonomía, etiquetado, navegación). Son numerosos los autores que han expuesto las características de esta disciplina, por lo que no hay una única definición consensuada y definitiva. No obstante, en todas ellas el denominador común son los conceptos de estructura, organización, contenido, acceso y usuarios.

Mario Pérez-Montoro (2010) recoge que la AI se encuentra estrechamente vinculada a las disciplinas del diseño gráfico de la web, la usabilidad, la interacción persona-ordenador o la gestión de contenidos, puesto que son complementarias y necesarias en el proceso de implementación de un sitio web.

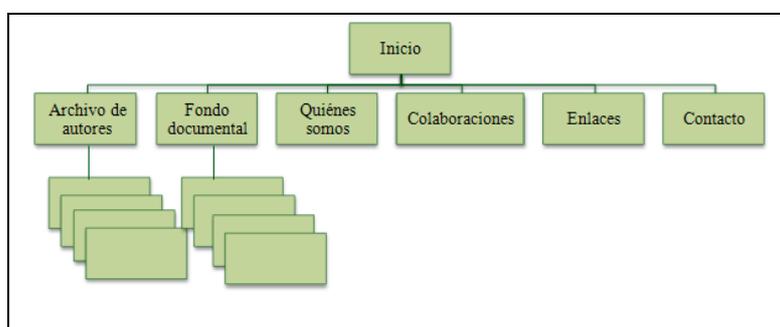
Para el sitio web Vestigium se pretende que las páginas sean útiles para el usuario, tenga un buen acceso a los contenidos y se alcance una interacción que satisfaga las necesidades de información. Para ello, como veremos en las páginas que siguen, se basa en algunas de las recomendaciones que emplea la AI en cuanto a:

- La organización de la información: definiendo un mapa del sitio y los tipos de páginas.
- Las áreas o bloques de contenido distribuidos en las páginas.
- Los elementos de navegación: menús, taxonomías, y vocabularios controlados.
- Los tipos de contenido que alberga.

Organización de la información

En relación con la organización de la información, el mapa del sitio web evidencia el orden de los contenidos. La estructura organizacional es jerárquica, con un elemento raíz o página inicial única de la que cuelgan las páginas que organizan el conjunto de contenidos del web. Posteriormente se definirán los sistemas de navegación que permitirán a los usuarios explorar cada una de las páginas que forman el sitio. La siguiente imagen ilustra la estructura organizacional del sitio web.

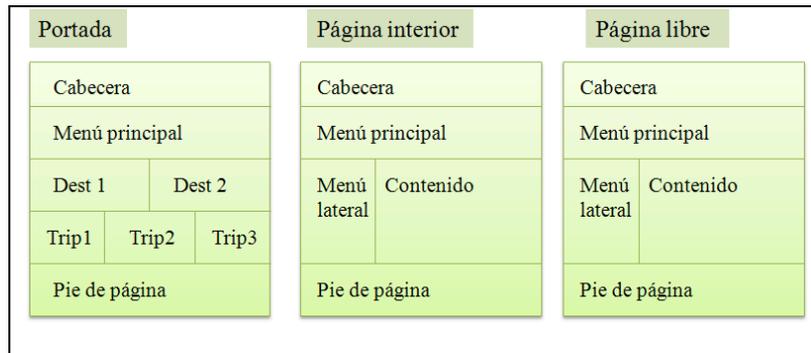
Ilustración 68: Mapa del sitio web Vestigium



Fuente: Elaboración propia en base a la estructura del sitio web Vestigium

Los tipos de páginas que contiene Vestigium son tres, como muestra la ilustración siguiente.

Ilustración 69: Tipos de páginas de Vestigium y áreas de contenido de la estructura interna



Fuente: Biblioteca Vestigium

- Portada: es la página inicial o *home*. Es única. Como ilustra la imagen 69 contiene el bloque de encabezamiento (cabecera), el bloque de menú principal, y los bloques de contenido destacado, ambos situados en el centro de la página (en la ilustración 69 se identifican como Dest1 y Dest2). Los tres bloques trípticos (en la ilustración 69 se identifican como Trip1, Trip2, Trip3), y el bloque de pie de página. En la ilustración siguiente se presenta la visualización de la portada de Vestigium.

Ilustración 70: Portada de Vestigium



Fuente: Biblioteca Vestigium

- Páginas internas: van asociadas a cada uno de los ítems del menú principal del encabezamiento por lo que, al abrirlas, aparece siempre seleccionada una entrada del menú. Estas páginas comparten el mismo menú lateral. Contiene el bloque de encabezamiento (cabecera), el bloque de menú principal, bloque de contenido y el bloque de pie de página. Son las páginas para el

“Archivo de autores, Fondo documental, Quiénes somos, Colaboradores, Enlaces y Contacto”. La ilustración siguiente muestra la página interna sobre “Quiénes somos”.

Ilustración 71: Ejemplo de página interna de Vestigium.



Fuente: Biblioteca Vestigium

- Páginas libres: tienen la misma estructura que las páginas internas pero no van asociadas a ninguna entrada del menú principal. En Vestigium son las páginas de cada uno de los autores, de cada uno de los documentos y de cada uno de los estudios bibliométricos. Es decir, son las

páginas para cada uno de los tipos de contenido de Vestigium. Por ejemplo, la imagen siguiente muestra la página libre sobre un documento de uno de los autores de Vestigium.

Ilustración 72: Página libre sobre un documento de Vestigium



Fuente: Biblioteca Vestigium

Es importante mantener una coherencia en todas las páginas, mediante una línea de colores continuada en el diseño y con una estructura uniforme, el encabezamiento, el menú principal y el pie de página fijos, para indicar a los usuarios que continúan navegando por un único sitio web.

Áreas de contenido

En cuanto a las áreas de contenido, los elementos de navegación y los tipos de contenido se trabajan en Drupal desde la capa de “estructura”, ítem del menú de la interfaz de administración. La ilustración siguiente presenta el menú de la interfaz de administrador con el ítem estructura seleccionado.

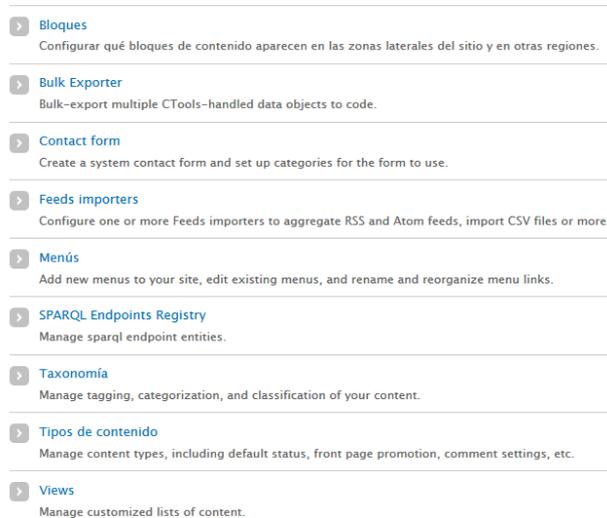
Ilustración 73: Menú de la interfaz de administrador de Drupal



Fuente: Interfaz de administrador de Vestigium

En la interfaz de “estructura” se gestionan los elementos: bloques de contenido, menús de navegación, taxonomía, tipos de contenido, vistas, *feeds* y otros elementos que incorpora el núcleo por defecto. La ilustración siguiente presenta estos elementos de la interfaz de estructura.

Ilustración 74: Bloques de trabajo en el apartado “estructura” de la interfaz de administrador de Vestigium



Fuente: Item “estructura” en la interfaz de administrador de Vestigium

El número de ítems que contiene el apartado “estructura” se ha visto ampliado desde el momento en el que se activaron e instalaron nuevos módulos (por ejemplo, RDF, SPARQL Endpoints, *Schema.org*), para adaptar el sitio a unas necesidades concretas relacionadas con la semantización de los contenidos.

A continuación se describen los trabajos realizados en cada uno de los apartados de “estructura” del sitio web.

1. Bloques

Esta sección permite configurar los bloques de contenido y su distribución en las páginas del sitio. Los bloques son

contenedores de información que se sitúan en diferentes zonas según la estructura que se quiera dar al sitio web y dependiendo también de la plantilla o tema que se utilice. Los bloques pueden mostrar contenido dinámico y funciones interactivas. Algunos de ellos se definen fijos en todas las páginas del sitio, y otros bloques sólo estarán visibles en algunas páginas específicas.

Desde la página de administración de los bloques, se activan y se configuran las características de la distribución de los contenidos de forma particular. Entre las opciones de configuración se encuentran la descripción del bloque, el título, el cuerpo y la ubicación. En las opciones de visibilidad se determinan las páginas específicas en las que cada bloque va a ser visible. También es posible hacer visibles los bloques según el tipo de contenido, los roles y los usuarios.

Los bloques de contenido creados son:

- Encabezamiento: en opciones de visibilidad se configura fijo para todas las páginas del sitio. Contiene el nombre del web, el logotipo de la biblioteca Vestigium, el menú principal y el buscador. El buscador es un módulo, *Search*, que debe ser activado desde el apartado Módulos para que aparezca en este listado.
- Destacado (*Featured*): es el bloque central de texto destacado. Contiene la información de presentación del

sitio web y del proyecto Vestigium.

- Contenido: el bloque central de contenido presenta una imagen de varios de los autores que se incluyen en la Biblioteca Vestigium, como muestra la ilustración siguiente.

Ilustración 75: Bloque de contenido central en la home de Vestigium



Fuente: Vestigium

- Barra lateral (*Sidebar*): constituye un menú lateral que contiene cuatro bloques de contenidos. Estos cuatro elementos son Vistas, listados que se crean y gestionan desde el módulo Vistas (*View*) que se verá más adelante. En opciones de visibilidad se configura que sea visible únicamente en las páginas internas, y no visible en la portada ni en las páginas libres. Los cuatro bloques de contenido en la barra lateral son:
 - Autores por localización: agrupa los autores que contiene el sitio en tres clasificaciones: Castellón, Valencia y Alicante.

- Autores por campo de actividad: agrupa los autores en dos clasificaciones: Científicos y Humanistas.
 - Documentos por formato: agrupa los documentos que contiene el sitio en tres clasificaciones: Vídeo, Texto y Otros formatos.
 - Estudios Bibliométricos: agrupa los documentos en dos clasificaciones: Humanistas y Científicos.
-
- Tríptico primero (*Triptych first*): tiene por título “Autores recientes”, que enlaza al listado de autores. Este elemento es una Vista que se crea y gestiona desde el apartado Vistas. Su visualización se muestra en la ilustración siguiente.
 - Tríptico medio (*Triptych middle*): con el título de “Documentos recientes”, enlaza al listado de obras. Este elemento es una Vista que se crea y gestiona desde el apartado Vistas. Se puede ver en la ilustración siguiente.
 - Tríptico último (*Triptych last*): con el título de “Estudios recientes”, enlaza al listado de estudios bibliométricos. Este elemento es una Vista que se crea y gestiona desde el apartado Vistas. Se puede ver en la ilustración siguiente.

Ilustración 76: Bloques de contenido “tríptico primero, medio y último” en la home de Vestigium



Fuente: Vestigium

- Pie de página: con los logotipos de las instituciones que forman parte del proyecto Vestigium, la información de contacto y el enlace a la política de privacidad del sitio.

Elementos de navegación y recuperación

El sistema de navegación y orientación es uno de los elementos más importantes para favorecer la usabilidad del sitio web y la interacción con el usuario. Proporciona elementos para desplazarse por las páginas del sitio de una forma sencilla con el objetivo de facilitar al usuario la información necesaria para saber dónde se encuentra en cada momento, de manera que puedan encontrar rápidamente la información que necesitan. Elementos como las taxonomías y vocabularios facilitan la recuperación de la información en el propio sitio web desde el

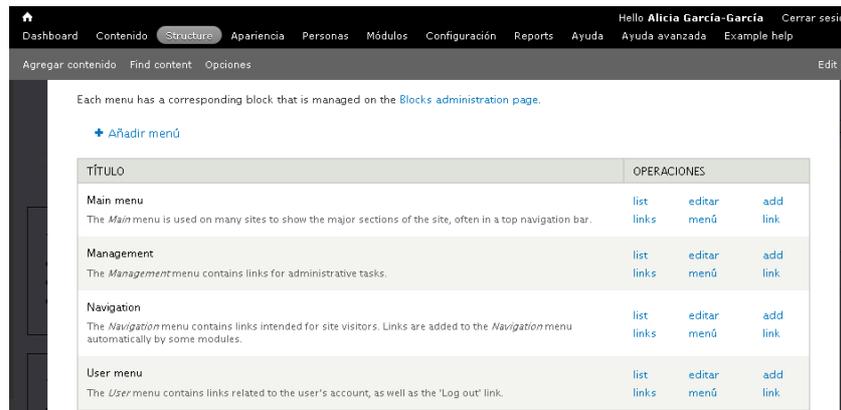
punto de vista de la AI; por ejemplo, cuando se interroga al sistema a través del propio buscador. Así también se emplean los vocabularios de valores para que el contenido sea indizado por los buscadores generales y nuestro sitio web sea recuperado ante una demanda de información.

1. Menús

Un menú es una lista de ítems que son enlaces que dirigen al usuario a otra página. Se organizan en vertical o en horizontal. Los menús son los principales elementos de navegación en un sitio web. Son los que permiten que el usuario se desplace de una página a otra, y facilitan su orientación, es decir, saber de dónde vienen, dónde se encuentran y a dónde pueden dirigirse en un sitio web. Son estructuras que ordenan y agrupan los contenidos bajo las categorías de una clasificación.

Desde la interfaz de administrador se gestiona la creación y eliminación de menús. Cada menú que se crea corresponde a un bloque y es gestionado desde la página de administración de bloques. Existen los menús únicamente visibles en la interfaz de administrador, que no son visibles desde la web pública.

Ilustración 77: Interfaz de administración de la estructura con los menús activados



Fuente: Interfaz de administración de Vestigium

Como se puede observar en esta ilustración, existen cuatro menús:

- Menú de administración del gestor de contenidos: permite al administrador gestionar contenido, estructura, módulos, personas, apariencia, configuración, ayuda.
- Menú de usuario: contiene los enlaces a la cuenta de usuario, como por ejemplo, cerrar sesión.
- Menú de navegación: para los usuarios administradores, permite crear contenido e importar vocabularios y contenidos.

Y el único menú visible desde la web pública:

- Menú principal: es el menú horizontal situado en el encabezamiento. Es el sistema de navegación constante en todas las páginas del sitio. Está formado por los ítems: Inicio, Archivo de autores, Fondo documental, Quiénes somos, Colaboraciones, Enlaces, Contactar, que son enlaces a cada una de esas páginas.

Un elemento importante a tener en cuenta cuando se crea la navegación, es resaltar visualmente el ítem del menú de la página en la que se encuentra el usuario, como recurso de contextualización para mostrar su ubicación en la jerarquía.

2. Taxonomías

La taxonomía, como vimos en el capítulo anterior, tiene que ver con los mecanismos para establecer las categorías en las que se pueden clasificar los contenidos. La taxonomía del sitio web es un vocabulario formado por un conjunto de términos para categorizar los contenidos. Desde el punto de vista de la AI, describir y categorizar los contenidos tiene como función facilitar la navegación por el sitio para localizar la información que el usuario necesita.

En el capítulo segundo se ha visto la importancia de los metadatos para aportar información de contexto de descripción de objetos digitales, con miras a facilitar la identificación, la localización y la recuperación de la información electrónica; así

como la necesidad de la aplicación de herramientas de control terminológico y de los sistemas de clasificación del conocimiento (taxonomías, tesauros y ontologías). De forma que en este apartado se crean las taxonomías necesarias. Y los tesauros y ontologías se implementarán mediante módulos específicos. Por otro lado, también se seleccionaron los vocabularios de valores RDF, ontologías OWL o SKOS (En adelante Vocabularios RDF) que más se ajustan a los contenidos de Vestigium, procedentes del inventario que propone el W3C y el proyecto LOV, como se expuso en el capítulo anterior.

Es posible que con un vocabulario sea suficiente, aunque en nuestro caso los contenidos se pueden clasificar en más de una categoría por lo que se crean varias taxonomías. Drupal crea por defecto un vocabulario denominado *Tags* (etiquetas) para los términos. Estas taxonomías en el ámbito de la web semántica que utiliza RDF se convierten en lo que comentamos en el capítulo anterior, el vocabulario SKOS. Como dijimos SKOS es el modelo de datos que se utiliza para transformar vocabularios de metadatos y vocabularios de valores (Schema, DC, FOAF) a la web semántica. En Vestigium se emplean conjuntamente varios vocabularios RDF según el tipo de contenido.

A continuación se crean las siguientes taxonomías para clasificar los contenidos por varios campos:

- Taxonomía Localización: para clasificar los contenidos según cuatro etiquetas:

Alicante, Castellón, Otras localizaciones, Valencia

- Taxonomía Campo de actividad: clasifica según las etiquetas:

Humanistas, Otros, Científicos

- Taxonomía Tipos de documentos: clasifica según las etiquetas:

Audio, Vídeo, Texto y Otros formatos

Estas taxonomías permiten clasificar el contenido de Vestigium para facilitar su consulta, como se muestra en la siguiente imagen, en la que los autores están clasificados por la taxonomía localización y recupera por la etiqueta “Castellón”.

Ilustración 78: Taxonomías de Vestigium



Fuente: Vestigium

Los vocabularios RDF se asignan a los campos de descripción de un tipo de información concreta, para lo que hay que determinarla. Estos tipos de contenido se verán en su apartado.

Conjuntos de elementos de metadatos y vocabularios de valores

Del inventario de Conjuntos de Datos, Vocabularios Controlados y Conjuntos de Elementos de Metadatos que recogió el Grupo Incubador de Datos Vinculados de Bibliotecas del W3C (2011) se han seleccionado aquellos que puedan ser útiles para emplear en Vestigium en función de los tipos de materiales que va a alojar. Se han clasificado en tres grupos, no todos procedentes del informe mencionado, que se describirán en detalle a continuación.

Conjuntos de elementos de metadatos

Un conjunto de elementos de metadatos, cuyo concepto se definió en el capítulo anterior, es una serie de elementos de información con los que se describen los objetos digitales (también llamados entidades). Definen las clases y los atributos

de las entidades que describen. Como se ha visto son la clave en las cuestiones de normalización e interoperabilidad. Vestigium publica sus contenidos con lenguaje RDF, por lo que los elementos de metadatos son el elemento predicado que vimos en la tripleta RDF y declaran la relación entre el sujeto y el objeto. Como se observa en la siguiente ilustración, en la tripleta el sujeto es una URI, el predicado es elemento de metadato “creator” del esquema DC, y el objeto el valor que adopta el recurso.

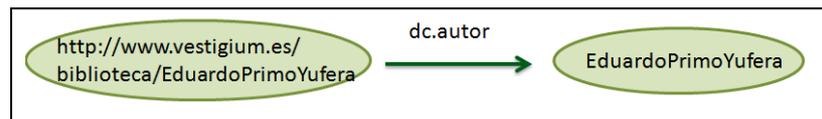
Ilustración 79: Ejemplo de un triplete de RDF de un autor de Vestigium

Sujeto:	http://www.vestigium.es/biblioteca/EduardoPrimoYufera
Predicado:	http://purl.org/dc/elements/1.1/creator
Objeto:	EduardoPrimoYufera

Fuente: Biblioteca Vestigium

La siguiente ilustración muestra la representación gráfica de la tripleta RDF que representa un autor de Vestigium.

Ilustración 80: Representación gráfica del triplete RDF de la imagen anterior



Fuente: Elaboración propia a partir de una tripleta RDF de un autor de Vestigium

A continuación se recogen los conjuntos de elementos de metadatos que se utilizan en Vestigium y que, como hemos visto, acabarán publicados como vocabularios RDF.

- **Dublin Core Metadata Terms**

En capítulos anteriores ya mencionamos con mayor o menor extensión el *Dublin Core Metadata*. En el capítulo primero se describió la versión inicial del conjunto de metadatos (*DC Metadata Elements*) de una forma detallada, y en el segundo capítulo se mencionó nuevamente, especificando el uso del conjunto completo del vocabulario (*DC Metadata Terms*) para Vestigium. En este apartado lo desarrollaremos.

DCMI *Metadata Elements* (1995) contó en sus inicios con los quince elementos originales para describir los recursos. Utilizaba el namespace `<http://purl.org/dc/elements/1.1.>`. Desde 2008 se ha ido ampliado el número de elementos. El conjunto completo del vocabulario es DCMI *Metadata Terms* (2012), que emplea el namespace `<http://purl.org/dc/terms/1.1.>`. Antes la codificación de los metadatos con los elementos DC definía los recursos como cadenas de valores literales, pero en esta versión los términos tienen asignados unas URI únicas para su identificación. Describe los recursos a través de los numerosos elementos DC y sus propiedades, basándose en RDF. Al estar basado en RDF

asocia a los valores literales a las URI, por lo que permite describir los datos como LD, lo que fija su valor.

La tabla siguiente muestra la diferencia entre el elemento de metadato creador como cadena de valor literal y como URI, aplicado a un autor de Vestigium. En principio es una cadena de caracteres, un literal o *string*, pero si la autoridad está disponible como LD, entonces el valor del metadato es una URI permanente para fijar su significado. Como se trata de un autor, con producción escrita, el valor lo encontraremos en VIAF, el conjunto de datos de autoridades de varias bibliotecas nacionales.

Tabla 13: Ejemplo de vocabulario de metadatos DCMI Terms y DCMI Elements

Dc:creator	String	Juan José Barcia Goyanes
Dcterms:creator	URI	http://viaf.org/viaf/55499576

Fuente: Vestigium

La siguiente tabla recoge los términos de *DCMI Terms* seleccionados para describir los campos de información de los tipos de contenido en Vestigium, como autor, estudio bibliométrico y documento.

Tabla 14: Esquema de descripción Dublin Core seleccionado

Nombre del elemento	Definición
Title	Nombre dado al recurso.
Creator	Entidad que es responsable principal de la elaboración del contenido del recurso.
Subject	Tema del contenido del recurso.
Description	Descripción del contenido del recurso.
Publisher	Entidad responsable de que el recurso esté disponible.
Contributor	Entidad responsable de realizar contribuciones al contenido de un recurso.
Date	Fecha de un hecho relativo al ciclo de vida del recurso.
Type	Naturaleza o género del contenido del recurso.
Format	Formato de archivo, medio físico o dimensiones del recurso.
Identifier	Referencia inequívoca al recurso dentro de un contexto determinado.
Source	Referencia a un recurso del cual deriva el que se está describiendo.
Language	Idioma del contenido intelectual del recurso.
Relation	Referencia a un recurso relacionado.
Coverage	Extensión o el alcance del contenido del recurso, la aplicabilidad espacial del recurso o la jurisdicción bajo la cual el recurso es relevante.
Rights	Información sobre los derechos contenidos en y sobre el recurso.

Fuente: Elaboración propia basado en Dublin Core Metadata Initiative (DCMI). *Dublin Core Metadata Terms*, 2012
<<http://dublincore.org/documents/dcmi-terms/>>

- **Schema.org**

Schema.org es una iniciativa puesta en marcha en 2011 por los motores de búsqueda generales Bing, Yahoo y Google con la intención de proporcionar un conjunto común de metadatos para estructurar las páginas web. Consiste en una colección de esquemas de metadatos que permiten a los diseñadores web incluir metadatos estructurados en las páginas web para describirlas y favorecer así que sean leídas por los tres buscadores, que las indexen y las recuperen en los resultados de las consultas de los usuarios. Ofrece un mecanismo para el marcado semántico del código HTML haciendo que una página web sea procesable por una aplicación o agente software para la extracción de información semántica (Pastor-Sánchez, 2012).

Es de carácter generalista, por lo que algunos de los elementos de varios conjuntos de metadatos se solapan. El namespace es `<xmlns:schema=http://schema.org/>`.

Se emplea en Vestigium para describir algunos campos, y asegurar así que los contenidos descritos sean recuperados por estos tres buscadores. En la imagen siguiente vemos un ejemplo de contenido autor que describe algunos campos de información

con Schema.org. En capítulos anteriores se examinaron otros estándares, como MADS o FOAF, pero Schema.org fue escogido por su versatilidad y extensión en el ámbito de la web semántica.

Ilustración 81: Extracto de algunos campos de descripción que emplean los metadatos del modelo Schema.org en el tipo de contenido Autor

Tipo schema:Person
Nombre completo schema:name, schema:givenName property
Biografía/Historia schema:description property
Fotografías schema:image property
Fecha de nacimiento schema:birthDate property
Lugar de nacimiento schema:addressLocality property
Fecha de defunción schema:deathDate property

Fuente: Biblioteca Vestigium

Vocabularios de valores

Los vocabularios se utilizan para asignar valores a los elementos de metadatos. Tal y como vimos en el capítulo anterior, de las fuentes presentadas (Linked Open Vocabularies, Vocabularios

controlados publicados en el Inventario del W3C, Proyecto HIVE ES) se han seleccionado los vocabularios de valores disponibles como LD que se han utilizado en Vestigium.

- **Lista de Encabezamientos de Materia**

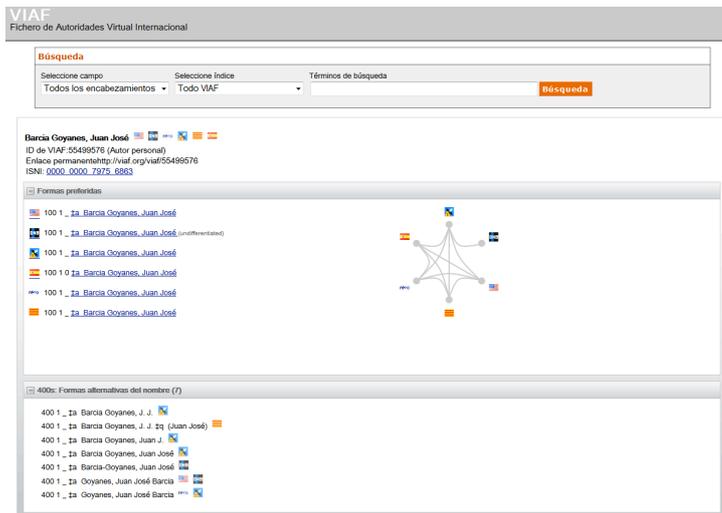
Para asignar valores a los metadatos de materia, seleccionamos la Lista de Encabezamientos de Materia (LEM) de las bibliotecas públicas de la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación Cultura y Deporte para aplicar en Vestigium. Recordemos que una lista de encabezamientos de materia recoge ordenadas alfabéticamente las materias o temas del documento para ayudar a recuperar por el campo materia y por los contenidos de temáticas afines gracias a las relaciones semánticas entre los términos. La Dirección General de Libro, Archivos y Bibliotecas puso en 2011 a disposición de la comunidad el conjunto de datos con los 17.323 términos de la LEM transformados a la estructura SKOS (Martínez Conde, 2011). Además, los datos se han enriquecido semánticamente al vincular con otras listas de encabezamientos de materia, un campo de equivalencia (skos:closeMatch). Se ha establecido el vínculo con los términos equivalentes de la *Library of Congress Subject Headings* (LCSH), que incorpora además un fichero en zip de los registros MARC de autoridades de materia, es decir, aquellos basados en el campo 150 en los

- **Virtual International Authority File (VIAF)**

El Fichero de Autoridades Virtual Internacional (VIAF), como se expuso en el capítulo anterior, es un registro con los ficheros de autoridad de las bibliotecas que forman parte del proyecto de la OCLC.

En la imagen que sigue se presenta la interfaz de VIAF y los resultados de una búsqueda de un autor que incorpora Vestigium, Juan José Barcia Goyanes. Es posible ver que recoge las variantes del nombre y los enlaces con las bibliotecas en las que existe esta autoridad.

Ilustración 83: Interfaz de búsqueda de VIAF



Fuente: VIAF<<http://viaf.org>>

En Vestigium se utiliza como valor para el elemento de

metadatos “autor” Juan José Barcia Goyanes.

Tabla 15: URI

Dc:creator	String	Juan José Barcia Goyanes
Dcterms:creator	URI	http://viaf.org/viaf/55499576

Fuente: Vestigium

- **GeoNames**

Es una base de datos que recoge nombres geográficos identificados por un URI estable, una descripción del recurso en RDF y todos sus nombres alternativos. Combina elementos de otros vocabularios como FOAF y DC Terms y se construye con la estructura que vimos en el apartado anterior para ontologías, la de OWL. Enlaza con numerosas fuentes LOD, entre las que destaca DBpedia. Es un proyecto de la compañía suiza Unxos GmbH (Switzerland) que fue fundado por Marc Wicky desarrollado por Christophe Boutreux. Cuenta con numerosos embajadores y colaboradores en muchos países del mundo, y numerosas comunidades en España. Este repositorio recoge en la actualidad diez millones de nombres geográficos, 2,8 millones de lugares con población y 5,5 millones de nombres alternativos (GeoNames, 2013). La ilustración que sigue presenta la interfaz de consulta.

Ilustración 84: Interfaz de búsqueda de GeoNames



Fuente: GeoNames <<http://www.geonames.org>>

La siguiente ilustración muestra la visualización en RDF de una consulta en GeoNames del nombre geográfico de Valencia y las alternativas del nombre en otros idiomas.

Ilustración 85: Ejemplo de un extracto de nombre geográfico de Valencia y algunos de sus nombres alternativos en un registro de GeoNames

```
- <rdf:RDF>
- <gn:Feature rdf:about="http://sws.geonames.org/2593113/">
  <rdfs:isDefinedBy rdf:resource="http://sws.geonames.org/2593113/about.rdf"/>
  <gn:name>Valencia</gn:name>
  <gn:officialName xml:lang="ko">발렌시아 지방</gn:officialName>
  <gn:officialName xml:lang="ja">ヴァレンシア</gn:officialName>
  <gn:alternateName xml:lang="ja">コムニダ・バレンシアーナ</gn:alternateName>
  <gn:alternateName xml:lang="fr">Communauté de Valence</gn:alternateName>
  <gn:officialName xml:lang="fr">Communauté Valencienne</gn:officialName>
  <gn:officialName xml:lang="es">Comunidad Valenciana</gn:officialName>
  <gn:officialName xml:lang="ca">Comunitat Valenciana</gn:officialName>
  <gn:alternateName xml:lang="it">Comunità Valenzana</gn:alternateName>
  <gn:officialName xml:lang="de">Landes Valencia</gn:officialName>
  <gn:alternateName xml:lang="ca">Pais Valencià</gn:alternateName>
  <gn:alternateName xml:lang="es">Pais Valenciano</gn:alternateName>
  <gn:alternateName xml:lang="en">Region of Valencia</gn:alternateName>
  <gn:alternateName xml:lang="es">Reino de Valencia</gn:alternateName>
  <gn:officialName xml:lang="en">Valencia</gn:officialName>
  <gn:shortName xml:lang="es">Valencia</gn:shortName>
```

Fuente: GeoNames <<http://www.geonames.org>>

Ya comentamos que en la web semántica el geoposicionamiento de los contenidos publicados está cobrando una gran fuerza. En el caso de Vestigium optamos por escoger Geonames por su

gran extensión entre la comunidad. Así como también la gran cantidad de datos que proporciona para los términos Castellón, Valencia y Alicante, que son relevantes para Vestigium. Sus antecedentes pueden rastrearse en el instrumento para control y recuperación de la información geográfica que produce el Instituto de Investigación Getty, el Geographic Names, cuya interfaz de consulta se muestra en la imagen que sigue.

Ilustración 86: Interfaz de búsqueda de nombres geográficos del GRI



Fuente: Getty Research Institute. Getty Thesaurus of Geographic Names <<http://www.getty.edu/research/tools/vocabularies/tgn/>>

No obstante, la producción de este recurso, Geonames, en un formato compatible con la web semántica asegura su utilidad para todos los campos del conocimiento. De nuevo, el hecho de la transversalidad de la naturaleza de su información produce un uso masivo como se aprecia en la siguiente ilustración.

En la tabla siguiente se presentan los tipos de contenido que se visualizan como páginas en Vestigium, y los términos de los esquemas de metadatos y vocabularios de valores asignados a los campos de información que describen estos tipos de contenido.

Tabla 16: Tipos de contenidos que se visualizan como páginas web dentro del sitio Vestigium

Tipo de contenido	Campos de información	Vocabulario RDF	Predicados RDF elementos de metadato
Página informativa	2 campos: título y cuerpo de texto	foaf:Document	Dc:title content: encoded
Artículo de opinión	4 campos: título, etiquetas, cuerpo e imagen	sioc:Item foaf:Document	Dc:title; dc:subject; content:encoded; og:image; rdfs:seeAlso
Autores	17 campos	schema:Person	schema:name; schema:image
Documentos	16 campos	schema:Book	Schema:name; schema:descriptio n
Estudios bibliométricos	2 campos: título y descripción	sioc:Item, foaf:Document	Dc:title content:encoded

Fuente: Vestigium

Para cada uno de los cinco tipos de contenido que se incluyen en Vestigium (páginas informativas, artículos de opinión, autores, documentos, estudios bibliométricos), se asignan términos de los vocabularios de metadatos y de valores. Se establece la relación entre los campos necesarios para describir el objeto de información, y los elementos de los vocabularios de metadatos para describir estos campos. Los vocabularios de metadatos se seleccionan del listado presentado anteriormente del inventario de Conjuntos de Elementos de Metadatos y Vocabularios de Valores del Informe Final del Grupo Incubadora de LD de Bibliotecas del W3C.

Para el tipo de contenido AUTOR se establecen 17 campos de descripción, a los que se le asignan elementos específicos de metadatos de los vocabularios Schema.org y OWL, como se puede observar en la tabla que sigue.

Tabla 17: Campos de información del tipo de contenido autor y elementos de metadatos de varios vocabularios

Campos de información de cada página de autor	Elementos de vocabularios de metadatos
Nombre completo	schema:name, schema:givenName
Fotografías	schema:image
Fecha de nacimiento	schema:birthDate
Lugar de nacimiento	schema:addressLocality

Fecha de defunción	schema:deathDate
Lugar de defunción	edm:deathPlace
Actividad laboral	Foaf:organization
Campo de actividad	Foaf:interest, dc:subject
Lugar dónde desarrolla su actividad	schema:homeLocation
Biografía/Historia	schema:description
Fuentes	dc:source
Ocupación	schema:affiliation
Documentos	Foaf:document
Estudios bibliométricos	Dc:title /qb:dataset
Otros accesos electrónicos	Dc:relation
Destacado	sioc:item
Mismo recurso (autor, documento) en otras fuentes	owl:sameAs

Fuente: Vestigium

Para el tipo de contenido DOCUMENTOS se definen 16 campos de descripción y los términos empleados para describirlos se utilizan del vocabulario de metadatos Schema.org, tal y como recoge la tabla siguiente.

Tabla 18: Campos de información del tipo de contenido documentos y elementos de metadatos del vocabulario schema.org

Campos de información de cada documento	Elementos de vocabularios de metadatos
Título del documento	schema:name
Descripción	schema:description
Lugar de publicación	http://purl.org/library/placeOfPublication
Fecha de publicación	schema: datePublished
Identificador	schema:isbn
Tipo	dc:type
Formato	schema:bookFormat
Recursos	Sioc:item
Materia	schema:about
Fuente	dc:source
Lenguaje	schema:inLanguage
Relación	dc:relation
Cobertura	dc:coverage
Derechos	schema:copyrightHolder
Autores	schema:author
URL	

Fuente: Vestigium

Para el tipo de contenido Página informativa se definen dos campos de descripción y los términos empleados para

describirlos se utilizan del vocabulario de metadatos DC y foaf, tal y como recoge la tabla siguiente.

Tabla 19: Campos de información del tipo de contenido página informativa y elementos de metadatos de varios vocabularios

Campos de información de la Página informativa	Elementos de vocabularios de metadatos
Título	Dc:title
Cuerpo de texto	content: encoded

Fuente: Vestigium

Para el tipo de contenido artículos de opinión se definen cuatro campos de descripción y los términos empleados para describirlos se utilizan del vocabulario de metadatos DC y foaf, tal y como recoge la tabla que sigue.

Tabla 20: Campos de información del tipo de contenido artículos de opinión y elementos de metadatos de varios vocabularios

Campos de información del Artículo de opinión	Elementos de vocabularios de metadatos
Título	Dc:title
Etiquetas	dc:subject
Cuerpo	content:encoded
Imagen	og:image

Fuente: Vestigium

Módulos

Los módulos son programas que proporcionan funcionalidades a un sitio web para adaptarse a las necesidades específicas de cada proyecto. Todas las funciones que permite Drupal, desde la redacción de contenidos, la publicación, la gestión, los idiomas, la inclusión de contenidos multimedia, elementos de localización, de búsqueda, elementos de mejora del posicionamiento, vinculación a redes sociales, *feeds* y todas las capacidades para componer un sitio web se incorporan mediante los módulos (Tomlinson, 2012). Éstos pueden ser de tres tipos:

- Los que forman el núcleo de Drupal: son un número de módulos básicos para la puesta en marcha de un sitio web.
- Los módulos contribuidos o contributivos: aquellos que expanden las capacidades del núcleo que han sido diseñados y desarrollados por la comunidad de Drupal. Existen miles de módulos clasificados en categorías según las funciones que proporcionan (utilidad, contenido, administración, comunidad, e-comercio, administración de usuarios, multilinguaje, SEO, RDF, prevención de *spam*).
- Los módulos customizados: son módulos escritos de forma personalizada para realizar tareas que solucionen

necesidades muy concretas. Se escriben nuevos módulos adaptados cuando las tareas que debe realizar son demasiado específicas que no hay módulos contributivos que las resuelvan.

Los módulos se instalan y gestionan desde la página Módulos de la interfaz de administrador. El proceso de instalación se puede realizar proporcionando la URL del archivo a instalar desde el sitio web de descarga de módulos de Drupal <https://drupal.org/project/project_module>. O bien, desde el sitio web de descargas de Drupal, se descarga el archivo en el ordenador y se sube al servidor del sitio web de Vestigium.

Desde el apartado Módulos se activan y desactivan las funcionalidades que queremos que tenga el sitio web. A continuación veremos los módulos básicos que vienen integrados en su núcleo y detallaremos aquellos que se descargan y activan específicamente en Vestigium para alcanzar las funcionalidades semánticas concretas que requerimos.

Al instalar Drupal vienen activados por defecto unos módulos que son requeridos por Drupal para su buen funcionamiento:

- Bloques (*Block*): es el encargado de gestionar los bloques de contenido, tal y como se ha visto en el apartado anterior.
- *Filter*: para el filtrado de contenido antes de mostrarlo.

- *Menu*: permite a los administradores personalizar el menú de navegación del sitio.
- *Node*: permite que se envíe contenido al sitio y que se exponga en las páginas.
- *Number*: define tipos de campos numéricos.
- *System*: para la gestión de la configuración general del sitio por los administradores.
- *Taxonomy*: permite la categorización del contenido.
- *Text*: define tipos de campo de texto simple.
- *User*: resulta necesario para administrar el registro de usuarios y el sistema de inicio de sesión.

Existen otros módulos que también forman parte del núcleo y que pueden ser activados. Algunos de los más importantes que se activan son:

- *Color*: que nos permitirá más adelante cambiar el color del tema seleccionado para ajustar las características de visualización según nuestros intereses.
- *Contact*: que habilita el uso de formularios de contacto, que es una de las páginas que cuelgan del menú principal del web.
- *Multilinguaje y Content translation*: va a permitir que el contenido sea traducido a diferentes idiomas. Ayuda a

proporcionar traducciones automáticas.

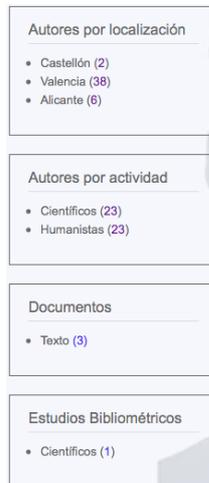
- *Corresponding node references*: permite crear enlaces bidireccionales entre nodos de contenido, es decir, en Vestigium establece enlaces que referencian tipos de contenido entre sí, por ejemplo, desde la página de autor, se define un campo de enlace a la página de su estudio bibliométrico.
- *Contextual links*: proporciona enlaces contextuales para llevar a cabo acciones relacionadas con los elementos de un página.
- El buscador (*Search*): para la búsqueda por palabras en el propio sitio web se configura, como se ha visto, desde el apartado de bloques de contenido, ubicándolo en la región del encabezamiento.
- *Feeds*: es un módulo para importar datos de fuentes externas. Permite importar cualquier tipo de contenido: autores, nodos, usuarios, taxonomías, desde diversas fuentes de datos y en varios formatos: csv, xml. Una posibilidad que ofrece el uso de este módulo junto con *Views*, es que permite mostrar el contenido que se importa en el mismo formato en el que se visualiza en la fuente de origen. Otra de las capacidades de este módulo es que es posible programar importaciones de información periódicas de forma automática. Lo que

puede resultar útil para importar contenido y mostrarlo en Vestigium.

- *Views*: las vistas permiten crear presentaciones muy variadas sobre el contenido. Para gestionar las vistas del contenido es necesario este módulo y su dependiente *Chaos Tool Suite*, conocido como *CTools*. *Views* es una herramienta para seleccionar y presentar el contenido mediante listas. Por ejemplo, mostrar los documentos de audio, los estudios bibliométricos, los últimos autores incorporados, los artículos más recientes publicados, o la galería fotográfica.

Las vistas se pueden incrustar como bloques de contenido en una página, como en nuestro caso que se sitúan en los cuatro bloques de la barra lateral de las páginas internas. La ilustración siguiente muestra la visualización de estos cuatro bloques en las páginas internas.

Ilustración 88: Vistas que organizan el contenido en cuatro bloques en las páginas internas



Fuente: Vestigium

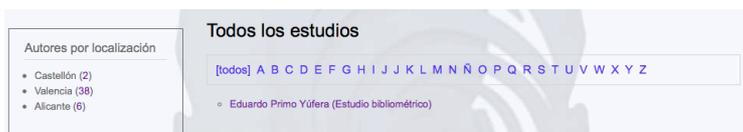
Pero además, estas vistas pueden visualizarse como páginas. En nuestro caso en la página de inicio, los tres bloques trípticos son vistas que se visualizan como páginas al clickar en cada uno de ellos, como muestran las dos siguientes imágenes.

Ilustración 89: Tres vistas en la home de Vestigium



Fuente: Vestigium

Ilustración 90: Visualización de la Vista Estudios bibliométricos como página



Fuente: Vestigium

Al crear las vistas se debe indicar la ruta de acceso de cada una de ellas para mostrarla como una página, no sólo como un bloque de contenido. Se crean todas las vistas que se han recogido en la tabla siguiente y se añaden las rutas de enlace que redirigen desde los tres bloques de contenido de la portada hacia las páginas.

Tabla 21: Vistas de los tres bloques de la home (autores, publicaciones y estudios bibliométricos) con URL de direccionamiento para visualizar como página

Vista	Ruta	Visualización
Autores Científicos	http://www.vestigium.es/biblioteca/cientificos	Página
Autores Humanistas	http://www.vestigium.es/biblioteca/humanistas	Página
Otros Autores	http://www.vestigium.es/biblioteca/otros-autores	Página
Autores de Alicante	http://www.vestigium.es/biblioteca/alicante	Página

Autores de Castellón	http://www.vestigium.es/biblioteca/castellon	Página
Autores de Valencia	http://www.vestigium.es/biblioteca/valencia	Página
Autores de otras localizaciones	http://www.vestigium.es/biblioteca/otras-localizaciones	Página
Documentos de audio	http://www.vestigium.es/biblioteca/audio	Página
Documentos de texto	http://www.vestigium.es/biblioteca/texto	Página
Documentos de vídeo	http://www.vestigium.es/biblioteca/video	Página
Documentos de otro tipo	http://www.vestigium.es/biblioteca/otrosformatos	Página
Autores recientes	http://www.vestigium.es/biblioteca/autores-recientes	Página y Bloque en la página de inicio
Documentos recientes	http://www.vestigium.es/biblioteca/documentos-recientes	Página y Bloque en la página de inicio
Estudios recientes	http://www.vestigium.es/biblioteca/estudios-recientes	Página y Bloque en la página de inicio
Estudios de científicos	http://www.vestigium.es/biblioteca/estudios-de-cientificos	Página

Todos los autores	http://www.vestigium.es/biblioteca/todos-los-autores	Página
Todos los documentos	http://www.vestigium.es/biblioteca/todos-los-documentos	Página
Todos los estudios	http://www.vestigium.es/biblioteca/todos-los-estudios	Página

Fuente: Biblioteca Vestigium

También se crean las vistas que aparecen en los bloques de la barra lateral. La tabla siguiente recoge las vistas de los cuatro bloques de las páginas interiores.

Tabla 22: Vistas de los cuatro bloques de las páginas interiores

Autores por campo de actividad	No tiene ruta de enlace porque es un bloque y no una página	Bloque en barra lateral de página interior
Autores por localización	No tiene ruta de enlace porque es un bloque y no una página	Bloque en barra lateral de página interior
Documentos por formato	No tiene ruta de enlace porque es un bloque y no una página	Bloque en barra lateral de página interior
Estudios bibliométricos	No tiene ruta de enlace porque es un bloque y no una página	Bloque en barra lateral de página interior

Fuente: Biblioteca Vestigium

Una vez creadas las vistas, aparecen visibles en el listado de bloques, desde donde se ubican en la barra lateral (*Sidebar*) de las páginas internas. De esta forma queda todo relacionado.

- *Taxonomy Menu*: transforma las taxonomías en menús, haciéndolos equivalentes a las clasificaciones. Es decir, posibilita editar los vocabularios y agregar los enlaces de los términos de la taxonomía como ítems de los menús.
- *Path*: permite cambiar el nombre a las URL, elemento importante para definir URI estables y permanentes, y mejorar el SEO. En Vestigium se utiliza para modificar las URL. Para cumplir el primer y segundo principio de buenas prácticas LD se tiene en cuenta las siguientes recomendaciones de definición de URI para hacerlas estables y permanentes (Berners-Lee, 1998):

Las URI son la identidad de cada página del sitio web, y las utilizan los motores de búsqueda para clasificar y entender mejor las páginas. Por ello los términos que se emplean deben ser palabras clave que representen bien el contenido, evitando palabras vacías de significado que puedan tener un impacto negativo en el SEO.

Además de no incluir detalles o términos relacionados con la implementación del sitio web en las URL (por ejemplo, Drupal) porque no aporta información con valor de perdurabilidad, pues podría realizarse una migración a otro gestor CMS en el futuro.

Así, en el mundo LD es importante cuidar las palabras que forman las URI porque ayudan a entender el tipo de entidad o recurso que están identificando.

Los siguientes ejemplos de URI lo evidencian:

<http://www.vestigium.es/vestigium_drupal/barciagoyanes>

<<http://www.vestigium.es/biblioteca/barciagoyanes>>

La primera no aporta información descriptiva del recurso y por lo tanto no ayuda a los motores de búsqueda a identificar el recurso e incluye detalles irrelevantes como Drupal. En cambio la segunda URI permite identificar que Vestigium es una biblioteca digital y es muy probable que “barciagoyanes” sea un autor.

Este tipo de URI con la base <<http://www.vestigium.es/biblioteca/>> es beneficioso para la visibilidad y estabilidad del recurso porque siguen las buenas prácticas de LD. En Vestigium se han pretendido URI limpias, estables y legibles por los sistemas automáticos.

- Módulo RDF

Ahora vamos a describir uno de los módulos que distinguían a Drupal del resto de CMS, el que es capaz de producir contenidos en RDF en su versión 7. Este módulo está incorporado en su núcleo y aporta el lenguaje RDF para la representación de nodos, taxonomías y otras informaciones de forma estructurada. Este soporte nativo proporciona al sistema la posibilidad de enriquecer el contenido con metadatos para permitir a otras

aplicaciones, como los motores de búsqueda y agregadores, entender mejor las relaciones y sus atributos.

Además, existen otros módulos complementario al RDF del núcleo, que complementan las capacidades del núcleo para formar parte de la web semántica y publicar los contenidos cumpliendo tres de los cuatro principios LD, que veremos a continuación.

Los módulos RDF enriquecen el contenido que se publica con unos metadatos que aportan a los contenidos la semántica necesaria para que sean legible por máquina. Esto se traduce en que las aplicaciones son capaces de comprender mejor los atributos de los objetos de información y las relaciones entre éstos.

La siguiente ilustración presenta el módulo RDF que forma parte del núcleo original del sistema Drupal 7.

Ilustración 91: Módulo de RDF que forma parte del núcleo original de Drupal versión 7

ACTIVADO	NOMBRE	VERSIÓN	DESCRIPCIÓN	OPERACIONES
<input checked="" type="checkbox"/>	RDF	7.0	Enriquece su contenido con metadatos para que otros programas (como motores de búsqueda o agregadores) entiendan mejor sus relaciones y atributos. Necesitado por: RDFx (activado), External RDF Vocabulary Importer (activado)	Ayuda

Fuente: Vestigium, interfaz de administrador

Una vez activado este módulo RDF del núcleo se pueden poner en funcionamiento los complementarios, para los que en

Vestigium contamos con la ayuda de un informático: External RDF Vocabulary Importer, RDF UI, RDFx, así como los relacionados con la generación de un punto de consulta SPARQL que son SPARQL API, SPARQL Endpoint y SPARQL Endpoints Registry.

- *External RDF Vocabulary Importer*: permite importar los vocabularios que necesita Vestigium de fuentes externas y exponerlos. Una vez importados hace las equivalencias con las clases y propiedades en la estructura de contenidos de Drupal. Como se ha visto en el capítulo segundo la utilización de vocabularios estandarizados incluye nuestros contenidos en la web semántica. Además, la reutilización de vocabularios es uno de los elementos importantes en la utilización de tecnologías semánticas. Como hemos ido desgranando, los vocabularios incluidos en Vestigium son DC, FOAF, SKOS, Schema, DC TERMS.
- RDF UI: permite a los administradores gestionar los mapeados o equivalencias de la estructura de datos del sitio, permite borrar, modificar o especificar mapeados para nuevos campos y tipos de contenido. Esta herramienta es clave en la administración de la coherencia interna de los contenidos. Dado que se usan diferentes vocabularios que son usados para múltiples tipos de contenidos en los websites a lo largo de la red,

un administrador de equivalencias permite controlar en cierta medida la consistencia en la descripción de la información.

- RDFx: esto complementa al anterior módulo. Amplía la funcionalidad de la API (aplicación) de mapeado de campos de RDF del núcleo de Drupal para proveer una mayor integración de formatos y otras capacidades. RDFx proporciona APIs para que los desarrolladores puedan manipular datos RDF, así como otros formatos de salida de datos desde Drupal como pueden ser RDF/XML, nTriples o Turtle. En el caso de Vestigium se han implementado para que se puedan reutilizar lo más posible estos datos.
- SPARQL API: es la aplicación que permite el uso de preguntas SPARQL.
- SPARQL Endpoint: con él se exponen los datos locales RDF en un SPARQL endpoint.
- SPARQL Endpoints Registry: proporciona un registro de SPARQL Endpoints.

La selección y activación de estos módulos favorecen seguir las buenas prácticas de LD al crear contenido, para lo que, como hemos comentado, contamos con la ayuda de un informático.

Producir contenidos enriquecidos mediante la asignación de metadatos, permite a otras aplicaciones como agregadores y motores de búsqueda entender mejor el tipo de contenido que es y las relaciones existentes entre los datos. El módulo RDF del núcleo y los complementarios facilitan exponer los datos internos como contenidos estructurados con RDF, donde se define tanto la estructura (sintaxis) de la información como el significado (semántica). Es decir, queda expresado el tipo de recurso y el tipo de relación que mantiene con otros recursos de información.

Al crear contenido en Vestigium, por ejemplo, una página sobre un autor, con el módulo de lenguaje RDF la información que se publica sobre este autor es estructurada y comprensible por las aplicaciones. Tal y como se ha visto en el capítulo segundo, se cumplen tres de los cuatro principios de buenas prácticas de LD al utilizar el lenguaje RDF y URI como nombres de las cosas. De este modo el sujeto es una URI que identifica el recurso por su nombre y localización de forma unívoca y permanente. El predicado es una URI que identifica el elemento del vocabulario DC de metadatos que determina la relación que existe entre sujeto y objeto. El objeto es el valor que tiene el sujeto, como se puede ver en la siguiente ilustración.

Ilustración 92: Ejemplo de enlaces RDF de un triplete de un autor de Vestigium

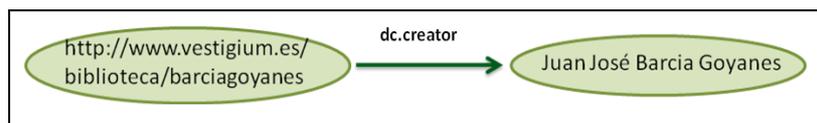
Sujeto:	< http://www.vestigium.es/biblioteca/barciagoyanes >
Predicado:	< http://purl.org/dc/elements/1.1/creator >
Objeto:	Juan José Barcia Goyanes

Fuente: Biblioteca Vestigium

La información del triplete RDF se interpreta de la siguiente forma: la página web de la biblioteca Vestigium cuya URI es <<http://www.vestigium.es/biblioteca/barciagoyanes>> es sobre el autor (elemento *creator* del vocabulario DC) llamado Juan José Barcia Goyanes. De esta forma, los científicos y humanistas valencianos entran en los nodos de la web semántica de una forma clara y precisa con la intención de ser el punto de referencia para los autores que incluya Vestigium.

La representación gráfica de este triplete, cuya información identifica el tipo de recurso del que se trata (página sobre un autor) y el valor que tiene (Juan José Barcia Goyanes), se muestra en la ilustración siguiente.

Ilustración 93: Representación gráfica del triplete RDF de la imagen 93

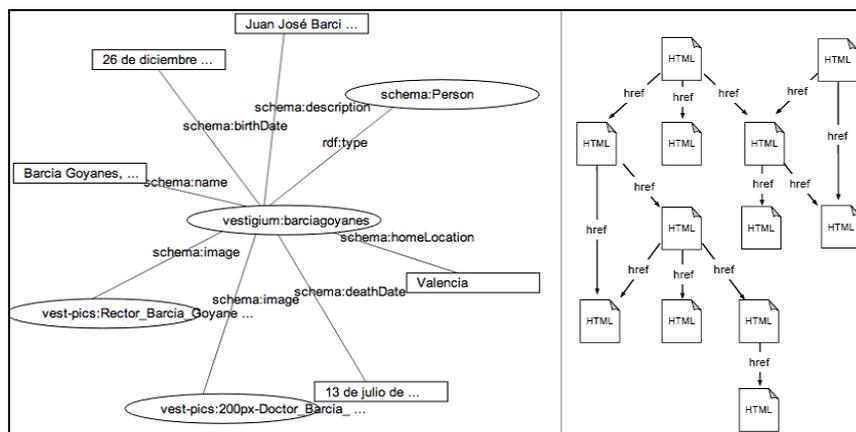


Fuente: Elaboración propia a partir de la triplete RDF de un autor en Vestigium

La información de este recurso propio puede ampliarse para crear una extensa red de información relacionada. El triplete inicial se amplía agregando nuevas estructuras de tripletas descritas con este mismo lenguaje RDF. Su composición identifica el tipo de recurso, los valores asignados por modelos de vocabularios de metadatos combinados, y las relaciones claramente identificadas entre sujeto y objeto.

Como se muestra en el ilustración siguiente, la información sobre el recurso se ha ampliado agregando tripletas con distintos modelos de vocabularios como Schema.org, Dublin Core y OWL que ampliará las posibilidades de ser entendida y reutilizada. Muestra a la izquierda la representación para la web semántica y a la derecha las relaciones con HTML, que producen la opacidad al no haber utilizado los vocabularios.

Ilustración 94: Representación gráfica de las relaciones entre datos de un documento HTML (grafo dirigido), y representación gráfica de las relaciones entre datos abiertos enlazados (grafo)



Fuente: Modificación basada en Castells, Pablo. La web semántica, 2003

- *Schema.org vocabulary*: es otro módulo que complementa al módulo RDF del núcleo para su buen funcionamiento. Este módulo *Schema.org vocabulari* permite describir el contenido utilizando el vocabulario Schema.org. Con este vocabulario es posible describir gran variedad de tipos de ítems (personas, objetos, documentos, cosas). Tal y como se ha visto en el apartado de tipos de contenido de Vestigium en este mismo capítulo, los ítems “autor”, “documentos” y “estudios bibliométricos” emplean varios campos del vocabulario Schema.org para describir el valor de los datos. Por ejemplo:

- Autor: nombre completo → schema:name
- Documento: materia → schema:about

En definitiva, hemos visto que alguno de los módulos que forman el núcleo de la aplicación, son clave para generar contenidos como LD, por ejemplo, *Path* para renombrar las URI, y el módulo RDF que produce tripletas. También la incorporación de otros módulos complementarios como *SPARQL Endpoint* y *Schema.org vocabulari* para asignar metadatos a los contenidos que se crean, que extienden las funcionalidades para seguir los cuatro principios de buenas prácticas de LD.

Apariencia

La capa superior del sitio web es la “Apariencia”. Ésta es la presentación que tiene el sitio web en Internet, y que será visible para todos los visitantes. Por ello es importante establecer un diseño visual atractivo para los usuarios. Además se han ido teniendo en cuenta las recomendaciones de usabilidad. Deben quedar claramente reflejadas la planificación de la estructura y la distribución del contenido.

El aspecto y la presentación visual de los contenidos es un elemento importante para crear una sensación positiva en los usuarios y para distinguir el sitio web del resto de sitios. La apariencia se modifica mediante temas (*themes*), tal y como explicamos. Con los temas se determinan los colores a utilizar en la página, las fuentes para el texto y los títulos, la distribución de imágenes y texto.

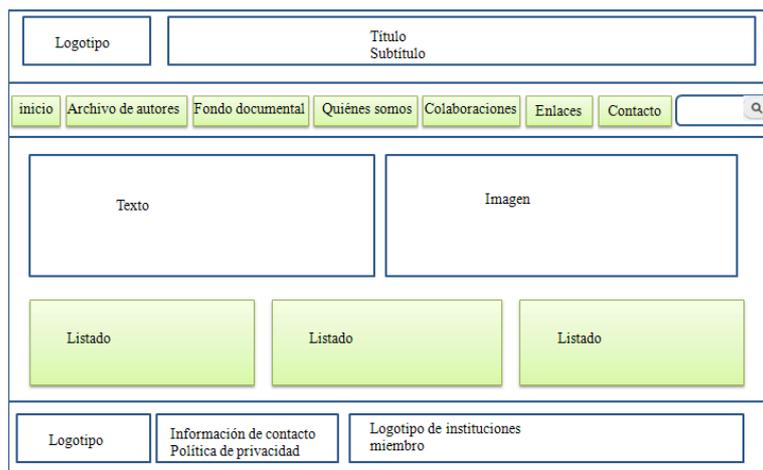
Drupal integra en su núcleo cuatro temas (*Bartik, Seven, Stark y Garland*), pero es posible aplicar otros gratuitos del repositorio proporcionados por la comunidad de colaboradores. Se establece como predeterminado el tema *Bartik*. Para seleccionar un tema es recomendable elegir uno sencillo y flexible que permita modificar cada uno de los elementos y adaptarlos a las características de nuestro sitio web. La instalación de los temas es sencilla. Se realiza desde “instalar nuevo tema” en la página de “Apariencia” en la que incluimos la URL del tema seleccionado; o bien, descargando el archivo desde la web de Drupal y subiéndolo al servidor.

En la página de “Apariencia” se muestran todos los temas descargados, divididos en los activados y los desactivados. Es posible tener activados varios temas al mismo tiempo, configurando desde las herramientas las opciones particulares de cada uno de los temas activados, y las opciones globales que se aplicarán para todos los temas.

En las opciones globales se determinan elementos como la imagen propia del logotipo, el título del sitio. En las opciones del tema predeterminado se establecen otros elementos como los colores del encabezamiento, el color del menú principal, la imagen de fondo, el título y el subtítulo, el color del pie de página, el color del texto.

Para los tres tipos de páginas que componen el sitio de Vestigium (portada, página interior y página libre), se diseñó una presentación específica. Para ello se contó con la aportación del diseño de María Ayala-Gascón. La ilustración 95 presenta el diseño de la página de portada.

Ilustración 95: Diseño de portada



Fuente: Elaboración propia basada en el diseño de María Ayala

La siguiente ilustración presenta el diseño de las páginas internas.

Ilustración 96: Diseño de página interna



Fuente: Elaboración propia basada en el diseño de María Ayala

La ilustración siguiente presenta la visualización de la portada.

Ilustración 97: Portada



Fuente: Vestigium

La siguiente ilustración muestra la visualización de la página interna.

Ilustración 98: Página interna

VESTIGIUM
BIBLIOTECA DIGITAL DE CIENTÍFICOS Y HUMANISTAS VALENCIANOS

[Inicio](#) | [Archivo de autores](#) | [Fondo documental](#) | [Quiénes somos](#) | [Colaboraciones](#) | [Enlaces](#) | [Contacto](#) [Buscar](#)

Todos los autores

[todos] [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [Ñ](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

- Andrés Estellés, Vicent
- Barcia Goyanes, Juan José
- Benlliure, Mariano
- Calatayud, Miguel
- Catalá Lucas, Rafael Miguel
- Cavanilles, José Antonio
- Corts Grau, José
- Carín Ortiz de Taranco, Felipe María
- Grisolia García, Santiago
- Lombart Rodríguez, Antonio
- Martínez Ruiz, José Augusto Trinidad
- Palop Gómez, José
- Sánchez Adell, Josep
- Segona, Tomás
- Viña Giner, José
- Amiches y Barrera, Carlos
- Benlliure, José
- Blasco Ibáñez, Vicente
- Candela, Manuel
- Cavadas Rodríguez, Pedro Carlos
- Corma Canós, Avelino
- Fuster i Ortells, Joan
- Gil Albert Simón, Juan
- Hernández, Miguel
- López Piñero, José María
- Miró Ferrer, Gabriel Francisco Victor
- Primo Yúfera, Eduardo
- Sanclis Grau, Josep
- Sorolla, Joaquín

Autores por localización

- Castellón (1)
- Valencia (23)
- Alicante (5)

Autores por actividad

- Científicos (9)
- Humanistas (20)

Documentos

- Vídeo (1)
- Texto (2)
- Otros formatos (1)

Estudios Bibliométricos

- Humanistas (1)
- Científicos (2)

©2011-2014 Vestigium
contacto@bibliotecavestigium.es
[Política de privacidad](#)

Universidad Católica de Valencia San Vicente Mártir

VNIVERSITAT ID VALÈNCIA

UNIVERSIDAD POLITÉCNICA DE VALÈNCIA

CSIC

Fuente: Vestigium

Contenidos propios y externos

Hasta aquí, se ha diseñado la estructura del sitio web, la distribución de los elementos, la organización de la información, las áreas de contenido y los tipos, se ha definido también el diseño visual y se ha implementado. A continuación, se procede a crear los contenidos. Para ello se va a hacer una distinción entre contenidos propios y contenidos externos procedentes de otras fuentes.

1. Crear contenidos propios

En el apartado “contenido” de la interfaz de administrador de Drupal, se crean los contenidos propios para cada tipología de páginas del sitio web. Como se ha visto, son tres los tipos de páginas: portada, página interior y página libre.

1.1 En la página de portada aparece:

- En el encabezamiento, sobre el menú principal y el buscador: se incrusta el logotipo de Vestigium diseñado por María Ayala. Es el que muestra la ilustración que sigue.

Ilustración 99: Logotipo de Vestigium



Fuente: Vestigium

- En el bloque de Texto Destacado se introduce como título “Bienvenidos al Archivo Digital Valenciano” y el siguiente texto:

“El grupo de investigación sobre científicos y humanistas valencianos de la Universidad Católica de Valencia San Vicente Mártir lleva a cabo una investigación que abarca el estudio biográfico y bibliométrico de científicos y humanistas valencianos del siglo XX, así como la gestión de la biblioteca digital que preserva la literatura vinculada a estas personalidades. En el desarrollo de la biblioteca digital colaboran investigadores de la Universidad Católica de Valencia, Universitat de València, Universidad Politécnica de Valencia, y del Centro Superior de Investigaciones Científicas (CSIC).”

- El siguiente apartado está formado por los tres bloques de contenido (*Triptych first, middle y last*), como se ha visto anteriormente en este mismo capítulo, con tres listados de nombres de autores, nombres de documentos y nombres de estudios bibliométricos, que enlazan a las páginas “Autores recientes” “Documentos recientes” y “Estudios recientes” respectivamente.
- En el pie de página se introducen los logotipos y enlaces

de las instituciones colaboradoras: Universidad Católica de Valencia, Universitat de València, Universidad Politecnica de Valencia y CSIC. Junto con la información de contacto de Vestigium y el enlace a la política de privacidad del sitio, como presenta la ilustración siguiente.

Ilustración 100: Pie de página del sitio web con el logotipo y las entidades colaboradoras



Fuente: Vestigium

1.2 En las Páginas interiores se introduce el siguiente contenido:

- Archivo de autores: tal y como se expuso en el apartado Módulos, esta página interior es una vista que muestra un listado con los enlaces a las páginas de cada uno de los autores.
- Fondo documental: del mismo modo que la página Archivo de autores, se trata de una vista que recoge el listado de todos los documentos y estudios bibliométricos que custodia el sitio.
- Quiénes somos: se incluye el siguiente texto: “El equipo Vestigium está formado por investigadores de las

universidades: Universidad Católica de Valencia, Universidad de Valencia, Universidad Politécnica de Valencia y CSIC”.

- Colaboraciones: en esta página se introduce el siguiente texto: “El equipo de la Vestigium invita a todos aquellos particulares, estudiantes, investigadores, centros culturales y otras instituciones patrimoniales interesados en el patrimonio científico y humanístico nacional a colaborar con la aportación de contenidos (artículos académicos o reseñas de libros, fotografías, imágenes o reproducciones) así como enlaces a contenidos relacionados, así como a formar parte de la red de contenidos linked data de la que participa esta biblioteca digital.”
- Enlaces: en esta página se recogen un conjunto de accesos a otras fuentes de información relacionadas, principalmente revistas especializadas. Se incluyen los títulos de las revistas y las URL. Un elemento importante a la hora de poner accesos a fuentes externas a nuestro sitio web es especificar que los enlaces externos se abran en una nueva página. Esto va a favorecer que la ventana de nuestro sitio web permanezca abierta y aunque el usuario continúe navegando por otros recursos, pueda volver cómodamente al sitio de origen. Además, este

sistema ayuda al usuario a identificar cuándo se encuentra en el mismo sitio web Vestigium, pues los enlaces se abren en la misma ventana. Por el contrario cuando se ha accedido a un sitio web distinto los enlaces se abren en ventanas nuevas.

- Contacto: se trata de un formulario de contacto para que los usuarios puedan comunicarse con la organización a través de un correo electrónico.

1.3 Páginas libres.

Cada una de las páginas libres se crea para un autor, un artículo o un estudio bibliométrico. A continuación se implementará información en cada uno de los distintos tipos de página cuyos contenidos originales fueron ya publicados por una parte del grupo de investigación en: María Ayala Gascón, Rafael Aleixandre Benavent y Asunción Gandía Balaguer “Indicadores de actividad científica en investigadores singulares: perfil bibliométrico de Eduardo Primo Yúfera, expresidente del CSIC” *Revista Española de Documentación Científica*, 2012. Vol. 35 n. 2, pp. 209-237,

- Página para un autor: se crea una página para Eduardo Primo Yúfera, como presenta la siguiente ilustración
cuya URI es

<<http://www.vestigium.es/biblioteca/EduardoPrimoYuferra>> con información biográfica extraída de la publicación.

Ilustración 101: Página de un autor



Fuente: Biblioteca Vestigium

- Página para un artículo: se crea una página de tipo documento. Se cumplimentan los campos con la información de la publicación seleccionada anteriormente. Se vincula la página de Vestigium creada sobre la publicación “Indicadores de la actividad” con el texto completo del artículo, que se encuentra almacenado en un repositorio institucional de acceso abierto, Digital.CSIC. Se vincula también mediante un campo de enlace con la página del autor en Vestigium. La visualización la muestra la ilustración siguiente.

Ilustración 102: Página de un documento

The image shows a screenshot of the Vestigium website. At the top, there is a logo for 'VESTIGIUM' and the text 'BIBLIOTECA DIGITAL DE CIENTÍFICOS Y HUMANISTAS VALENCIANOS'. Below this is a navigation menu with links: Inicio, Archivo de autores, Fondo documental, Quiénes somos, Colaboraciones, Enlaces, and Contacto. A search bar is also present.

The main content area features a sidebar on the left with four categories: 'Autores por localización' (Castellón: 2, Valencia: 38, Alicante: 0), 'Autores por actividad' (Científicos: 23, Humanistas: 23), 'Documentos' (Texto: 3), and 'Estudios Bibliométricos' (Científicos: 1).

The main article is titled 'Indicadores de actividad científica en investigadores singulares: perfil bibliométrico de Eduardo Primo Yúfera, expresidente del CSIC'. The author is listed as 'Primo Yúfera, Eduardo'. The article type is 'Artículo' and the format is 'Texto'. It was published in April-June 2012 in the 'Revista Española de Documentación Científica'.

The description provides a detailed overview of Eduardo Primo Yúfera's scientific career, mentioning his role as founder and director of the Institute of Agrochemistry and Technology of Foods, his work as president of the Spanish Council of Scientific Research (CSIC), and his extensive publication record in various scientific journals.

At the bottom of the page, there is a footer with contact information, logos for the University of Valencia and the Polytechnic University of Valencia, and the CSIC logo.

Fuente: Biblioteca Vestigium

- Página para un estudio bibliométrico. Contiene las figuras, gráficos, tablas, y datos empleados en el estudio. Cuenta con un campo de enlace al artículo al que pertenece. El ejemplo de estudio bibliométrico sobre Eduardo Primo Yúfera se observa en la ilustración siguiente.

Ilustración 103: Página de un estudio bibliométrico

The screenshot shows the Vestigium website interface. At the top, there is a logo for 'VESTIGIUM' and the text 'BIBLIOTECA DIGITAL DE CIENTÍFICOS Y HUMANISTAS VALENCIANOS'. Below this is a navigation menu with links: Inicio, Archivo de autores, Fondo documental, Quiénes somos, Colaboraciones, Enlaces, and Contact. A search bar is also present.

The main content area is titled 'Eduardo Primo Yúfera (Estudio bibliométrico)'. It includes the following sections:

- Autores por localización:** Castellón (2), Valencia (38), Alicante (6).
- Autores por actividad:** Científicos (23), Humanistas (23).
- Documentos:** Texto (3).
- Estudios Bibliométricos:** Científicos (1).

The 'Recursos:' section contains a table of attachments:

Adjunto	Tamaño
TABLAS-Finales.pdf	128.22 KB
FIGURA_1.jpg	88.18 KB
FIGURA_2.jpg	65.08 KB
FIGURA_3.JPG	142.43 KB
FIGURA_4.JPG	134.23 KB
FIGURA_5.JPG	107.39 KB

Below the table, there is an 'Artículo:' section with the text: 'Indicadores de actividad científica en investigadores singulares: perfil bibliométrico de Eduardo Primo Yúfera, expresidente del CSIC'.

The footer contains copyright information: ©2011-2014 Vestigium, contacto@bibliotecavestigium.es, and a link to the 'Política de privacidad'. It also features logos for the Universitat de València, Universitat de València, and Universidad Politécnica de Valencia, along with the CSIC logo.

Fuente: Biblioteca Vestigium

Hasta aquí se ha desarrollado el sitio web y se han completado las páginas con información. En el desarrollo interno se han seleccionado los módulos necesarios para publicar los contenidos con lenguaje RDF y asignar URI para cumplir los tres primeros principios de buenas prácticas de LD.

Para cumplir también con el cuarto y último principio LD, hay que incluir enlaces a otras URI para conectar los datos y que no queden aislados, de modo que se puedan compartir con otras

fuentes externas. En este sentido la información se amplía con contenidos externos que van a aportar un contexto y un valor añadido a nuestro sitio web.

2. Enlazar LD de otras fuentes (linking datasets)

Los conjuntos de datos suelen estar distribuidos en diversas fuentes. Para enriquecer la información de Vestigium con información adicional de los autores, hay que buscar y seleccionar los conjuntos que sean adecuados y estén relacionados con los contenidos propios, y después enlazar con los nuevos datos descubiertos, como se vio en el capítulo anterior con las fuentes LOD.

Como vimos en el capítulo anterior, el método de búsqueda y descubrimiento de los datos se puede realizar a través de la consulta a tres tipos de recursos:

1. Interfaces de consultas a las fuentes LOD.
2. SPARQL Endpoints.
3. Buscadores web semánticos.

1. A través de las interfaces de consultas a catálogos LOD.

Tal y como se vio en el capítulo anterior, del listado propuesto de recursos LOD del Data Hub y de la nube de LOD se seleccionan las siguientes fuentes, que se han interrogado con consultas sobre algunos de los autores que incluye Vestigium:

- En la fuente VIAF, como se vio en el segundo capítulo, se han realizado búsquedas de cada uno de los autores registrados en Vestigium. La ficha del autor en VIAF tiene una URI estable que lo identifica de forma unívoca, y mediante el campo owl:sameAs se ha enlazado con la ficha del autor en Vestigium. En la siguiente tabla se muestran los dos identificadores del mismo recurso “autor” en dos fuentes distintas.

Tabla 23: URI del recurso autor en Vestigium y URI del mismo recurso en VIAF

<http://www.vestigium.es/biblioteca/barciagoyanes>
<http://viaf.org/viaf/55499576/>

Fuente: Vestigium

Los recursos enlazan a través de las URI en el campo owl:SameAs de la ontología OWL, que identifica un mismo objeto digital en fuentes diversas. Tal y como se vio en el apartado “Tipos de contenido”, para el contenido de tipo AUTOR el campo de descripción visible para los usuarios es

“Mismo recurso en otras fuentes”.

2. A través de interfaces de consulta SPARQL Endpoints

Utilizando lenguaje de consulta SPARQL Protocol and RDF Query Language. En Vestigium se han activado los módulos SPARQL API, Endpoint y Endpoint Registry, como se ha visto en el apartado Módulos de este mismo capítulo. Esto permite ejecutar consultas dinámicas a almacenes remotos de RDF y la incorporación a nuestro sitio web de datos externos. Para el caso de Vestigium la implementación corre a cargo de un informático, si bien la selección de las fuentes se realizó durante nuestro proceso de investigación: DBpedia, BNE y Síndice. Y a continuación se consultan estas fuentes:

- DBpedia se puede consultar mediante DBpedia SPARQL Endpoint también mediante el software Silk Link Discovery Framework. Realizamos la consulta en el SPARQL Endopint para personas que hayan nacido en Alicante y el resultado, como se observa en la ilustración siguiente, muestra los autores alicantinos de los que existen contenidos en DBpedia. Esta consulta se realiza para Castellón y Valencia.

Ilustración 104: Interfaz de consulta SPARQL de DBpedia con la búsqueda de autores nacidos en Alicante

SPARQL Explorer for <http://dbpedia.org/sparql>

```

SPARQL:
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT ?name ?birth ?death ?person WHERE {
  ?person dbo:birthPlace :Alicante .
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  ?person dbo:deathDate ?death .
  FILTER (?birth < "1900-01-01"^^xsd:date) .
}
ORDER BY ?name

```

Results:

SPARQL results:

name	birth	death	person
"Carlos Arniches"@en	"1866-10-10+02:00"^^xsd:date	"1943-04-15+02:00"^^xsd:date	:Carlos_Arniches ↗
"Eleuterio Maisonnave Y Cutayar"@en	"1840-09-05+02:00"^^xsd:date	"1890-05-04+02:00"^^xsd:date	:Eleuterio_Maisonnave_y_Cutayar ↗
"Francisco Bouligny"@en	"1736-09-03+02:00"^^xsd:date	"1800-11-24+02:00"^^xsd:date	:Francisco_Bouligny ↗
"Francisco Domingo Joseph Bouligny"@en	"1736-09-03+02:00"^^xsd:date	"1800-11-24+02:00"^^xsd:date	:Francisco_Bouligny ↗
"Óscar Esplá"@en	"1886-08-04+02:00"^^xsd:date	"1976-01-05+02:00"^^xsd:date	:%C3%93scar_Espl%C3%A1 ↗
"Óscar Esplá y Triay"@en	"1886-08-04+02:00"^^xsd:date	"1976-01-05+02:00"^^xsd:date	:%C3%93scar_Espl%C3%A1 ↗

Powered by [OpenLink Virtuoso](#) and [GraphDB](#)

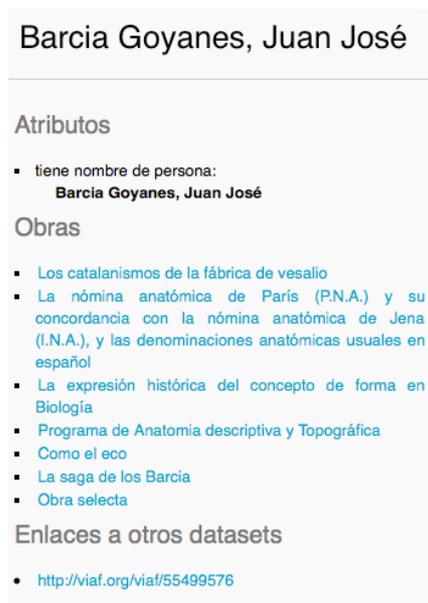
Fuente: DBpedia SPARQL <<http://dbpedia.org/sparql>>

- En fuentes como la BNE, la búsqueda de contenidos semánticos ha mejorado notablemente para los usuarios finales. En los inicios del proyecto se consultaba a través de un SPARQL Endpoint <datos.bne.es/sparql> y con una interfaz de consulta <bne.linkeddata.es/>, que en 2014 ha sido sustituida por la nueva versión <datos.bne.es>, una interfaz visual más amigable para los usuarios, que explota las relaciones de los datos de fuentes externas y presenta la información. Las consultas pueden realizarse empleando lenguaje natural, entorno a

tres entidades: autores, obras y temas.

La siguiente ilustración es el resultado de una búsqueda en la versión de 2011 <bne.linkeddata.es> sobre un autor que se recoge en Vestigium. Recupera la información que se muestra en la ilustración que sigue: la ficha del autor, los atributos que identifican que es una persona, las obras del autor y las URI de enlace con otros datasets en otras fuentes de datos.

Ilustración 105: Resultado de búsqueda de un autor en bne.linkeddata.es



Barcia Goyanes, Juan José

Atributos

- tiene nombre de persona:
Barcia Goyanes, Juan José

Obras

- [Los catalanismos de la fábrica de vesalio](#)
- [La nómina anatómica de París \(P.N.A.\) y su concordancia con la nómina anatómica de Jena \(I.N.A.\), y las denominaciones anatómicas usuales en español](#)
- [La expresión histórica del concepto de forma en Biología](#)
- [Programa de Anatomía descriptiva y Topográfica](#)
- [Como el eco](#)
- [La saga de los Barcia](#)
- [Obra selecta](#)

Enlaces a otros datasets

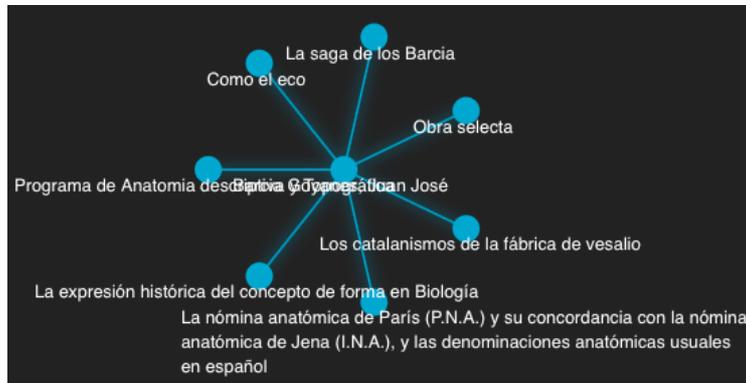
- <http://viaf.org/viaf/55499576>

Fuente: Bne.linkeddata.es

También se mostraba la información en forma de grafo, en el que los nodos son los valores de los datos y los arcos las

relaciones, como presenta la ilustración siguiente.

Ilustración 106: Gráfico con los resultado de búsqueda de un autor en bne.linkeddata.es



Fuente: Bne.linkeddata.es

La siguiente imagen es el resultado de una búsqueda en la nueva versión de 2014 <datos.bne.es> sobre el mismo autor.

Ilustración 107: Resultado de búsqueda de un autor en datos.bne.es

The screenshot shows the 'datos.bne.es' website interface. At the top, there is a navigation menu with 'INICIO', 'AUTORES', 'OBRAS', 'TEMAS', and 'Ayuda'. A search bar contains the text 'Buscar un autor, obra o tema'. Below the search bar, the author's name 'Barcia Goyanes, Juan José' is displayed. Underneath, three circular statistics are shown: 'Autor de 15 Obras', 'Tema en 1 Obra', and 'Participa en 10 Obras'. A section titled 'Sus obras' lists six individual works, each with a red 'Obra' button. The works listed are: 'Como el eco (1 versión)', 'La expresión histórica del concepto de forma en Biología (1 versión)', 'La nomenclatura anatómica de París (P.N.A.) y su concordancia con la nomenclatura anatómica de Jena (I.N.A.) y las denominaciones anatómicas usuales en español (1 versión)', 'La saga de los Barcia (1 versión)', 'Los catalanismos de la fábrica de vesalio (1 versión)', and 'Obra selecta (1 versión)'.

Fuente: datos.bne.es

La información que proporciona como resultado está integrada en una única interfaz, aunque los contenidos provienen de fuentes externas. En Vestigium, mediante un nuevo campo owl:sameAs en la ficha de autor se enlaza con la URL del autor en la BNE. Como muestra la siguiente tabla enlaza información sobre un mismo autor en varias fuentes.

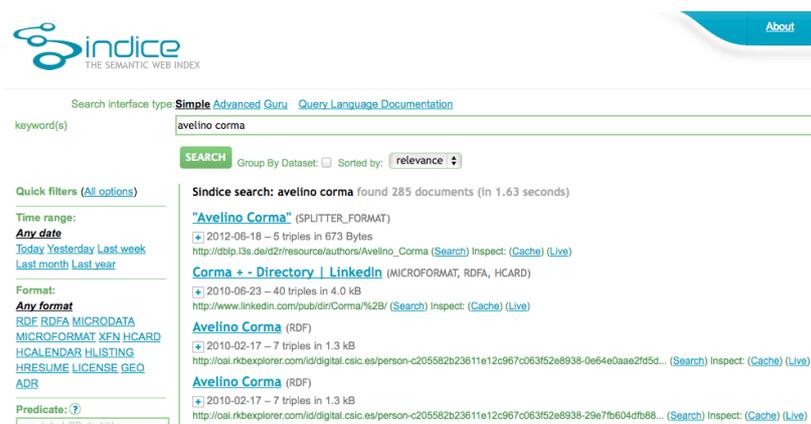
Tabla 24: URI del recurso autor en Vestigium y URI del mismo recurso en VIAF y en BNE

Barcia Goyanes, Juan José < http://www.vestigium.es/biblioteca/barciagoyanes >
owl:sameAs < http://viaf.org/viaf/55499576 >
owl:sameAs < http://datos.bne.es/resource/XX873273 >

Fuente: Vestigium

- Síndice: realizamos una búsqueda de un autor de Vestigium en Síndice, por ejemplo, “Avelino Corma”. Los resultados que recupera son numerosos datasets formados por tripletas en lenguaje RDF y procedentes de fuentes diversas, como se puede observar en la imagen siguiente.

Ilustración 108: Interfaz de búsqueda de Síndice con los resultados de buscar el autor “Avelino Corma”



Fuente: Síndice <<http://sindice.com>>

La búsqueda para el autor “Sorolla” recupera diferentes fuentes que exponen conjuntos de tripletas con información sobre el autor. Una de ellas es DBpedia con 54 tripletas RDF, cuyo enlace es http://dbpedia.org/page/Joaqu%C3%ADn_Sorolla que se conecta en Vestigium mediante la etiqueta owl:sameAs, que corresponde al campo “mismo recurso” visible en la página del autor en Vestigium. En Dbpedia, la ficha de descripción de Sorolla proporciona numerosos enlaces al mismo recurso en otras fuentes y al mismo recurso en otros idiomas, como recoge la siguiente ilustración.

Ilustración 109: URI del mismo recurso en DBpedia en diferentes idiomas

owl:sameAs	<ul style="list-style-type: none">▪ http://de.dbpedia.org/resource/Joaquín_Sorolla▪ http://fr.dbpedia.org/resource/Joaquín_Sorolla_y_Bastida▪ http://it.dbpedia.org/resource/Joaquín_Sorolla▪ http://pt.dbpedia.org/resource/Joaquín_Sorolla▪ http://ru.dbpedia.org/resource/Соролья,_Хоакин▪ http://pl.dbpedia.org/resource/Joaquín_Sorolla_y_Bastida▪ http://es.dbpedia.org/resource/Joaquín_Sorolla▪ freebase:Joaquín_Sorolla▪ http://nl.dbpedia.org/resource/Joaquín_Sorolla_y_Bastida▪ http://eu.dbpedia.org/resource/Joaquín_Sorolla▪ http://www.wikidata.org/entity/Q351746▪ http://yago-knowledge.org/resource/Joaquín_Sorolla
------------	--

Fuente: DBpedia http://dbpedia.org/resource/Joaqu%C3%ADn_Sorolla

El enlace también proporciona acceso a otros recursos de información sobre el autor en fuentes externas, mediante el campo dbpedia-owl:wikiPageExternalLink como muestra la ilustración que sigue.

Ilustración 110: URI de información relacionada con el recurso en fuentes externas

dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none">▪ http://museosorolla.mcu.es/colecc_pintura.html▪ http://obrasocial.bancaja.es/cultura/coleccion/series.aspx?id=7/▪ http://www.joaquin-sorolla-y-bastida.org/the-complete-works-1-150-3-2.html▪ http://www.youtube.com/watch?v=pJmDD3lwpNM▪ http://www.hispanicsociety.org/▪ http://www.museodelprado.es/en/exhibitions/exhibitions/at-the-museum/joaquin-sorolla-1863-1923/▪ http://www.sorollapaintings.com▪ http://www.artrenewal.org/pages/artist.php?artistid=1393
----------------------------------	---

Fuente: Dbpedia <http://dbpedia.org/resource/Joaqu%C3%ADn_Sorolla>

De este modo, desde una búsqueda en Síndice, se ha navegado a otro recurso que es DBpedia, que proporciona numerosos enlaces a otras fuentes externas. Enlazando desde Vestigium a estos recursos, estamos conectándonos a una red de enlaces de información relacionada con el autor, lo que proporciona un contexto informativo y un valor añadido a la información inicial que existe en Vestigium.

3. Consultas a los buscadores web semánticos generales.

- En WolframAlfa, se realiza la siguiente búsqueda: “Sorolla” y los resultados que recupera son numerosos datos sobre el autor procedentes de diversas fuentes y que se integran en una misma interfaz. A continuación la imagen siguiente con los resultados de la búsqueda muestra algunos de los bloques de información agregados.

Ilustración 111: Interfaz de búsqueda de WolframAlpha con los resultados de la búsqueda de Sorolla

The image shows the WolframAlpha search interface. At the top, the search bar contains the text 'sorolla'. Below the search bar, the results are displayed under the heading 'Joaquín Sorolla (painter)'. A table titled 'Basic information:' provides the following data:

field	value
full name	Joaquín Sorolla y Bastida
date of birth	Friday, February 27, 1863 (151 years ago)
place of birth	Valencia, Valencia
date of death	Friday, August 10, 1923 (age: 60 years) (90 years ago)
place of death	Madrid, Madrid

Below the table is a 'Timeline:' section showing a horizontal axis from 1860 to 2000, with a blue bar indicating the lifespan of Joaquín Sorolla from 1863 to 1923. Underneath, 'Notable artworks:' are listed: 'Self Portrait (1909)', 'Children on the Seashore (1903)', 'And They Still Say Fish are Expensive! (1894)', and 'Academic Life Study (1887)'. At the bottom, there is a 'Wikipedia page hits history:' section and a 'Log scale' button.

Fuente: WolframAlpha <<http://www.wolframalpha.com/input/?i=sorolla>>

Los diferentes campos de datos detallan la fuente externa de la que se obtiene la información. De este modo permite navegar y descubrir nuevos contenidos relacionados.

Así pues, es posible realizar búsquedas en los diferentes recursos semánticos y recuperar conjuntos de datos estructurados en RDF, que se conectan mediante la URI estable. En este sentido, desde la página de un autor en Vestigium se podría navegar a información del mismo recurso en otra fuente

externa como DBpedia (campo owl:sameAs), y desde ésta a su vez, continuar la navegación a información relacionada con el recurso en otras fuentes externas (mediante los campos owl:ExternalLink y rdfs:seeAlso).

Evaluación de Vestigium

Se han seguido los cuatro principios de buenas prácticas de LD en el desarrollo del sitio web Vestigium para que al publicar contenido, éste sea descrito en RDF y contenga URI estables de identificación que puedan ser descubiertas y enlazadas por otras fuentes. El resultado del sitio web se visualiza para los usuarios en el navegador en la dirección www.vestigium.es. La apariencia de una página de autor es la que muestra la siguiente ilustración.

Ilustración 112: Página sobre un autor en Vestigium



VESTIGIUM

BIBLIOTECA DIGITAL DE CIENTÍFICOS Y HUMANISTAS VALENCIANOS

Inicio | Sobre nosotros | Contacto | Ayuda

Buscar

Autores por localización

- Castellón (1)
- Valencia (23)
- Alicante (5)

Autores por actividad

- Científicos (9)
- Humanistas (20)

Documentos

- Vídeo (1)
- Texto (2)
- Otros formatos (1)

Estudios Bibliométricos

- Humanistas (1)
- Científicos (2)

Barcia Goyanes, Juan José



Biografía/Historia:
Juan José Barcia Goyanes (Santiago de Compostela, 26 de diciembre de 1901 - Valencia, España, 13 de julio de 2003) fue un médico anatomista, pionero de la Neurología, de la Psiquiatría y de la Neurocirugía en España, Decano y Rector en Valencia, escritor y humanista español. Graduado de la Universidad de Valencia, es miembro de la quinta generación de médicos, que se han seguido prolongando en el tiempo, la mayoría de ellos profesores de Universidad y dedicados al cultivo de las ciencias neurológicas. En dos ocasiones acompaña a su padre, que era médico militar al norte de África, donde inicia su interés por la cultura árabe. Aquel entonces políglota, hablaba español, gallego, valenciano, portugués, alemán, francés, inglés, italiano, sueco, danés, ruso, griego moderno, árabe y polaco y conocía el griego clásico, el latín, hebreo, sánscrito y persa. Habiendo publicado en la mayoría de estos idiomas. Barcia estudia el Bachillerato y la carrera de Medicina en Santiago de Compostela, España. Inicia la carrera de medicina en enero de 1918, cuando contaba 16 años, al lado de su abuelo Juan Barcia Caballero y desde el primer año fue alumno interno de Anatomía. Terminó la licenciatura en 1922, ya que había adelantado un año, con excelentes calificaciones recibiendo el «premio Rodríguez Abaytua» de la Real Academia de Medicina al mejor expediente de su país. En 1925 leyó su tesis doctoral titulada *Variaciones atávicas y las detenciones en el desarrollo, estudiadas en el maxilar superior de los locos*.

Actividad laboral:

Barcia Goyanes obtuvo por oposición la plaza de profesor auxiliar temporal de Anatomía siendo alumno y colaborador del profesor Rodríguez Cadalso, que siempre consideró su maestro en esa rama. Durante ese tiempo fue médico del sanatorio de Angel Bañar Cortés, destacado cingano de Santiago de Compostela que lo inició en la Cirugía. En 1926, cuando todavía no había cumplido los 25 años obtuvo por oposición la cátedra de Anatomía de la Universidad de Salamanca. En 1929 se trasladó a Valencia para ocupar la cátedra de Anatomía. En ese año obtuvo por oposición la plaza de Jefe de Servicio de Enfermedades Nerviosas del hospital de Valencia. El Hospital General de Valencia era de hecho el hospital Clínico. La Facultad de Medicina estaba adosada a él y los distintos servicios eran dirigidos por los catedráticos de las diferentes asignaturas, con excepción del servicio de Medicina y Cirugía que estaban duplicados y el mencionado de Neuropsiquiatría. Esta es la razón, por la que Barcia Goyanes enseñó Neurología hasta la inauguración del hospital Clínico en las instalaciones actuales en 1966. En 1930 transformó el servicio de Enfermedades Nerviosas en Servicio de Neuropsiquiatría y Neurocirugía, por lo tanto el primero de España. El 22 de noviembre de 1931, Barcia Goyanes realizó la primera operación neuroquirúrgica en España de que se tenga referencia, abriendo el duramáter, siguiendo las técnicas de Cushing y Dandy. El estallido de la guerra Civil le sorprende en Santander, en cuya Universidad de Verano había iniciado un curso junto a su entonces alumno Lain Entralgo sobre medicina Psicosomática. Al terminar la Guerra Civil Española, vuelve a Valencia y allí continúa su actividad hasta su muerte. Al crearse el Consejo Superior de Investigaciones científicas, es nombrado Jefe de la Sección de Neurología del Instituto Cajal. Fundó la *Revista Médica Española*,¹ la *Revista Española de Neurofisiología y Neurocirugía*. Fundó con otros la *Sociedad Luso-Española de Neurocirugía*. Fue además presidente en dos legislaturas de la *Sociedad Española de Neurología* y de la *Sociedad Española de Anatomía* y miembro de honor de varias sociedades europeas y americanas. Dirigió numerosas tesis doctorales. Fue decano de la Facultad de Medicina de Valencia, España, desde 1945 hasta 1964 y durante su gestión se produjo el cambio decisivo en la Facultad. Logró la creación de la nueva Facultad de Medicina. Inauguró el nuevo hospital Clínico. Rescató la biblioteca Histórica, que había fundado su predecesor en el cargo el Rector Nicolás Ferrer Julié y, con la colaboración del profesor López Piñero, creó la nueva biblioteca y Hemeroteca. Barcia Goyanes fue rector de la Universidad Literaria de Valencia en 1965 hasta su jubilación en 1971 y durante su rectorado se produjo una gran expansión de la Universidad. Organizó numerosos Congresos y Reuniones, de los que cabe destacar el Primer Congreso de Historia del País Valenciano, tan controvertido en la época que se organizó. Estuvo en posesión de numerosas condecoraciones y distinciones. Y por último, fue un excelente poeta y deportista.

Ocupación:

Ex-decano y ex-Rector Universidad de Valencia

Lugar dónde desarrolla su actividad:

Valencia

Campo de actividad:

Humanistas

Fecha de nacimiento:

26 de diciembre de 1901

Lugar de nacimiento:

Santiago de Compostela, España

Fecha de defunción:

13 de julio de 2003

Lugar de defunción:

Valencia, España

Documentos:

El mito de Vesalio

Estudios bibliométricos:

Juan José Barcia Goyanes (Estudio bibliométrico)

Mismo recurso (autor, documento, etc.) en otros conjuntos de datos:

<http://datos.bne.es/resource/O873273>
<http://viaf.org/viaf/55495676>
<http://hdl.handle.net/10261/3924>

©2011-2014 Vestigium
contacto@bibliotecavestigium.es
Política de privacidad



Fuente: Vestigium

En cuanto a los resultados para que los clientes, agentes web semánticos y las aplicaciones puedan recuperar los documentos en RDF y puedan enlazar con las URL de los recursos de información disponibles en Vestigium, se realiza una prueba mediante la herramienta Síndice. Esta herramienta, además de indizar documentos que utilicen cualquier formato semántico y localizar documentos que tengan sentencias RDF sobre un determinado recurso (Pastor-Sánchez et al., 2013), permite realizar un análisis sobre cualquier documento web para identificar la estructura interna de descripción del recurso y mostrar los elementos empleados.

Al hacer el análisis sobre la URI de la página de un autor de Vestigium los resultados muestran que:

1. Vestigium se ha desarrollado con Drupal y crea contenidos estructurados con lenguaje RDF.
2. Al crear un nuevo contenido, asigna una URI unívoca a cada objeto digital.
3. Identifica qué tipo de recurso es cada objeto digital que tiene asignada una URI. Como se puede ver en la imagen que sigue, la URI analizada es de tipo persona (identificado con el vocabulario Person de schema.org).

4. Identifica el mismo objeto digital en diferentes fuentes externas a Vestigium. Por ejemplo, el objeto digital analizado en Síndice es la URI <<http://www.vestigium.es/biblioteca/barciagoyanes>> y los otros identificadores URI del mismo recurso en distintas fuentes son en la BNE <<datos.bne.es/resource/XX873273>>, en VIAF <<viaf.org/viaf/55499576>> y en digital CSIC <<hdl.handle.net/10261/3924>>, como muestra la siguiente imagen.

Ilustración 113: Resultado de Síndice sobre el análisis de la URL de un contenido de autor creado en Vestigium



Fuente: Síndice

5. Emplea diferentes vocabularios de metadatos para la descripción estructurada de los recursos. La siguiente imagen es el resultado del análisis de Síndice sobre la URI de un autor, que muestra las tripletas RDF, compuestas por el sujeto, que es una URI, el predicado, que son los elementos de los vocabularios de metadatos y de valor que describen las

propiedades del recurso autor, y el objeto de la tripleta, que es el valor de cada característica o propiedad. Por ejemplo, en la imagen se puede leer que para el recurso autor, cuya URI es <http://www.vestigium.es/biblioteca/barciagoyanes> (sujeto), la propiedad afiliación (predicado RDF del vocabulario sorq:affiliation) tiene el valor “exdecano y exrector de la Universidad de Valencia” (objeto).

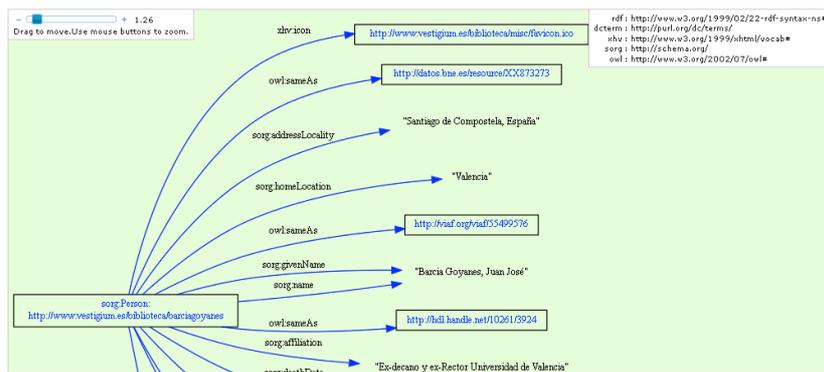
Ilustración 114: Resultado de Síndice, predicados RDF empleados en la ficha de autor de Vestigium

subject	predicate	object
<http://www.vestigium.es/biblioteca/barciagoyanes>	dcterms:title	"Barcia Goyanes, Juan José Vestigium. Biblioteca Digital de Humanistas y Científicos Valencianos"@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	xhv:icon	<http://www.vestigium.es/biblioteca/misc/favicon.ico>
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:name	"Barcia Goyanes, Juan José"@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:givenName	"Barcia Goyanes, Juan José"@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	rdf:type	soq:Person
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:image	""@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:image	""@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:description	"Juan José Barcia Goyanes (Santiago de Compostela, 26 de diciembre de 1901 - Valencia, España, 13 de julio de 2003) fue un médico anatomista, pionero de la Neurología, de la Psiquiatría y de la Neurocirugía en España, Decano y Rector en Valencia, escritor y humanista español. Graduado de la Universidad de Valencia, es miembro de la quinta generación de médicos, que se han seguido prolongando en el tiempo, la mayoría de ellos profesores de Universidad y dedicados al cultivo de las ciencias neurológicas. En dos ocasiones acompaña a su padre, que era médico militar al norte de África, donde inicia su interés por la cultura árabe. Auténtico políglota, hablaba español, gallego, valenciano, portugués, alemán, francés, inglés, italiano, sueco, danés, ruso, griego moderno, árabe y polaco y conocía el griego clásico, el latín, hebreo, sánscrito y persa. Habiendo publicado en la mayoría de estos idiomas. Barcia estudia el Bachillerato y la carrera de Medicina en Santiago de Compostela, España. Inicia la carrera de medicina en enero de 1918, cuando contaba 16 años, al lado de su abuelo Juan Barcia Caballero y desde el primer año fue alumno interno de Anatomía. Terminó la licenciatura en 1922, ya que había adelantado un año, con excelentes calificaciones recibiendo el «premio Rodríguez Abaytúa» de la Real Academia de Medicina al mejor expediente de su país. En 1925 leyó su tesis doctoral titulada Variedades atávicas y las detenciones en el desarrollo, estudiadas en el maxilar superior de los locos."@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:affiliation	"Ex-decano y ex-Rector Universidad de Valencia"@es
<http://www.vestigium.es/biblioteca/barciagoyanes>	soq:homeLocation	"Valencia"@es

Fuente: Síndice

Otro producto que genera Síndice como resultado del análisis que realiza de la URI es un grafo. Este grafo representa las tripletas RDF y los elementos que la componen. La siguiente imagen muestra el grafo del recurso analizado, en el que el nodo principal es la URI del objeto digital y enlaza mediante predicados RDF a cada uno de los atributos o características, que son el valor que adquiere el predicado RDF. Estos valores pueden ser términos, por ejemplo, “Valencia” para el predicado RDF *sorg:homeLocation*, o pueden ser una o más URI a información relacionada con el recurso en otras fuentes y sitios web externos, para el predicado RDF *owl:sameAs*.

Ilustración 115: Grafo generado por Síndice sobre el recurso autor Barcia Goyanes en Vestigium



Fuente: Síndice

CONCLUSIONES

Las conclusiones siguientes son una recapitulación del estudio realizado.

Uno. Una biblioteca digital con contenidos de acceso abierto ya no es suficiente en estos momentos. La red se encuentra saturada de todo tipo de contenidos y la atención del usuario cada vez es más difícil de captar. Son necesarias tecnologías semánticas que proporcionen un contexto informativo a la información que se ofrece. Mediante enlaces persistentes a otras fuentes que también forman parte de la red de datos, se conecta directamente con datos relacionados y se amplía la información sobre el recurso. Esta información de contexto habitualmente no suele ser proporcionada por las bibliotecas. Las tecnologías semánticas hacen posible enlazar las colecciones bibliográficas con datos geográficos y de autoridades, creando una red de navegación más allá del recurso originado en la biblioteca.

Dos. Los gestores de contenido facilitan la producción de objetos digitales de información propios (como las obras o las biografías de los autores) publicándolos como datos enlazados. El caso particular del gestor Drupal versión 7, constituye una herramienta eficaz para publicar contenidos cumpliendo con los cuatro principios de buenas prácticas de datos enlazados, participando así de la web semántica. Este gestor está compuesto por elementos básicos que exponen los datos en

RDF. Sin embargo, continúa siendo necesario instalar otros complementos adicionales (como el módulo RDF y el módulo de consulta SPARQL) para avanzar en la integración de otros contenidos ajenos en la propia biblioteca.

Tres. La publicación de contenidos con el cms Drupal 7 y los módulos RDF complementarios consigue identificar unívocamente los objetos digitales en la web. Lo que constituye uno de los principios en los que se basa la web semántica. Los objetos digitales que componen el sitio web Vestigium (páginas biográficas de investigadores y humanistas valencianos, enlazadas con las páginas de los documentos como artículos y tesis, de los propios autores o que los estudian, enlazados con estudios bibliométricos sobre las publicaciones de estos autores) están publicados con dicho cms y los módulos RDF. Además, tienen asignada una URI persistente de forma que quedan identificadas todas las variantes de un mismo objeto digital, independientemente del idioma o de la fuente en la que se encuentre.

Cuatro. La publicación de datos como LD los conecta con información de fuentes externas mediante enlaces persistentes. Muchas de las fuentes identificadas que publican conjuntos de datos como LOD son bibliotecas nacionales de numerosos países, principalmente anglosajones, organizaciones culturales

internacionales y centros de relevancia internacional. Por tanto, esta información de contexto cuenta con una calidad excelente, que enriquece los contenidos. Proporciona información con carácter semántico y un contexto informativo de enlaces por los que navegar hacia datos relacionados de otras fuentes. Ello supone un gran avance con respecto al mundo de los diccionarios biográficos que utilizan la web como medio de difusión, por no mencionar la revolución que supone con respecto a las tradicionales obras de referencia en papel.

Cinco. La reutilización y la combinación de vocabularios de metadatos y vocabularios de valores ya creados extiende ampliamente los campos de descripción de los objetos digitales. Se han reutilizado herramientas de control que están siendo usadas en la web, tal y como se promueve en el entorno semántico. Se han seleccionado los vocabularios de metadatos que mejor describen las características de cada uno de los tipos de contenidos que administra Vestigium. Se han reutilizado ontologías, así como también los vocabularios de valores para describirlos: schema.org, FOAF, DC, OWL, y SIOC.

Seis. Los beneficios en términos de acceso, interoperabilidad y reutilización están incidiendo en diversos colectivos profesionales que por ahora se refleja en conectar información. Para establecer estas conexiones sobre los objetos identificados hay que utilizar a escala mundial las mismas tecnologías para

describir, para comunicar y para enlazar. La normalización, por tanto, tiene un peso forzosamente ineludible para que todos describan los datos de la misma manera, utilicen vocabularios compatibles con todas las instituciones y los diversos contenidos hablen el mismo idioma para comunicarse entre sí. Describir de forma semántica los contenidos posibilita la interpretación por parte de las aplicaciones (índices semánticos, buscadores semánticos, aplicaciones de consulta RDF), lo que facilita la integración de los datos, mejora el posicionamiento SEO y la visibilidad y acceso al recurso. Por ahora esta tecnología, al igual que ocurrió con el acceso abierto a la literatura científica, se está consolidando. Se afianza su utilización gracias a la influencia de sus promotores (W3C), su eficacia y el hecho de que las instituciones culturales participen de esta iniciativa, de modo que la cantidad de datos que se encuentran disponibles en la web semántica aumenta inexorablemente día a día.

Siete. Las limitaciones de la web semántica. En el momento actual la web semántica se encuentra en un primer estadio en el que determinadas comunidades están publicando datos enlazados. Parece ser que por ahora algunas capacidades que se unian a la WS, como la inferencia de conocimiento, deben esperar a una mayor madurez tecnológica. Otras, se dirigen a proporcionar funcionalidades a los sistemas automáticos, aunque el usuario final no puede percibirlos. Para hacer visibles las

aportaciones de la WS a los usuarios finales es necesario el desarrollo de aplicaciones de visualización de contenidos semánticos. La WS precisa aún de desarrollos tecnológicos por parte de tecnólogos especialistas para generar aplicaciones que mejoren diversos aspectos como pueden ser nuevos servicios y productos semánticos que reutilicen los conjuntos de LD publicados y los hagan visibles.

Ocho. Las limitaciones de Vestigium. Las dificultades de la implementación de Vestigium fueron varias. En primer lugar establecer la equivalencia entre los campos de descripción de los recursos con los ítems de los vocabularios de metadatos semánticos. En segundo lugar, fue compleja la selección y puesta en marcha de los módulos, tanto para crear contenido LD, como para importarlo de fuentes externas. En este sentido es posible crear contenidos como LD y enlazarlos con datos de fuentes externas. Sin embargo, para importar contenidos, por ejemplo registros de autoridades de VIAF y registros bibliográficos de BNE, y visualizarlos desde la interfaz del propio sitio web Vestigium, es necesaria una aplicación desarrollada para tal efecto o la adaptación de un módulo apropiado. Como se ha visto, algunas instituciones pioneras en este aspecto, (como por ejemplo la BNE, la FAO (UN) o la Fundación Ignacio Larramendi) emplean un software propio en sus bibliotecas digitales, repositorios y catálogos. En este

sentido el proyecto Vestigium tiene que trabajar en la posibilidad que ofrece Drupal de utilizar un módulo disponible de importación de LD para visualizar datos externos en la interfaz propia, aunque es necesario un periodo de prueba y un proceso de adaptación para llegar a alcanzar estas funcionalidades. Por otro lado, la ausencia de enlaces entrantes refleja que se trata de un proyecto incipiente y por el momento con poca visibilidad.

Consideraciones finales

La integración de Vestigium en la web semántica, no ha sido una cuestión de pura implementación tecnológica. Cada uno de los nodos del gráfico mostrado sobre LOD lleva asociado un estudio previo de las opciones que pueden ser pertinentes para las características de los contenidos, además de una crítica toma de decisiones y la implementación técnica del proyecto. De esos tres procesos nuestra investigación abarca los dos primeros, si bien también se participó en la tercera fase. Esta tesis muestra que las tareas de un gestor de información actualmente suponen labores cercanas a lo que en TIC se denomina CEO o director ejecutivo para ciertos momentos del desarrollo de un producto de información. Las decisiones que se tomaron, por tanto, son la conclusión de las labores previas de investigación y atañen desde la elección del CMS hasta los vocabularios.

La red de datos enlazados constituye una iniciativa de cooperación internacional para crear conexiones entre recursos de información de diferentes fuentes por los que navegar. Por el momento en Vestigium se puede hablar de publicar contenidos RDF y conectar con otras fuentes y ofrecer así un contexto enriquecido. Por tanto, Vestigium forma parte de esta red de fuentes de información LOD aportando contenidos fiables y enlazables de la herencia cultural científica y humanista valenciana.

Trabajo Futuro

Este trabajo ofrece una aproximación al estado actual de la publicación de información bibliotecaria como LD. Los resultados pueden constituir un punto de partida para continuar con el trabajo a niveles mucho más amplios.

Hasta ahora, el trabajo desarrollado está más enfocado a conectar información y compartir recursos con otros productores de datos ya que nos encontramos en una fase aún inicial de la web de datos en la que los primeros trabajos se enfocan hacia técnicos, bibliotecarios e instituciones.

Un acercamiento de este tipo permitiría abordar aspectos que han quedado fuera de este trabajo, y que podrían encauzarse a través de cuatro posibles líneas de trabajo.

Una futura línea de trabajo sería la vinculada al desarrollo de una aplicación de importación de contenidos de registros de autoridades de fuentes como VIAF, DBpedia y Wikimedia, de las que se podría reutilizar información biográfica e imágenes de autoridades. También los registros bibliográficos de fuentes como la BNE e Hispana que permitirían reutilizar información bibliográfica y objetos bibliográficos digitalizados. Además de la visualización de estos contenidos reutilizados desde la propia interfaz del sitio web Vestigium.

Otra posible línea de trabajo se vincula a futuros desarrollos y mejoras en relación con la presentación de la información a los usuarios finales. En el caso Vestigium los siguientes pasos habrían de dirigirse hacia el desarrollo de interfaces gráficas orientadas, tanto de consulta, como de visualización de datos para los distintos tipos de usuarios. Actualmente son numerosos los tipos de visualizaciones que pueden realizarse del análisis de datos bibliográficos.

Otra futura línea de investigación es convertir la herramienta Vestigium en la base para un estudio de la historia de la ciencia en profundidad. Con la colaboración de especialistas del área deben abordarse los contenidos propios de los procesos

culturales locales con las aportaciones de los científicos y humanistas valencianos al avance de la ciencia. Esto supone una puesta en valor que podrá acometerse cuando la masa crítica de autores introducidos permita ser abordada mediante métodos de análisis de redes sociales, redes de colaboración científica o de las referencias externas.

Una última línea de trabajo, que iría muy ligada a lo anteriormente dicho sería la relacionada con la preservación de las humanidades digitales y las manifestaciones culturales en formato digital. Un campo muy diverso y que supone una oportunidad estratégica de ampliar terreno en el ámbito digital, gracias a las infraestructuras que suponen las bibliotecas digitales académicas y de investigación.

Cuando este proyecto de red de LOD esté definitivamente consolidado, cuente con muchas más instituciones de las que actualmente forman parte, y la mayoría de los actuales silos de información que hasta ahora se encuentran aislados en la red se conecten, estaremos en el camino de conseguir el *commons* o bien común del conocimiento.

BIBLIOGRAFÍA

- AENOR. *Información y documentación. Conjunto de elementos de metadatos Dublin Core. UNE-ISO 15836*. Madrid: AENOR, 2007
- AENOR. *Norma Internacional ISO 690-2*, Madrid: AENOR, 1997
- AENOR. “ISO 5127-6:1983. Documentación e información. Vocabulario Parte 6: Lenguajes Documentales”. *Rev. Esp. Doc. Cient.* 1997 vol. 20, n. 4, <<http://redc.revistas.csic.es/index.php/redc/article/view/576/651>>
- Agenjo Bullón, Xavier. Los archivos y Linked Open Data: normativa de Europea y del W3C. En: Los archivos diocesanos: nuevos retos ante la era virtual. Madrid, 11 y 12 de diciembre de 2012. <<http://www.slideshare.net/DIGIBIS/los-archivos-y-linked-open-data-normativa-de-europeana-y-del-w3c-de-xavier-agenjo>>
- Agenjo Bullón, Xavier; Hernández, Carrascal, Francisca. Informe final del grupo incubador de datos vinculados de bibliotecas, 2011 <http://www.larramendi.es/LAM/Incubator/lld/XGR-lld-20111025.html#Benefits_of_the_Linked_Data_Approach>
- Agenjo Bullón, Xavier; Hernández, Carrascal, Francisca. *La Biblioteca Virtual Larramendi: función y planteamiento*. 2010 <<http://www.digibis.com/en/images/PDF/biblioteca-virtual-larramendi-en-la-web-3-0.pdf>>

- Aguillo, Isidro. "Internet invisible o Infranet: definición, clasificación y evaluación". *VII Jornadas Españolas de Documentación*. Bilbao, Universidad del País Vasco, 2000

- Aguillo, Isidro. *Internet invisible: Los contenidos son la clave*. CINDOC-CSIC, 2003. <http://internetlab.cindoc.csic.es/cursos/Internet_Invisible2003.pdf>

- Alemu, Getaneh; Stevens, Brett; Ross, Penny. "Semantic Metadata Interoperability in Digital Libraries: A Constructivist Grounded Theory Approach" In: *ACM/IEEE Joint Conference on Digital Libraries, Ottawa (Canada), 13 June 2011* <<http://hdl.handle.net/10760/15829>>

- Alemu, Getaneh; Stevens, Brett; Ross, Penny. "Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries: a social constructivist approach" *New Library World*, 2012, Vol113, n. 1/2, pp. 38-54

- Alexander, Keith; Cyganiak, Richard; Hausenblas, Michael; Zhao, Jun. "Describing Linked Datasets with the VOID Vocabulary" 2011 <<http://www.w3.org/TR/void/>>

- Alexander, Keith; Cyganiak, Richard; Hausenblas, Michael; Zhao, Jun. *VoiD Guide - Using the Vocabulary of Interlinked Datasets* <<http://vocab.deri.ie/void/guide/2009-01-29>>

- Alonso, Julio; Subirats, Imma; Martínez Conde, M^a Luisa. *Informe APEI sobre acceso abierto*, 2008
<<http://www.apei.es/wp-content/uploads/2013/11/InformeAPEI-Accesoabierto.pdf>>

- Álvarez, Martín. El W3C y la web semántica, En: *VII Feria de Tiendas Virtuales*, Huesca 17 de Abril de 2012
<<http://www.w3c.es/Presentaciones/2012/0417-WebSemantica-MA/0417-WebSemantica-MA.pdf>>

- American Library Asociation, *History*, 2014
<<http://www.ala.org/aboutala/history>>

- Andrade, Jesús Alberto. “Crisis de las revistas impresas”. En: *Opcion*, vol. 25 n. 59 Maracaibo ago. 2009

- Anglada, Lluís. “Políticas de acceso abierto en los países del sur de Europa”. *Blok de bid*, nov. 2013
<<http://www.ub.edu/blokdebid/es/content/politicas-de-acceso-abierto-en-los-paises-del-sur-de-europa>>

- Anibaldi, Stefano and Jaques, Yves and Celli, Fabrizio and Stellato, Armando and Keizer, Johannes *Migrating bibliographic datasets to the Semantic Web: The AGRIS case. Semantic Web.*, 2013. In EFITA, Torino (Italy), 23-27 June 2013.

- Arriola Navarrete, Oscar. “Open access y software libre: un área de oportunidad para las bibliotecas”. *Biblioteca universitaria*, 2011, vol. 14, n. 1, pp. 26-40.
<<http://eprints.rclis.org/16172/>>

- Aupí, Vicente; Brines, Rafael. *Eduardo Primo Yúfera. La investigación al servicio de la humanidad*. Valencia: Universidad Politécnica de Valencia, 1994
- Ayala-Gascón, María. *Eduardo Primo Yúfera: Un estudio bio-bibliométrico*. Tesis Doctoral. Valencia: Universidad Católica de Valencia San Vicente Martir, 2013
- Baker, Thomas; et al. Principles of Metadata Registries: A White Paper of DELOS Registries Working Group. Retrieved July 21, 2003, <<http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf>>
- Baker, Thomas; et al. *Library Linked Data Incubator Group Final Report*, 2011 <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>
- Barrueco, José Manuel. “Experiencias cooperativas en archivos abiertos en Ciencias Sociales”. En: *IV Semana de la ciencia*, Madrid, 2004
- Barrueco, José Manuel; Subirats Coll, Imma. “Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo”. *El profesional de la información*, 2003, vol. 12, n 2, pp. 99-106. <<http://eprints.rclis.org/4093/>>
- Barrueco, José Manuel; García Testal, Cristina. “Aplicación de LOD para enriquecer la colección patrimonial de la Universitat de València Somni”. En:

XIII Jornadas Españolas de Documentación: Fesabid 2013. Toledo 24 y 25 de mayo, 2013
<<http://www.slideshare.net/fesabid/aplicacion-de-lod-para-enriquecer-la-coleccion-patrimonial-de-la-universitat-de-valencia-somni-jos-manuel-barrueco-cruz-cristina-garca-testal>>

- Barrueco Cruz, José Manuel; Caballos Villar, Almudena; Campos Rodríguez, Ángeles; Casaldàliga, Núria; Combarro Felpeto, Pilar; Cívico Martín, Rafaela; Domènech Luisa; García Gil, Ma Angeles; Losada, Marina; Morillo Moreno, José Carlos. *Guía para la evaluación de repositorios institucionales de investigación*, Madrid: FECYT, RECOLECTA y CRUE 2014
<http://recolecta.fecyt.es/sites/default/files/contenido/documentos/GuiaEvaluacionRecolecta_v.ok_0.pdf>
- Bechhofer, Sean; Harmelen, Frank van; Hendler, Jim; Horrocks, Ian; McGuinness, Deborah L.; Patel-Schneider, Peter F.; Andrea Stein, Lynn. OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004 <<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>>
- Bethesda Statement on Open Access Publishing, 2003.
<<http://www.earlham.edu/~peters/fos/bethesda.htm>>
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003.
<http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf>
- Berners-Lee, Tim. *Linked Data - Design Issues*, 2006,
<<http://www.w3.org/DesignIssues/LinkedData.html>>

- Berners-Lee, Tim. *On the next web*. Ted.com, 2009. <http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html>
- Berners-Lee, Tim. *Tejiendo la Red: el inventor del World Wide Web nos descubre su origen*. Madrid: siglo XXI de España editores, 2000.
- Bibliographic Ontology. *Bibliographic Ontology Specification*, 2009 <<http://biblontology.com/>>
- Biblioteca Nacional Española. Historia de la catalogación, 2014 <<http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/NormasNacionales/ReglasDeCatalogacion/HistoriaDeLaCatalogacion/>>
- Biblioteca Nacional Española. Historia de la Biblioteca Nacional de España, 2014 <http://www.bne.es/webdocs/LaBNE/docs/historia_BNE.pdf>
- Biblioteca Digital Valenciana. Presentación, <<http://bv2.gva.es/es/cms/elemento.cmd?id=estaticos/paginas/presentacion.html>>
- Biomed Central <<http://www.biomedcentral.com/openaccess/repositories>>
- Bizer, Christian; Heath, Tom; Berners-Lee, Tim “Linked Data-The Story So Far,” *International Journal on Semantic Web and Information Systems*, 2009, vol. 5 n. 3, pp. 1-22 <<http://tomheath.com/papers/bizer-heath->

berners-lee-ijswis-linked-data.pdf>

- Blasco Gil, Yolanda. *La facultad de derecho de Valencia durante la restauración (1875-1900)*, Valencia: Universitat de València, 2000
- Borillo Domenech, Ricardo; Gumbau Mezquita, José Pascual. “Arquitecturas orientadas a servicios para la preservación de objetos digitales”. En: *El documento electrónico: aspectos jurídicos, tecnológicos y archivísticos*, Alicante: Servicio de publicaciones de la Universitat Jaume I, 2008. pp. 215-236
- Brickley, Dan; Miller, Libby. *FOAF Vocabulary Specification*, 2014 <<http://xmlns.com/foaf/spec/>>
- Brickley, Dan; Guhna, R.V. *RDF Vocabulary Description Language 1.0: RDF Schema*, 2003 <<http://www.w3.org/TR/2003/PR-rdf-schema-20031215/>>
- Brickley, Dan; Guha, R.V. *RDF Schema specification*, 2014. <<http://www.w3.org/TR/PR-rdf-schema>>
- Byrne, Gillian; Goddard, Lisa. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine*, 2010, <<http://www.dlib.org/dlib/november10/byrne/11byrne.print.html>>
- Bush, Vannevar. “As We May Think”. *The Atlantic Monthly*, July 1945 (Traducción al castellano en Vannevar Bush. “Cómo podríamos pensar”. *Revista de Occidente*, marzo 2001 N° 239)

<<http://sindominio.net/biblioweb/pensamiento/vbush-es.pdf>>.

- FAO. AIMS. AGROVOC
<<http://aims.fao.org/es/standards/agrovoc/functionalities/search>>
- Caracciolo, Caterina; Stellato, Armando; Morshed, Ahsan; Johannsen, Gudrun; Rajbhandari, Sachit; Jaques, Yves; Keizer, Johannes. "The AGROVOC Linked Dataset". *Semantic Web*, 2013, vol. 4, n. 3, pp. 341-348.
<<http://eprints.rclis.org/20648/>>
- Caracciolo, Caterina and Morshed, Ahsan and Stellato, Armando and Johannsen, Gudrun and Jaques, Yves and Keizer, Johannes "Thesaurus Maintenance, Alignment and Publication as Linked Data: The AGROVOC Use Case", 2012 . *In 6th Metadata and Semantics Research Conference (MTSR 2012)*, Cádiz, Spain, 28-30 November 2012 <<http://eprints.rclis.org/17734/>>
- Caro Castro, Carmen. "Vocabularios estructurados, Web Semántica y Linked Data: oportunidades y retos para los profesionales de la documentación" En: *Arquivologia, Biblioteconomia e Ciência de Informação: Identidades, Contrastes e Perspectivas de Interlocução*. Eduardo Ismael Murgia, Mara Eliane Fonseca Rodrigues (orgs.). Niterói: Editora da UFF, 2012, p. 139-155.
- Castells, Pablo. *La web semántica*. Madrid: Universidad Autónoma de Madrid, 2003
<<http://arantxa.ii.uam.es/~castells/publications/castells-uclm03.pdf>>

- Chan, Lois Mai; Zeng, Marcia Lei. "Metadata interoperability and standardization – a study of methodology part I: achieving interoperability at the schema level" *D-Lib Magazine*, 2006, vol. 12 n. 6 <<http://www.dlib.org/dlib/june06/chan/06chan.html>>

- Córdoba González, Saray. Acceso abierto al conocimiento en la Universidad de Costa Rica. Universidad de Costa Rica, 2011 <<http://www.vinv.ucr.ac.cr/docs/multimedia/repositorios-ucr.pdf>>

- Codina, Lluís; Marcos, Mari-Carmen; Pedraza, Rafael. *Web semántica y sistemas de información documental*. Gijón: Trea, 2009

- Codina, Lluís; Rovira, Cristòfol. "Web semántica: visión global y análisis comparativo", en *Tendencias en documentación digital*. Gijón: Trea, 2006.

- Codina, Lluís. "Internet invisible y web semántica: ¿el futuro de los sistemas de información en línea?" *Revista Tradumàtica – Traducció i la Tecnologia de la Informació i la Comunicació*, Noviembre 2003 vol. 2 La Documentació <<http://fti.uab.es/tradumatica/revista>>

- Coyle, Karen; Thomas Baker. "Guidelines for Dublin Core Application Profiles." *Dublin Core Metadata Initiative*, 2009 <<http://dublincore.org/documents/profile-guidelines/index.shtml>>

- Coyle, Karen. Understanding metadata and its purpose. *Journal of Academic Librarianship*, 2005, v. 31, n. 2, pp 160-163 <<http://www.kcoyle.net/jal-31-2.html>>
- Cutter, Charles A. Rules for a Dictionary Catalog. 4rth. Ed. Washington: GPO, 1904.
- Currás, Emilia. *Ontologías, taxonomías y tesauros: Manual de construcción y uso*. Gijón, Trea, 2005.
- Cyganiak, Richard; Jentzsch, Anja. *LOD Cloud diagram state*, 2011 <<http://lod-cloud.net/state/>>
- Cyganiak, Richard; Jentzsch, Anja. *Linking Open Data cloud diagram*, 2011 <<http://lod-cloud.net/>>
- Cyganiak, Richard; Wood, David; Lanthaler, Markus. *RDF Concepts and abstract Sintaxis*, 2014 <<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>>
- Cyganiak, Richard; Reynolds, Dave. *The RDF Data Cube Vocabulary*, 2014. <<http://www.w3.org/TR/vocab-data-cube/>>
- Cyganiak, Richard; Zhao, Jun; Alexander, Keith; Hausenblas, Michael; “Vocabulary of Interlinked Datasets (VoID)” *DERI Vocabularies*, 2011 <<http://vocab.deri.ie/void>>
- D'Arcus, Bruce; Giasson, Frédérick. Bibliographic Ontology Specification Specification Document - 4 November 2009 <<http://purl.org/ontology/bibo/>>
- Daudinot Founier, Isabel. “Organización y recuperación de información en Internet” *ACIMED 2006*, vol. 14, n. 5.

<http://bvs.sld.cu/revistas/aci/vol14_5_06/aci06506.htm>
>

- Dbpedia. About, 2014, <<http://wiki.dbpedia.org/About>>
- De la Rosa, Antonio. *Introducción a XML para Documentalistas*. "Hipertext.net", núm. 1, 2003. <<http://www.hipertext.net>>
- DIGIBÍS. Bibliotecas Virtuales producidas por DIGIBÍS, 2014<<http://www.digibis.com/bibliotecas-virtuales/producidas-por-digibis.html>>
- Dirección General del Libro, Archivos y Bibliotecas. *Reglas de Catalogación*. Madrid: Ministerio de Cultura, Secretaría General Técnica, 1995
- Dublin Core Metadata Initiative (DCMI). *Library Application Profile*, 2004 <<http://dublincore.org/documents/library-application-profile/>>
- Dublin Core Metadata Initiative (DCMI). *Qualified Dublin Core*, 2000 <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>
- Dublin Core Metadata Initiative (DCMI). *Dublin Core Metadata Terms*, 2012 <<http://dublincore.org/documents/dcmi-terms/>>
- Dublin Core Metadata Initiative (DCMI). *Dublin Core Metadata Element Set*. 2012 <<http://dublincore.org/documents/dces/>>

- Duval, Erik; Hodgins, Wayne; Sutton, Stuart; Weibel, Stuart L. "Metadata principles and practicalities". *D-Lib Magazine*, 2002 vol. 8 n. 4 <<http://www.dlib.org/dlib/april02/weibel/04weibel.html>>
- Elba García, Nelida. Caballero, Sergio. *Metadatos: necesidad e importancia de integrar estándares*. Argentina: Universidad Nacional de Misiones, 2009 <http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109_10a.pdf>
- Elsevier <<http://www.elsevier.com/about/open-access/open-archives>>
- Eito Brun, Ricardo. *Lenguajes de marcas para la gestión de recursos digitales*. Gijón: Ediciones Trea, 2008
- Escolano Rodríguez, Elena. "Principios internacionales y las Reglas de catalogación". En: I Congreso sobre Principios de Catalogación: influencia y panorama europeo, Madrid, 14 de abril de 2005 <http://www.bne.es/es/Servicios/NormasEstandares/DocumentosProfesionales/PrincipiosInternacionalesDeCatalogacion/Docs/congreso_catalogacion_1.pdf>
- España. Ley 37/2007, de 16 de noviembre, sobre reutilización de la información en el sector público. Boletín Oficial del Estado, núm. 276 de 17 de noviembre de 2007, páginas 47160 a 47165 <<http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>>
- España. Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Boletín Oficial del Estado,

núm. 131 de 2 de junio de 2011, páginas 54387 a 54455
<<http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf>>

- Estivill Rius, Assumpció. “Momentos estelares de la catalogación en el cincuentenario de los Principios de París“ Bid, 2012 n. 28, <<http://bid.ub.edu/28/estivill2.htm>>
- Estivill Rius, Assumpció. Resource description and access, RDA: un nuevo retraso para preparar mejor el cambio *El profesional de la información*, 2011 vol. 20, n. 6, pp. 694-700
- European Commission. *Programas Marco*, 2013 <http://europa.eu/rapid/press-release_IP-13-1232_es.htm>
- European Commission. *Programa de financiación europea Horizon 2020* <<http://ec.europa.eu/programmes/horizon2020/>>
- European Comisión. DRIVER: *Digital Repository Infrastructure Vision for European Research*, 2014 <<http://www.driver-repository.eu/>>
- European Comision. *Open Access Pilot in FP7*, 2008 <http://www.uv.es/operuv/guies/open-access-pilot_en.pdf>
- European Commission. European IPR Helpdesk. “*Fact Sheet: Open access to publications and data in Horizonte 2020: Frequently Asked Questions (FAQ)*” 2014

- <http://www.iprhelpdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf>
- European RDA Interest Group (EURIG). Reunión anual de EURIG, Viena 25 de abril de 2014
<<http://www.bne.es/webdocs/Inicio/Perfiles/Bibliotecarios/EURIGViena2014.pdf>>
 - European Research Advisory Board. *Final Report. Scientific Publication: Policy on Open Access*, 2006
<http://ec.europa.eu/research/eurab/pdf/eurab_sci_pub_report_recomm_dec06_en.pdf>
 - Europeana Open Linked Data 2014
<<http://labs.europeana.eu/api/linked-open-data/introduction/>>
 - FAO. AGRIS, 2014 <<http://agris.fao.org/agris-search/index.do>>
 - Ferreya, Diego. “Explotación de repositorios OAI a través de la armonización de vocabularios controlados” En: *2do Encuentro Nacional de Catalogadores Biblioteca Nacional. Buenos Aires, Argentina 25, 26, 27 Noviembre, 2009*
<http://www.bn.gov.ar/descargas/catalogadores/presentaciones/251109_11b.pdf>
 - FOAF Project: *Friend of a Friend: introducing FOAF, 2000* <<http://www.foaf-project.org/original-intro>>
 - Food and Agriculture Organization of the United Nations (FAO). *FAO Authority Description Concept Scheme*
<<http://aims.fao.org/community/pages/authority-control-content-model>>

- Folmer, Erwin Johan Albert. *Quality of semantic standards*, 2012
<http://doc.utwente.nl/81390/1/thesis_E_Folmer.pdf>
- Fox, Edward. "Open archives initiative". *D-lib magazine*, 2000 vol. 6 n. 6
<<http://www.dlib.org/dlib/june00/06inbrief.html#FOX>>
- Franklin, Jack. "Open access to scientific and technical information: the state of the art". *Information Services and Use*, 2003, vol 23, n. 2,3 pp.67-86
- Fundación Cotec para la Innovación Tecnológica. *Informe COTEC 2013*. Madrid: Fundación Cotec para la Innovación Tecnológica, 2013
<http://www20.gencat.cat/docs/DAR/DE_Departament/DE02_Estadistiques_observatoris/27_Butlletins/02_Butlletins_ND/Fitxers_estatics_ND/2013_fitxers_estatics/0124_2013_RDi_COTEC.pdf>
- Fundación Española de Ciencia y Tecnología (FECYT). "Declaración de la Alhambra sobre acceso abierto" En: *Seminario para el acceso abierto a la información científica: Políticas de desarrollo en el Sur de Europa*, Granada 13 y 14 de mayo 2010,
<http://oaseminar.fecyt.es/Resources/Documentos/ADeclaration/AD_OA_Spanish.pdf>
- Furrie, Betty. *Understanding MARC Bibliographic*. 2003. Traducción de Ageo García Barbosa Conociendo MARC bibliográfico: catalogación legible por máquina, 2003. <<http://www.loc.gov/marc/umbspa/umbspa.html>>

- Fuster, Justo Pastor. *Biblioteca valenciana de los escritores que florecieron hasta nuestros días: con adiciones y enmiendas a la de D. Vicente Ximeno*, 2 vols, Valencia, 1827-1830.
- García-García, Alicia; Rodríguez-Gairín, Josep-Manuel; Saorín, Tomás; González, Luis-Millán; García-Massó, Xavi; Ferrer-Sapena, Antonia; Peset, Fernanda. “ODiSEA: International Registry on Research Data”. *Bid: textos universitaris de Biblioteconomia i Documentació*, 2012, n. 29 <<http://bid.ub.edu/29/garcia2.htm>>
- García Gutiérrez, Antonio. *Introducción a la Documentación informativa y periodística*. Madrid: MAD, 1999
- García-Marco, Francisco Javier. “Schema.org: la catalogación revisitada” *Anuario ThinkEPI*, 2013 vol. 7, pp. 169-172
- Garrido Arilla, María Rosa. *Teoría e historia de la catalogación de documentos*. Madrid: Síntesis.1996
- Garrido Arilla, María Rosa. *Contienda por el control documentario: etapas pretécnica y técnica en catalogación*. En: I Congreso Universitario en Ciencias de la Documentación: teoría, historia y metodología de la documentación en España, Madrid, 2000
- Garzon Farinos, Fernanda. *Marco y antecedentes para crear biografías en unabiblioteca digital interoperable sobre científicos y humanistas valencianos*. *Vestigium*. Trabajo de investigación inédito. UPV, 2011

- GeoNames.org. *GeoNames database*
<<http://www.geonames.org>>
- Generalitat Valenciana. *Investigación Científica y el Desarrollo Tecnológico en la Comunitat Valenciana: Informe anual 2011*, Valencia: Generalitat Valenciana, 2012
<http://www.vicepresidencia.gva.es/documents/80920710/80950153/Informe_anual_2011.pdf/5fb0643e-701a-459e-8172-28e7215db70b>
- Generalitat Valenciana. *Investigación Científica y el Desarrollo Tecnológico en la Comunitat Valenciana: Informe anual 2012*, Valencia: Generalitat Valenciana, 2013
<<http://www.vicepresidencia.gva.es/documents/80920710/80950153/Resumen+ejecutivo+2012-web.pdf/cfd79c8a-decd-424f-8a51-212169f66953>>
- Gil Urdiciáin, Blanca. *Manual de lenguajes documentales*. Madrid: Editorial Noesis, 1996
- Gilliland, Anne J. "Setting the Stage." In *Introduction to Metadata*. 2nd ed., v. 3.0, edited by Murtha Baca. Los Angeles: Getty Research Institute, 2008. <http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html>
- Gómez, Nancy Diana; Arias, Olga Margarita. "El cambio de paradigma en la comunicación científica". *Información, cultura y sociedad*, 2002. n. 6. <<http://eprints.rclis.org/archive/00003649>>
- González-Crespo, Rubén; Ferro Escobar, Roberto; Yanet Velazco, Sandra; Moreiro, José Antonio. "Propuesta de

una taxonomía y ontología para clasificación inteligente de grupos de investigación”. *Sociedad y Utopía. Revista de Ciencias Sociales*, Noviembre de 2012, n. 40. pp. 18-50

- González González, Enrique (1989). Hacia una definición del término humanismo, *Estudis: revista de historia moderna*, n. 15, p. 45-66
- Gómez Dueñas, Laureano Felipe. “La iniciativa de Archivos Abiertos (OAI) un nuevo paradigma en la comunicación científica y el intercambio de información”. *Revista Códice*, 2005 vol. 2, n. 1, pp. 21-48.
- Gómez-Pérez, Asunción. “Datos enlazados en la BNE” Biblioteca Nacional de España, 14 de diciembre de 2011 <<http://www.slideshare.net/bne/datos-enlazados>>
- Gruber, Thomas R. “A Translation Approach to Portable Ontology Specifications”. *Knowledge Acquisition*, 1993, Vol. 5, n. 2, pp.199-220
- Guralnick, Robert; Constable, Heather; Wiczorek, John [et al.] “Data's shameful neglect”. *Nature*, Sept. 2009, vol. 461, no. 145 <<http://www.nature.com/nature/journal/v461/n7261/full/461145a.html>>
- Guinchat, Claire; Menou, Michel. *Introducción general a las ciencias y técnicas de la información y documentación*. Montevideo: Unesco, 1983
- Haslhofer, Bernhard; Schandl, Bernhard. “Interweaving OAI-PMH Data Sources with the Linked Data Cloud”.

Int. J. Metadata, Semantics and Ontologies, 2009 vol. 5.
n.1 pp.17-31

- Harris, Steve; Seaborne, Andy. SPARQL 1.1 Query Language, W3C Recommendation 21 March 2013. 2013 <<http://www.w3.org/TR/sparql11-query/>>
- Heath, Tom; Bizer, Christian. *Linked data: evolving the Web into a global data space*. (1st edition). Synthesis lectures on the semantic web: theory and technology, February, vol. 1, n. 1 pp. 1-136. Morgan & Claypool 2011 <<http://linkeddatabook.com/editions/1.0/>>
- Hernández-Pérez, Tony; García-Moreno, María-Antonia (2013). “Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios”. *El profesional de la información*, 2013, mayo-junio, vol. 22, n. 3, pp. 259-263. <<http://dx.doi.org/10.3145/epi.2013.may.10>>
- Hernández-Pérez, Tony; David Rodríguez Mateos, Gema Bueno De la Fuente. “Open Access: el papel de las bibliotecas en los repositorios institucionales de acceso abierto”. *Anales de Documentación*, 2007 vol.10 <<http://revistas.um.es/analesdoc/article/view/1141>>
- Hillman, Diane; Coyle, Karen; Phipps, Jon; Dunsire, Gordon. “RDA vocabularies: process, outcomes, use” *D-Lib Magazine* 2010 <<http://dlib.org/dlib/january10/hillmann/01hillmann.html>>
- IEEE Computer Association. *IEEE Standard Glossary of Software Engineering Terminology* (No. 610. 12-1990) New York: IEEE Computer Society, 1990.

- International Federation of Library Associations and Institutions (IFLA). *ISBD descripción Bibliográfica Internacional Normalizada. Edición consolidada*, 2011
<<http://www.ifla.org/files/assets/hq/publications/series/44-es.pdf>>
- International Federation of Library Associations and Institutions (IFLA). *Statement on open access to scholarly literature and research documentation*, 2004
<<http://archive.ifla.org/V/cdoc/open-access04.html>>
- International Federation of Library Associations and Institutions (IFLA). *Declaración de principios internacionales de catalogación*, 2009
<http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-es.pdf>
- International Federation of Library Associations and Institutions (IFLA). *ISBD: (ER) (Electronic Resources) International standard bibliographic description for electronic resources*, 1997
<<http://archive.ifla.org/VII/s13/pubs/isbd.htm>>
- International Federation of Library Associations and Institutions (IFLA). *FRBR Requisitos funcionales de los registros bibliográficos*. Madrid: Ministerio de Cultura, Secretaría General Técnica, 2004
<<http://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf>>
- International Federation of Library Associations and Institutions (IFLA). *FRAD Requisitos Funcionales de los Datos de Autoridad*. Madrid: Ministerio de Cultura, Secretaría General Técnica, 2009

<http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-es.pdf>

- International Federation of Library Associations and Institutions (IFLA). *FRSAD Requisitos Funcionales para Datos de Autoridad de Materia*, 2011. <<http://www.ifla.org/files/assets/cataloguing/frsad/frsad-final-report-es.pdf>>
- International Conference on Cataloguing Principles (Paris:1961). Report.–London: International Federation of Library Associations, 1963, pp. 91-96. También disponible en: *Library Resources and Technical Services*, vol. 6 (1962), pp. 162-167; y *Statement of principles adopted at the International Conference on Cataloguing Principles, Paris, October, 1961.– Annotated edition / with commentary and examples by Eva Verona.– London: IFLA Committee on Cataloguing, 1971*
- Instituto Nacional de Estadística. Encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación (TIC) en los hogares en 2008 <<http://www.ine.es/>>
- Instituto Nacional de Estadística. Encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación (TIC) en los hogares en 2009 <<http://www.ine.es/>>

- Instituto Nacional de Estadística. Encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación (TIC) en los hogares en 2010 <<http://www.ine.es/>>
- Instituto Nacional de Estadística. Encuesta sobre Equipamientos y Uso de Tecnologías de la Información y Comunicación (TIC) en los hogares en 2011 <<http://www.ine.es/>>
- International Organization for Standardization. *Norma Internacional ISO/IEC 11179 Information Technology – Metadata Registries (MDR)*, 2012 <http://jtc1sc32.org/doc/N2201-2250/32N2224T-text_for_ballot-CD_11179-1.pdf>
- Isaac, Antoine; Summers, Ed. *SKOS Simple Knowledge Organization System Primer*, 2009 <<http://www.w3.org/TR/skos-primer/>>
- Isaac, Antoine; Waites, William; Young, Jeff; Zeng, Marcia. *Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets*, 2011 <<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>>
- Jacobs, Ian; Walsh, Norman. *Architecture of the World Wide Web*, 2004 <<http://www.w3.org/TR/webarch/>>
- Jaques, Yves and Anibaldi, Stefano and Celli, Fabrizio and Subirats-Coll, Imma and Stellato, Armando and Keizer, Johannes *Proof and Trust in the OpenAGRIS Implementation.*, 2012. In International Conference on Dublin Core and Metadata Applications 2012, Kuching,

Sarawak, Malaysia, 3-7 September 2012.

- Juárez Santamaría, Beatriz. “Uso de los metadatos en el orden documental”. *Información, producción y servicios*, 2007, vol. 10, n. 42 pp.8
- Joint Steering Committee (JSC). Annual Reports, 2013 <<http://www.rda-jsc.org/annrep.html>>
- Joint Steering Committee (JSC). *Report and Recommendations of the U.S. RDA Test Coordinating Committee*, 20 June, 2011 <<http://www.loc.gov/bibliographic-future/rda/source/rdatesting-finalreport-20june2011.pdf>>
- Kaplan, Deborah; Saunes, Anne; Wilczek, Eliot. “Archival description in OAI-ORE”. *Journal of Digital Information*, 2011, vol. 12, n. 2 <<http://journals.tdl.org/jodi/article/view/1814/1769>>
- Krichel, Thomas *The Semantic web and an introduction to RDF.*, 2002 <<http://hdl.handle.net/10760/4214>>
- Küster Boluda, Inés; Hernández Fernández, Asunción. “De la Web 2.0 a la Web 3.0: antecedentes y consecuencias de la actitud e intención de uso de las redes sociales en la web semántica” . *Universia Business Review*, 2013, n°. 37 , pp. 104-119 <http://ubr.universia.net/pdfs_web/UBR37_Web/06.pdf>
- Lacasa Otín, María Ángeles; Martínez Ezquerra, Piedad. “La IFLA: origen, evolución y situación actual”. *Boletín de la ANABAD*, 1993, vol 3, n. 2, pp.71-108

- Lamarca Lapuente, María Jesús. Hipertexto, el nuevo concepto de documento en la cultura de la imagen. Tesis doctoral. Universidad Complutense de Madrid, 2013
- Lazinger, S. *Digital preservation and metadata: history, theory, practice*. En glewood, Colorado: libraries Unlimited, 2001
- Library of Congress. *Bibliographic Framework as a web of data: linking data model and supporting services*, 2012 <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>
- Library of Congress. *MADS User Guidelines*, 2012, <<http://www.loc.gov/standards/mads/userguide/index.html>>
- Library of Congress, *MODS to MARC 21 mapping*, 2009 <<http://www.loc.gov/standards/mods/mods2marc-mapping.html>>
- Library of Congress. *Jefferson's Legacy: A Brief History of the Library of Congress*, 2006. <<http://www.loc.gov/loc/legacy/>>
- Library of Congress. *Library of Congress Subject Headings*. <<http://id.loc.gov/authorities/subjects.html>>
- Library of Congress. *MADS: Metadata Authority Description Schema*, 2013. <<http://www.loc.gov/standards/mads/>>
- Library of Congress. *MADS/RDF Primer*, 2012 <<http://www.loc.gov/standards/mads/rdf/>>

- Library of Congress. *MARC21: localización y accesos electrónicos*, 2007.
<<http://www.loc.gov/marc/marcspa.html>>
- Library of Congress. *MARC Standards. RDA in MARC*, 2012. <<http://www.loc.gov/marc/RDAinMARCspa-10-22.pdf>>
- Library of Congress. *MARC XML Schema*, 2014
<<http://www.loc.gov/standards/marcxml/>>
- Library of Congress. *METS: Metada Enconding and Transmission Standard*, 2014.
<<http://www.loc.gov/standards/mets/mets-home.html>>
- Library of Congress. *MODS: Metadata Object Description Schema*, 2013.
<<http://www.loc.gov/standards/mods/>>
- Library of Congress. *Response of the Library of Congress, the National Agricultural Library, and the National Library of Medicine to the RDA Test Coordinating Committee*.13 June 2011.
<<http://www.loc.gov/bibliographic-future/rda/source/rda-execstatement-13june11.pdf>>
- López Guillamón, Ignacio. “Evolución reciente de la catalogación”. *Anales de documentación*, 2004. vol. 7, pp.141-152
<<http://revistas.um.es/analesdoc/article/view/1661/1711>>
- López Piñero, José María; Terrada, M^a Luz; Aleixandre, Rafael; Valderrama, Juan Carlos; González, Gregorio; Navarro, Carolina. *Santiago Grisolia*.

Recuerdos. Valencia: Consell Valencià de Cultura, 2008

- Lynch, Clifford. A. "Institutional repositories: essential infrastructure for scholarship in the Digital Age" *ARL: Bimonthly report*, 2003, n. 226. <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>
- Mancebo, María Fernanda. *La Universidad de Valencia de la Monarquía a la República (1919-1939)*, Valencia: Instituto de Cultura Juan Gil-Albert, Universitat de València, 1994.
- Margaix Arnal, Dídac. "Redes sociales y web 2.0 en 2009" *Anuario Think Epi*, 2010, vol. 4, pp. 211-216.
- Martín Galán, Bonifacio; Méndez Rodríguez, Eva M^a. "Citas y notas en documentos HTML: una propuesta de organización y acceso". En: *Organización del Conocimiento en Sistemas de Información y Documentación: actas del III Encuentro de ISKO-España (3º. Getafe. 1997)*. Zaragoza: Fco. Javier García Marco, 1998, pp. 391-402.
- Martín, Sandra Gisela; Angelozzi, Silvina Marcela. "Análisis y comparación de metadatos para la descripción de recursos electrónicos en línea". En: *III Encuentro Internacional de Catalogadores*, Buenos Aires, Argentina, 2007. <<http://eprints.rclis.org/15906/1/2007%20Angelozzi-Martin%20Analisis%20y%20comparacion%202%20%281%29.pdf>>
- Martínez Conde, María Luisa. Hispana: directorio y recolector de recursos digitales. En *Humanidades*

digitales: edición y difusión: seminario de investigación, A Coruña, 2 y 3 de julio de 2012 <<http://www.bidiso.es/sielae/upload/estaticas/file/PDSHD7.pdf>>

- Martín-Gavilán, César *Temas de Biblioteconomía: El formato MARC: variedades geográficas y de aplicación. MARC* 21, 2008 <<http://eprints.rclis.org/14525/1/marc.pdf>>
- Martínez Equihua, Saúl. *Biblioteca Digital Conceptos, recursos y estándares* Buenos Aires: Alfagrama, 2007, 170 p.
- Melero, Remedios; Abad, Francisca; Abadal, Ernest; Amat, Carlos B.; Gimenez, Francesc; Rodríguez-Gairín, Josep Manuel; Rodríguez, Nerea. “Dulcinea: iniciativa para el análisis de los derechos de copyright y autoarchivo de revistas científicas españolas”. En: *XI Jornadas Españolas de Documentación* (Fesabid): 20, 21 y 22 de mayo de 2009, Auditorio Palacio de Congresos de Zaragoza, 2009 pp.117-122 <http://www.fesabid.org/zaragoza2009/Libro_Actas_Fesabid_2009.pdf>
- Melero, Remedios; Abad, Francisca. “Revistas open Access: características, modelos económicos y tendencias”. *BiD: textos universitaris de Biblioteconomia i Documentació*, juny 2008, n. 20 <<http://www.ub.edu/bid/20meler2.htm>>
- Melero, Remedios. “Acceso abierto a las publicaciones científicas: definición, recursos, copyright e impacto. *El profesional de la información*, 2005, vol. 14, n. 4 (jul.-

- ago.), pp. 255-266. <<http://eprints.rclis.org/6571/1/EPI-melero.pdf>>
- Méndez, Eva y Senso, José A. “Unidad de Autoformación. Introducción a los Metadatos: estándares y aplicación”. *SEDIC*, 2004 <<http://www.sedic.es/autoformacion/metadatos/tema1.htm>>
 - Méndez, Eva. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. *Jornades Catalanes de Documentació* (7. 1999. Barcelona) Barcelona: Col·legi de Bibliotecaris-Documentalistes de Catalunya, 1999, pp. 487-498
 - Méndez, Eva; Greenberg, Jane. “Linked data for open vocabularies and HIVE’s global framework”. *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 236-244. <<http://dx.doi.org/10.3145/epi.2012.may.03>>
 - Messina Declaration <<http://www.aepic.it/conf/Messina041/viewpaper5af5.pdf?id=49&cf=1>>
 - Miller, Paul. “Interoperability: What is it and Why should I want it?”. En: *Ariadne*, 2000 vol. 24. <<http://www.ariadne.ac.uk/issue24/interoperability>>
 - Miller, J. Steven. *Metadata and Cataloguin online resources: selected reference documents, web sites and articles* 2011 <<https://pantherfile.uwm.edu/ml/wwww/resource.html>>

- Ministerio de Educación, Cultura y Deporte. *Lista de Encabezamientos de Materia para Bibliotecas Públicas en SKOS*, 2013 <<http://id.sgcb.mcu.es/lem/>>
- Morato, Jorge; Sanchez-Cuadrado, Sonia; Ruiz-Robles, Alejandro; Moreiro-González, José-Antonio. “Visualización y recuperación de la información en la web semántica”. En: *El Profesional de la Información*, mayo-junio, 2014, v. 23, n. 3, pp.319-329
- Moreiro, José-Antonio. “Hacia la primacía de los conceptos sobre los términos en los vocabularios para la web semántica”. En: *Anuario ThinkEPI*, 2013, vol. 7, pp.173-177.
- Morville, Peter; Rosenfeld, Louis. *Information Architecture for the World Wide Web*. Sebastapol, CA: O'Reilly Media, 1998
- Moya-Anegón, Félix de; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; González-Molina, Antonio; et al. *Indicadores bibliométricos de la actividad científica española 2010*, Madrid: FECYT ,2013 <http://icono.fecyt.es/informespublicaciones/Documentos/indicadores%20bibliometricos_web.pdf>
- Munford, William A. *A history of the Library Association, 1877-1977*, p. 3. London: Library Association, 1977
- National Information Standards Organization (NISO). *Understanding metadata*. Bethesda, MD: NISO Pres. 2004

<<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>

- National Information Standards Organization (NISO). *ANSI/NISO Z39.19: Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Baltimore: NISO Pres. 2005
<http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf>
- National Library of Medicine (NLM). *MeSH: Medical Subject Headings Classification*, 2013
<<http://www.nlm.nih.gov/mesh/meshhome.html>>
- Nature Publishing Group
<http://www.nature.com/authors/author_resources/declaration.html>
- Nelson, Theodor. *Literary Machines*. Sausalito (CA): Mindful Press, 1993
- Nogales Flores, Tomás. *MARC en XML*. *Universidad Carlos III de Madrid*, 2004
<<http://www.bib.uc3m.es/~nogales/cursos/marcxml.html>>
- OKFN. *Medline RDF*, 2011
<http://eris.okfn.org/ww/2011/05/medline/>
- OKFN. *Linked Open Vocabularies (LOV)*
<<http://lov.okfn.org/dataset/lov/#VocabListTable>>

- *Library of Congress Subject Headings*,
LCSH<<http://id.loc.gov/authorities/subjects.html>>

- Online Computer Library Center (OCLC). Dewey Services, 2014
<<http://www.oclc.org/dewey/webservices.en.html>>

- Online Computer Library Center (OCLC). History of the OAIster, 2014 <<http://oclc.org/oaister/about.en.html>>

- Online Computer Library Center (OCLC). Expanding el concepto de control bibliográfico universal: Fichero de Autoridades Internacional Virtual (VIAF) <http://www.bne.es/export/sites/BNWEB1/es/LaBNE/Cooperacion/CooperacionInternacional/VIAF/docs/Presen-tacixn_oficial_del_proyecto_VIAF.pdf>

- Open Archives Initiative. Object Reuse and Exchange. <<http://www.openarchives.org/ore/>>

- Open Society Institute (OSI). *Budapest Open Access Initiative (BOAI)*, 2002
<<http://www.budapestopenaccessinitiative.org/read>>

- OpenDOAR <<http://www.opendoar.org/find.php>>

- Oren, Eyal; Delbru, Renaud; Catasta, Michele; Cyganiak, Richard; Tummarello, Giovanni. "Sindice.com: a document-oriented lookup index for open linked data". En: *Intl journal of metadata, semantics and ontologies*, 2008. v. 3, n. 1, pp. 37-52. <<http://dx.doi.org/10.1504/IJMSO.2008.021204>>

- Orduna-Malea, Enrique. “Reutilización e intercambio de objetos digitales compuestos en la Web: el proyecto OAI-ORE”. En: *Anuario ThinkEPI*, 2009 EPI SCP, pp.45-48
- Ortí y Figuerola, Francisco. *Memorias históricas de la fundación y progreso de la insigne universidad de Valencia*, Madrid, 1730
- Ortiz-Repiso Jiménez, Virginia. “Nuevas perspectivas para la catalogación: metadatos versus Marc” *Rev. Esp. Doc. Cien.* 1999, vol. 22, n. 2, <<http://redc.revistas.csic.es/index.php/redc/article/view/338/546>>
- O’Reilly, Tim. *What is web 2.0*, 2005 <<http://oreilly.com/web2/archive/what-is-web-20.html>>
- O’Reilly, Tim; Battelle, John. “Web Squared: web 2.0 five years on”. In: *Web 2.0 Summit, San Francisco 20-22 october 2009* <<http://www.web2summit.com/web2009/public/schedule/detail/10194>>
- Ouksel, Aris M.; Sheth, Amit. *Semantic Interoperability in Global Information Systems: A brief introduction to the research area and the special section*. ACM SIGMOD Record, 1999. vol. 28 n. 1, pp. 5-12
- Oxford Journals <http://www.oxfordjournals.org/for_authors/repositories.html>
- Pascual Bueno, José. *Inventario y estudio del archivo del Dr. Juan José Barcia Goyanes. Legado a la Generalitat*

Valenciana. Tesis Doctoral. Valencia: Universidad Católica de Valencia San Vicente Mártir, 2013.

- Pastor-Sánchez, Juan-Antonio; Martínez-Méndez, Francisco-Javier; Rodríguez-Muñoz, José-Vicente. “Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data ”. *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 245-253 <<http://hdl.handle.net/10760/17066>>
- Pastor-Sánchez, Juan Antonio. SKOS Simple Knowledge Organization System: Qué es skos? <<http://skos.um.es/acerca/index.php>>
- Pastor-Sánchez, Juan-Antonio. *Diseño de un sistema colaborativo para la creación y gestión de tesauros en Internet basado en SKOS*. Tesis en Red, 2009. ISBN 978-84-692-1932-4
- Pastor-Sánchez, Juan-Antonio; Martínez Méndez, Francisco Javier. “Manual de SKOS (simple knowledge organization system, sistema para la organización del conocimiento simple)”. *Anales de Documentación*, 2010, vol. 13, pp. 285-320.
- Pastor-Sánchez, Juan-Antonio. *Tecnologías de la web semántica*. Barcelona: Editorial UOC, 2011
- Pastor-Sánchez, Juan-Antonio. “Prospectiva de la web semántica: divergencia tecnológica y creación de mercados linked data”. *Anuario ThinkEPI*, 2012, v. 6, pp. 1269-275

- Pastor-Sánchez, Juan-Antonio. “Los cms como pieza fundamental en el despliegue de la web semántica”. *Anuario ThinkEPI*, 2012, v. 6, pp. 184-189
- Pastor-Sánchez, Juan-Antonio; Orduña-Malea, Enrique; Saorín, Tomás. “Marcado semántico automático en gestores de contenidos: integración y cuantificación”. *El profesional de la Información*, septiembre-octubre, 2013, vol. 22, n. 5, pp. 381-391
- Pedraza-Jiménez, Rafael; Codina, Lluís; Rovira, Cristòfol. “Web semántica y ontologías en el procesamiento de la información documental”. En: *El profesional de la información*, 1997, vol. 16, n. 6, pp. 569-578.
- Peis Redondo, Eduardo; Herrera Viedma, Enrique; Hassan Montoro, Yusef; Herrera, Juan Carlos. “Ontologías, metadatos, y agentes: recuperación semántica de la información”. En: *Actas de las II Jornadas de Tratamiento y Recuperación de la información (JOTRI)*, 2003 pp.157-165
- Peis, Eduardo; Moya, Félix de. “Sgml y servicios de información”. En: *El profesional de la información*, 2000, junio, v. 9, n. 6, pp. 4-17.<<http://www.elprofesionaldelainformacion.com/contenidos/2000/junio/1.pdf>>
- Peset Reig, Mariano; Peset Reig, Luis. *La Universidad española (siglos XVIII y XIX). Despotismo ilustrado y revolución liberal*. Madrid, Taurus, 1974.
- Peset, Fernanda; Albiñana, Ricardo; Morales, Silvia. “Internet invisible: un recurso terciario en la Red”. *El*

profesional de la información, vol. 9, n. 7-8, julio-agosto 2000.

<<http://www.elprofesionaldelainformacion.com/contenidos/2000/julio/4.pdf>>

- Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. “Open data y Linked open data: su impacto en el área de bibliotecas y documentación”. En: *El profesional de la información*, 2011, marzo-abril, v. 20, n. 2, pp. 165-173. DOI: 10.3145/epi.2011.mar.06
- Peset, Fernanda; Ortín Pérez, Alfredo. *Tendencias internacionales en el acceso libre a la documentación científica digital: OAI-Open Archives Initiative.*, 2007 In: UNSPECIFIED, (ed.) *Derecho, historia y universidades: estudios dedicados a Mariano Peset*. Oc 978-84-370-6864-0. v ll.978-84-370-6866-4. Universidad de Valencia, pp. 441-452. <<http://eprints.rclis.org/6470/>>
- Pérez-Montoro, Mario. *Arquitectura de la información en entornos web*. Gijón: Ediciones TREA, 2010
- Picco, Paola y Ortiz Repiso, Virginia. “RDA, el nuevo código de catalogación: cambios y desafíos para su aplicación” *Revista Española de Documentación Científica*, 2012, vol. 35, n. 1, enero-marzo, pp.145-173
- Pierre, Margaret St.; LaPlant, William P. “Issues in crosswalking content metadata standards”. *NISO*, 1998 <http://www.niso.org/publications/white_papers/crosswalk>

- Pinto Molina, María. “Introducción al análisis documental y sus niveles: el análisis de contenido”. *Boletín de la ANABAD*, 1989, vol 39, n. 2, pp. 323-342

- REBIUN. Declaración de CRUE/REBIUN en apoyo del modelo de acceso abierto. En *XII Asamblea General de REBIUN, la CRUE-Conferencia de Rectores de Universidades Españolas*. Universidad del País Vasco, 3-5 noviembre, 2004
<<http://biblioteca.upc.es/Rebiun/nova/archivosNoticias/89.pdf>>

- Registry of Open Access Repository
<<http://roar.eprints.org/>>

- Ríos Hilario, Ana Belén; Martín González, Yolanda. “Aplicación de los Requisitos funcionales de los registros bibliográficos (FRBR) en los catálogos en línea”. *Acimed: revista cubana de los profesionales de la información y la comunicación en salud*, 2005, Vol. 13, N. 4.
<http://bvs.sld.cu/revistas/aci/vol13_4_05/aci05405.pdf>

- Rodríguez Perojo, Keilyn; Ronda León, Rodrigo. *La web semántica: un nuevo enfoque para la organización y recuperación de información en el web*, 2011<http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm>

- Rosling, Hans. *The best stats you've ever seen*. TED, 2006
<http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen>

- Sánchez-Cuadrado, Sonia; Colmenero-Ruiz, María-Jesús; Moreiro-González, José-Antonio. “Tesauros: estándares y recomendaciones”. En: *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 229-235
- Sánchez-Cuadrado, Sonia; Morato-Lara, Jorge; Palacios-Madrid, Vicente; Llorens-Morillo, Juan; Moreiro-González, Antonio. “De repente, ¿todos hablamos de ontologías?”. En: *El Profesional de la información*, 2007 vol. 16, n. 6, pp. 562-568.
- Santana Arroyo, Sonia. “El modelo FRBR en el servicio de referencia y su implementación en los catálogos automatizados”. *Acimed*, 2011; Vol. 22, n.3 <<http://www.acimed.sld.cu/index.php/acimed/article/view/182/149>>
- Saorín-Pérez, Tomás. “Cómo *linked open data* impactará en las bibliotecas a través de la innovación abierta”. *Anuario ThinkE- PI*, 2012, v. 6, pp. 288-292 <<http://www.thinkepi.net/como-linked-open-data-impactara-en-las-bibliotecas-a-traves-de-la-innovacion-abierta>>
- Saorín-Pérez, Tomás; Pastor-Sánchez, Juan-Antonio. “Gestión documental y de contenidos web”. *Anuario ThinkEPI*, 2012, v. 6, pp. 232-239
- Saorín-Pérez, Tomás. “Drupal: fundamentos técnicos”. En: Tramullas, Jesús. *Drupal para bibliotecas y archivos*. Zaragoza: Universidad de Zaragoza, 2010

- Schema.org. *Getting started with schema.org* <<https://schema.org/docs/gs.html>>
- Senso, José A.; De la Rosa Piñero, Antonio. *Evolución del Dublin Core Metadata Initiative*. 2003 <<http://www.ugr.es/~jsenso/curriculum/dcmi.pdf>>
- Shotton, David; Peroni, Silvio. *CiTO, the Citation Typing Ontology*, 2014 <<http://purl.org/spar/cito>>
- Shotton, david. CiTO, the Citation Typing Ontology. *J Biomed Semantics*. 2010 Jun 22;1 Suppl 1:S6. <<http://www.ncbi.nlm.nih.gov/pubmed/20626926>>
- Sistema HIVE, 2012 <http://klignon.uc3m.es/hive-es/wiki/index.php/Sistema_HIVE>
- Slavic, Aida; Cordeiro, Maria Inês. Riesthuis, Gerhard. “El desarrollo de la Clasificación Decimal Universal: 1992-2008 y más allá”. *Revista española de Documentación Científica*, 2009, Vol. 32, n. 3. doi:10.3989/redc.2009.3.658
- Slype, Georges Van; Hípola, Pedro; Moya, Félix de. *Los lenguajes de indización: concepción, construcción y utilización de los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1991
- Smith, Michael K.; Welty, Chris; McGuinness, Deborah L. *OWL Web Ontology Language Guide*, W3C Recommendation 10 February 2004 <<http://www.w3.org/TR/owl-guide/>>

- Springer
<<http://www.springer.com/open+access/authors+rights?SGWID=0-176704-12-683201-0>>
- Studer, Rudi; Benjamins, Richard; Fensel, Dieter. “Knowledge Engineering: Principles and Methods”. *Data and Knowledge Engineering*, 1998, Vol. 25 n. 1-2 pp. 161-197
- Suber, Peter. *A very brief Introduction to Open Access*. 2004. <<http://www.earlham.edu/~peters/fos/brief.htm>>
- Suber, Peter. *Budapest Open Access Initiative: Frequently Asked Questions*, 2007 <<http://www.earlham.edu/~peters/fos/boaifaq.htm>>
- Subirats Coll, Imma. “La web semántica y su aplicación en servicios de información: El caso de la Organización de las Naciones Unidas para la Alimentación y la Agricultura”. En: *Tercer Simposio Nacional De Patrimonio Bibliográfico y Documental, Colombia 30 de septiembre y 1 de octubre 2013* <<http://www.slideshare.net/faoaims/simposio-nacional-bogota>>
- Taton, René. “Las biografías científicas y su importancia en la historia de las ciencias”. En Lafuente, A; Saldaña, J.J. (Coordinadores), *Historia de las ciencia*, Madrid: CSIC, 1987. pp. 73-85;
- TEI Consortium, TEI *Guidelines for Electronic Text Encoding and Interchange* <<http://www.tei-c.org/Guidelines/>>

- *The Valparaíso declaration for improved scientific communication in the electronic medium*, 2004
<<http://eprints.rclis.org/20069/>>

- Tillett, Barbara B. *RDA Antecedentes y Aspectos de su Implementación*. Washington: Library of Congress, 2009
<http://www.loc.gov/catdir/cpsd/RDA/RDAantecedentes_instructor.pdf>

- Tomlinson, Todd. *Manual imprescindible de Drupal 7*. Madrid: Anaya Multimedia cop., 2012

- Tramullas, Jesús. "Gestión de contenidos con *Drupal*: revisión de módulos específicos para bibliotecas, archivos y museos". *El profesional de la información*, septiembre-octubre, 2013, v.22, n. 5, pp. 425-431.

- Tummarello, Giovanni; Morbidoni, Christian; Puliti, Paolo; Piazza, Francesco; "A proposal for textual encoding based on semantic web tools", *Online Information Review*, 2008 Vol. 32 n. 4, pp.467-477

- UNESCO. *Tesaurus de la UNESCO*, 2010
<<http://databases.unesco.org/thessp/>>

- UNESCO. Convención sobre la Protección del Patrimonio Mundial Cultural y Natural. París, 1972

- UNESCO. *Recomendación sobre la Protección de los bienes culturales muebles*. París, 1978

- UNESCO. *Declaración de México. Conferencia Mundial sobre políticas culturales*. París, 1982

- UNESCO. *Convención para la salvaguarda del patrimonio cultural inmaterial*. París, 2003
- Unión Europea. *EuroVoc: Tesauro multilingüe de la Unión Europea*, 2014, <<http://eurovoc.europa.eu/>>
- Universidad de los Andes. Biblioteca José Enrique Díez. *FRBR y RDA*. <<http://biblioestandares.bn.cl/sites/biblioestandares.bn.cl/files/FRBR%20y%20RDA.pdf>>
- Universal Decimal Classification Consortium (UDCC). *Universal Decimal Classification, 2013* <<http://www.udcc.org/udcsummary/about.htm>>
- Universidad de Alicante. adhesión a la Declaración de Berlín. <<http://blogs.ua.es/blogvrtie/tag/conocimientoabierto/>>
- Universitat de València Estudi General adhesión a la Declaración de Berlín. <http://www.uv.es/jroderic/documentos/Victoria_Caracteristicas_Roderic.pdf>
- Universidad Miguel Hernández de Elche <<http://comunicacion.umh.es/2013/07/09/la-umh-lanza-el-repositorio-digital-institucional-rediumh/>>
- Universitat Politècnica de València adhesión a la Declaración de Berlín <<http://poliscience.blogs.upv.es/open-access/entorno-oa/politicas-de-accesoabierto/>>
- Vaishnav, Ashwini; Sonwane, Shashank S. *Mapping*

Metadata Standards. In: International Conference on Semantic Web and Digital Libraries (ICSD) Bangalore 21-23 February, 2007.

- Vallez, Mari; Rovira, Cristòfol; Codina, Lluís; Pedraza, Rafael. “Procedimientos para la extracción de palabras clave de páginas web basados en criterios de posicionamiento en buscadores” *Anuario Hypertext.net*, 2010, n. 8 mayo,
<http://www.upf.edu/hipertextnet/numero-8/extraccion_keywords.html>
- Van de Sompel, Herbert ; Lagoze, Carl (ed.) *The Open Archives Initiative Protocol for Metadata Harvesting*. <<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>>
- Vianello Osti, Marina. *El hipertexto entre la utopía y la aplicación: identidad, problemática y tendencias de la Web*. Gijón: TREA, 2004.
- Van Noorden, Richard “Chinese agencies announce open-access policies”. *Nature*, May 2014, 19 <<http://www.nature.com/news/chinese-agencies-announce-open-access-policies-1.15255#/related-links>>
- Vilches-Blázquez, Luis Manuel; Corcho, Óscar; Rodríguez Pascual, Antonio; Bernabe Poveda, Miguel Angel. “Web Semántica e Información Geográfica: Una interrelación necesaria ante las problemáticas actuales”. *Revista Mapping*, 2008, 125. pp. 76-81 <http://oa.upm.es/5133/2/RevistaMapping_Vilches_BlazquezLM.pdf>

- *Washington D.C. principles for free access to science*, 2004. <<http://www.dcpinciples.org/statement.htm>>

- *Water & Stone. Open Source CMS Market Share Report: 4th annual report on the industry*. Water & Stone, November 2011
<<http://waterandstone.com/portfolio/2011-open-source-cms-market-share-report/>>

- *Wellcome Trust. Position statement in support of open access publishing*. 2003
<<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Policy/index.htm>>

- *Wood, David; Zaidman, Marsha; Ruth, Luke. Linked Data: structured data on de web*. NY: Manning, 2014

- *World Wide Web Consortium (W3C). Describing Linked Datasets with the VoID Vocabulary*, 2011<<http://www.w3.org/TR/void/>>

- *World Wide Web Consortium (W3C). The RDF Data Cube Vocabulary*. January 2014<<http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/>>

- *World Wide Web Consortium (W3C). Use Case FAO Authority Description Concept Scheme*, 2010
<http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_FAO_Authority_Description_Concept_Scheme>

- *World Wide Web Consortium (W3C). Semantic Web*, 2013. <<http://www.w3.org/standards/semanticweb/>>

- World Wide Web Consortium (W3C). *XML Technology*, 2010a <<http://www.w3.org/standards/xml/>>
- World Wide Web Consortium (W3C). *Guía breve de tecnologías XML*, 2010b <<http://www.w3c.es/Divulgacion/GuiasBreves/TecnologiasXML>>
- World Wide Web Consortium (W3C), *Namespaces in XML*, 2009 <<http://www.w3.org/TR/2009/REC-xml-names-20091208/>>
- Ximeno, Vicente. *Escritores de reyno de Valencia: chronologicamente ordenados desde el año MCCXXXVIII de la Christiana Conquista hasta el de MDCCXLVIII*, 2 vols, Valencia, 1748-1749
- Zhang, Xiaolin “Development of open access in China: strategies, practices, challenges” *Insights*, 2014. vol. 27, n. 1, pp. 45-50 <<http://dx.doi.org/10.1629/2048-7754.111>>

SIGLAS Y ACRÓNIMOS

AACR	Anglo American Cataloguin Rules
AGRIS	International System for Agricultural Science and Technology
AI	Arquitectura de la Información
ALA	American Library Association
ANSI	American National Standards Institute
ARC	Augmentation Research Centre
ARPA	Agencia de Investigación de Proyectos Avanzados
BIBO	Bibliographic Ontology
BIVALDI	Biblioteca Valenciana Digital
BL	British Library
BNE	Biblioteca Nacional Española
BnF	French National Library
BOAI	Budapest Open Access Initiative
CAS	Chinese Academy of Sciences
CDU	Clasificación Decimal Universal

CERN	Organización Europea para la Investigación Nuclear
CILIP	Chartered Institute of Library and Information Professionals
CNI	Liga para la Información en Red
CORC	Cooperative Online Resource Catalog
CRUE	Conferencia de Rectores de las Universidades Españolas
CSIC	Consejo Superior de Investigaciones Científicas
CSS	Cascading Style Sheet
DLB	Federación de Bibliotecas Digitales
DC	Dublin Core
DCMI	Dublin CoreMetadata Initiative
DRIVER	Digital Repository Infrastructure Vision for European Research
DTD	Document Type Definition
EURAB	European Research Advisory Board
EURIG	European RDA Interest Group
FAO	Food and Agriculture Organization

FECYT	Fundación Española para la Ciencia y la Tecnología
FOAF	Friend of a Friend
FP7	Séptimo Programa Marco de Investigación
FRAD	Functional Requirements for Authority Data
FRBR	Functional Requirements for Bibliographic Records
FRSAD	Functional Requirements for Authority Data
FTP	File Transfer Protocol
GND	German National Library
HTML	HyperText Markup Language
HTTP	Hipertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IFLA	International Federation of Library Association and Institutions
IMLS	Institute of Museum and Library Services
INE	Instituto Nacional de Estadística

ISBD	International Standard Bibliographic Description
ISO	International Organization for Standardization
JSC	Joint Steering Committee
LA	Library Association UK
LC	Library of Congress EEUU
LD	Linked Data
LEM	Lista de Encabezamientos de Materia
LLD-XS	Grupo Incubador de Datos Vinculados de Bibliotecas del W3C
LOD	Linked Open Data
LOV	Linked Open Vocabularies
NAL	National Library of Agriculture
NCSA	National Center for Supercomputing Applications
NLM	National Library of Medicine
NSF	National Science Foundation
NSFC	National Natural Science Foundation of China

MADS	Metadata Authority Description Schema
MARC	Machine Readable Cataloguing
MeSH	Medical Subject Headings Thesaurus
METS	Metadata Encoding and Transmission Standard
MIT	Massachusetts Institute of Technology
MODS	Metadata Object Description Schema
OA	Open Access
OAI	Open Archives Initiative
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
OD	Open Data
OKFN	Open Knowledge Foundation
OPAC	Online Public Access Catalog
OSI	Open Source Initiative
OWL	Web Ontology Language

RDA	Resource Description Access
RDF	Resource Description Framework
REBIUN	Red de Bibliotecas Universitarias
SCI	Science Citation Index
SELL	Southern European Libraries Link
SKOS	Simple Knowledge Organization System
SW	Semantic Web
TDX	Tesis Doctorales en Red
TED	Technology Entertainment and Design
TEI	Text Encoding Initiative
UDCC	Universal Decimal Classification Consortium
UN	United Nations
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VIAF	Virtual International Authority File
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	Extensible Markup Language

GLOSARIO

Bibliographic Ontology (BIBO)	Ontología para describir citas y referencias bibliográficas
Cliente / Servidor	Modelo de interacción entre un PC y un servidor web: el “cliente” PC (Internet Explorer, Firefox, Netscape) solicita una página web al servidor. El servidor envía un archivo codificado HTML al cliente.
Datasets, Conjuntos de datos	Conjuntos de metadatos estructurados o descripciones de objetos digitales. La descripción de una entidad se compone de elementos de un conjunto de metadatos como MARC 21 o Dublin Core, y de un valor para cada elemento. Los valores de los elementos de metadatos se toman de los vocabularios controlados o son texto libre (W3C, 2011).
Cascading Style Sheet CSS	archivo secundario usado por una página web, que contiene instrucciones para el formato y la presentación del contenido del sitio web.
Dublin Core Metadata Element Set	Vocabulario compuesto por 15 propiedades para la descripción de recursos digitales. El vocabulario más utilizado en WS.

Dublin Core Metadata Element Terms	Extension del conjunto de términos Dublin Core Metadata Element Set para la descripción bibliográfica de publicaciones en la web.
Document Type Definition DTD	Definición de Tipo de Documento: es una descripción de estructura y sintaxis de un documento XML o SGML. Su función básica es la descripción del formato de datos. Es la manera de decir cuál es la sintaxis correcta del lenguaje.
Friending a Friend FOAF	Vocabulario empleado en la descripción de recursos de la WS para describir personas y las relaciones entre ellas.
Functional Requirements for Authority Data FRAD	Requisitos Funcionales de los Datos de Autoridad es un modelo conceptual para la descripción de autoridades.
Functional Requirements for Bibliographic Records FRBR	Requisitos Funcionales de los Registros Bibliográficos es un modelo conceptual para la descripción de los registros bibliográficos.
Functional Requirements for Authority	Requisitos Funcionales para Datos de Autoridad de Materia es un modelo conceptual para la normalización de los

Data FRSAD	campos de materia.
Hypertexto	Es un sistema de escritura no secuencial. Se trata de bloques de texto conectados entre sí mediante unos enlaces o hipervínculos que permiten navegar de un texto a otro.
HyperText Markup Language HTML	Lenguaje de marcas para la World Wide Web que mediante un conjunto limitado de marcas o etiquetas define la estructura del documento y todos los elementos de visualización.
Hipertext Transfer Protocol HTTP	Uno de los protocolos estandarizados más comunes de Internet para transferir datos
International Federation of Library Association and Institutions IFLA	Organismo Internacional no gubernamental encargada de la promoción, cooperación, investigación y desarrollo de todo lo relacionado con la gestión de la información y las bibliotecas. Responsable de las normas bibliográficas.
Internet profunda Internet invisible	Nombre inadecuado para referirse al sector de sitios y de páginas web que no pueden indizar los motores de búsqueda de uso público como Google o AltaVista (Codina, 2003).
International	La Descripción Bibliográfica Internacional

Standard Bibliographic Description ISBD	Normalizada es norma internacional que especifica los requisitos para la descripción bibliográfica de los recursos de información que pueden formar parte de las colecciones de las bibliotecas.
Linked Data LD	Un patrón para la descripción de conjuntos de datos mediante hiperenlaces legibles por máquina, lenguaje RDF e identificadores URI, que son los elementos básicos de la WS.
Linked Open Data LOD	Son datos enlazados publicados en la web con licencias que permiten la reutilización.
Linked Open Data Cloud Diagram	Diagrama que representa el conjunto de datos abiertos enlazados publicados en la web en el marco del proyecto Linking Open Data Project 2007-2011
Open Access OA	Movimiento de acceso libre al conocimiento. Esta filosofía incluye el acceso abierto a la ciencia y a la información, sin barreras económicas ni restricciones de derechos de autor, con miras a la difusión y a la reutilización del conocimiento a todos los usuarios para continuar con el ciclo de transferencia de la información.
Online Computer	Es una entidad de investigación sin ánimo de lucro de la que forman parte más de

Library Center OCLC	15.000 bibliotecas de múltiples países de todo el mundo. Mantienen el Worldcat, el mayor catálogo de biblioteca en línea de acceso público del mundo (OCLC, 2014).
Open DataOD	Es una iniciativa que promueve la liberación de datos, procedentes de la investigación, de organizaciones o gubernamentales, en formatos reutilizables.
Online Public Access Catalog OPAC	Es el catálogo digital de biblioteca de acceso libre.
Open Source Initiative OSI	Es la iniciativa que promueve un movimiento a favor del software de código fuente abierto.
Web Ontology Language OWL	Lenguaje de representación de ontologías, estándar del W3C, para dotar de significado las tripletas de RDF definiendo clases, propiedades y relaciones.
Resource Description Access RDA	Recursos, Descripción y Acceso. Es un modelo para la representación de recursos y de contenido de cualquier soporte de almacenamiento en entornos electrónicos. Están basadas en el modelo conceptual FRBR y FRAD para la descripción y el acceso de los recursos digitales.
Resource	Lenguaje de definición de metadatos para

Description Framework RDF	describir objetos de información distribuidos en la web de forma estructurada. Es el lenguaje basado en tripletas de la web semántica para identificar unívocamente y dotar de significado entidades informativas.
Simple Knowledge Organization System SKOS	Lenguaje de representación del conocimiento para migrar los sistemas de organización del conocimiento (tesauros, taxonomías) a los lenguajes de ontologías de la web semántica.
SPARQL Protocol and RDF Query Language	Lenguaje estandarizado de consulta para datos RDF. Es el equivalente del lenguaje SQL de consulta a bases de datos relacionales.
SPARQL Endpoint	Servicio de consulta al que se interroga con el lenguaje SPARQL para localizar conjuntos de datos descritos en RDF.
Semantic Web SW	Extensión de la World Wide Web basada en datos legibles por máquina, descritos con RDF, identificados con URI permanentes y descritos con vocabularios de metadatos.
Triplestore	Repositorio o base de datos RDF que almacena tripletas RDF.
Uniform Resource Identifier URI	Identificador global para recursos web. Identifican de forma unívoca los recursos. Un URI puede ser un URL y un URN.

Uniform Resource Locator URL	Es un localizador de recursos web, una dirección estandarizada para localizar un recurso. Está dentro del concepto más general URI.
Virtual International Authority File VIAF	El Fichero de Autoridades Virtual Internacional es un proyecto en colaboración entre bibliotecas nacionales y OCLC para normalizar y estandarizar las autoridades.
World Wide Web Consortium W3C	Consortio Internacional que producen recomendaciones sobre las tecnologías y protocolos de la World Wide Web.
World Wide Web WWW	Conocida como la web, es la red informática mundial para la transferencia de documentos de hipertexto a través de Internet.
Extensible Markup Language XML	Lenguaje de marcas para la codificación e intercambio de información de forma automatizada. Mediante el uso de etiquetas diferencian la estructura del documento de la información que contiene.

ANEXOS

Anexo 1: Materias por las que clasifica WOS

Acoustics (Acústica)	International Relations (Relaciones Internacionales)
Agricultural Economics & Policy (Política y Economía Agrícola)	
Agricultural Engineering (Ingeniería Agrícola)	Language & Linguistics (Lengua y Lingüística)
Agriculture, Dairy & Animal Science (Ciencia de los Animales, Productos Lácteos y Agricultura)	Law (Derecho)
Agriculture, Multidisciplinary (Agricultura, Multidisciplinar)	Limnology (Limnología)
Agronomy (Agronomía)	Linguistics (Lingüística)
Allergy (Alergia)	Literary Reviews (Críticas Literarias)
Anatomy & Morphology (Anatomía y Morfología)	Literary Theory & Criticism (Crítica y Teoría de la Literatura)
Andrology (Andrología)	Literature (Literatura)
Anesthesiology (Anestesiología)	Literature, African, Australian, Canadian (Literatura, Africana, Australiana, Canadiense)
Anthropology (Antropología)	Literature, American (Literatura, Americana)
Archaeology (Arqueología)	Literature, British Isles (Literatura, Islas Británicas)
Architecture (Arquitectura)	Literature, German, Dutch, Scandinavian (Literatura, Alemana, Holandesa, Escandinava)
Area Studies (Estudios de Área)	Literature, Romance (Literatura, Romance)
Art (Arte)	Literature, Slavic (Literatura, Eslava)
Asian Studies (Estudios Asiáticos)	Logic (Lógica)
Astronomy & Astrophysics (Astronomía y Astrofísica)	Management (Gestión)
Audiology & Speech-Language Pathology (Audiología y Patología del Habla y del Lenguaje)	Marine & Freshwater Biology (Biología Marina y de Agua Dulce)
Automation & Control Systems (Sistemas de Control y Automatización)	Materials Science, Biomaterials (Ciencia de los Materiales, Biomateriales)
Behavioral Sciences (Ciencias del Comportamiento)	Materials Science, Ceramics (Ciencia de los Materiales, Cerámica)
Biochemical Research Methods (Métodos de Investigación Bioquímica)	Materials Science, Characterization & Testing (Ciencia de los Materiales, Caracterización y Testeo)
Biochemistry & Molecular Biology (Bioquímica y Biología Molecular)	

Biodiversity Conservation (Conservación de la Biodiversidad)	Materials Science, Coatings & Films (Ciencia de los Materiales, Revestimientos y Película)
Biology (Biología)	
Biophysics (Biofísica)	Materials Science, Composites (Ciencia de los Materiales, Compuestos)
Biotechnology & Applied Microbiology (Biotecnología y Microbiología Aplicada)	Materials Science, Multidisciplinary (Ciencia de los Materiales, Multidisciplinar)
Business (Negocios)	
Business, Finance (Negocios, Finanzas)	Materials Science, Paper & Wood (Ciencia de los Materiales, Papel y Madera)
Cardiac & Cardiovascular Systems (Sistemas Cardíaco y Cardiovascular)	Materials Science, Textiles (Ciencia de los Materiales, Textil)
Cell & Tissue Engineering (Ingeniería de Tejidos y Células)	Mathematical & Computational Biology (Biología Computacional y Matemática)
Cell Biology (Biología Celular)	Mathematics (Matemáticas)
Chemistry, Analytical (Química, Analítica)	Mathematics, Applied (Matemáticas, Aplicadas)
Chemistry, Applied (Química, Aplicada)	Mathematics, Interdisciplinary Applications (Matemáticas, Aplicaciones Interdisciplinarias)
Chemistry, Inorganic & Nuclear (Química, Inorgánica y Nuclear)	Mechanics (Mecánica)
Chemistry, Medicinal (Química, Medicinal)	Medical Ethics (Ética Médica)
Chemistry, Multidisciplinary (Química, Multidisciplinar)	Medical Informatics (Informática Médica)
Chemistry, Organic (Química, Orgánica)	Medical Laboratory Technology (Tecnología Médica de Laboratorio)
Chemistry, Physical (Química, Física)	Medicine, General & Internal (Medicina, General e Interna)
Classics (Clásicos)	Medicine, Legal (Medicina, Legal)
Clinical Neurology (Neurología Clínica)	Medicine, Research & Experimental (Medicina, Experimental y de Investigación)
Communication (Comunicación)	Medieval & Renaissance Studies (Estudios Renacentistas y Medievales)
Computer Science, Artificial Intelligence (Informática, Inteligencia Artificial)	Metallurgy & Metallurgical Engineering (Metalurgia e Ingeniería Metalúrgica)
Computer Science, Cybernetics (Informática, Cibernética)	Meteorology & Atmospheric Sciences

Computer Science, Hardware & Architecture (Informática, Hardware y Arquitectura)	(Meteorología y Ciencias de la Atmósfera)
Computer Science, Information Systems (Informática, Sistemas de Información)	Microbiology (Microbiología)
Computer Science, Interdisciplinary Applications (Informática, Aplicaciones Interdisciplinarias)	Microscopy (Microscopía)
Computer Science, Software Engineering (Informática, Ingeniería de Software)	Mineralogy (Mineralogía)
Computer Science, Theory & Methods (Informática, Teoría y Métodos)	Mining & Mineral Processing (Minería y Tratamiento de Minerales)
Construction & Building Technology (Construcción y Tecnología de la Edificación)	Multidisciplinary Sciences (Ciencias Multidisciplinarias)
Criminology & Penology (Criminología y Penología)	Music (Música)
Critical Care Medicine (Medicina Intensiva)	Mycology (Micología)
Crystallography (Cristalografía)	Nanoscience & Nanotechnology (Nanociencia y Nanotecnología)
Cultural Studies (Estudios Culturales)	Neuroimaging (Neuroimagen)
Dance (Baile)	Neurosciences (Neurociencias)
Demography (Demografía)	Nuclear Science & Technology (Ciencia y Tecnología Nuclear)
Dentistry, Oral Surgery & Medicine (Odontología, Medicina y Cirugía Oral)	Nursing (Enfermería)
Dermatology (Dermatología)	Nutrition & Dietetics (Nutrición y Dietética)
Developmental Biology (Biología del Desarrollo)	Obstetrics & Gynecology (Obstetricia y Ginecología)
Ecology (Ecología)	Oceanography (Oceanografía)
Economics (Economía)	Oncology (Oncología)
Education & Educational Research (Educación e Investigación Educativa)	Operations Research & Management Science (Ciencia de la Gestión e Investigación de Operaciones)
Education, Scientific Disciplines (Educación, Disciplinas Científicas)	Ophthalmology (Oftalmología)
	Optics (Óptica)
	Ornithology (Ornitología)
	Orthopedics (Ortopedia)
	Otorhinolaryngology (Otorrinolaringología)
	Paleontology (Paleontología)
	Parasitology (Parasitología)
	Pathology (Patología)
	Pediatrics (Pediatría)
	Peripheral Vascular Disease

Education, Special (Educación, Especial)	(Enfermedad Vascular Periférica)
Electrochemistry (Electroquímica)	Pharmacology & Pharmacy (Farmacología y Farmacia)
Emergency Medicine (Medicina de Emergencia)	Philosophy (Filosofía)
Endocrinology & Metabolism (Endocrinología y Metabolismo)	Physics, Applied (Física, Aplicada)
Energy & Fuels (Energía y Combustibles)	Physics, Atomic, Molecular & Chemical (Física, Atómica, Molecular y Química)
Engineering, Aerospace (Ingeniería, Aeroespacial)	Physics, Condensed Matter (Física, Materia Condensada)
Engineering, Biomedical (Ingeniería, Biomédica)	Physics, Fluids & Plasmas (Física, Fluidos y Plasmas)
Engineering, Chemical (Ingeniería, Química)	Physics, Mathematical (Física, Matemática)
Engineering, Civil (Ingeniería, Civil)	Physics, Multidisciplinary (Física, Multidisciplinar)
Engineering, Electrical & Electronic (Ingeniería, Eléctrica y Electrónica)	Physics, Nuclear (Física, Nuclear)
Engineering, Environmental (Ingeniería, Ambiental)	Physics, Particles & Fields (Física, Partículas y Campos)
Engineering, Geological (Ingeniería, Geológica)	Physiology (Fisiología)
Engineering, Industrial (Ingeniería, Industrial)	Planning & Development (Planificación y Desarrollo)
Engineering, Manufacturing (Ingeniería, Fabricación)	Plant Sciences (Botánica)
Engineering, Marine (Ingeniería, Marina)	Poetry (Poesía)
Engineering, Mechanical (Ingeniería, Mecánica)	Political Science (Ciencias Políticas)
Engineering, Multidisciplinary (Ingeniería, Multidisciplinar)	Polymer Science (Ciencia de Polímeros)
Engineering, Ocean (Ingeniería, Oceánica)	Primary Health Care (Atención Primaria de Salud)
Engineering, Petroleum (Ingeniería, Petróleo)	Psychiatry (Psiquiatría)
Entomology (Entomología)	Psychology (Psicología)
Environmental Sciences (Ciencias Ambientales)	Psychology, Applied (Psicología, Aplicada)
	Psychology, Biological (Psicología, Biológica)
	Psychology, Clinical (Psicología, Clínica)
	Psychology, Developmental (Psicología, del Desarrollo)

Environmental Studies (Estudios Ambientales)	Psychology, Educational (Psicología, Educativa)
Ergonomics (Ergonomía)	Psychology, Experimental (Psicología, Experimental)
Ethics (Ética)	Psychology, Mathematical (Psicología, Matemática)
Ethnic Studies (Estudios Étnicos)	Psychology, Multidisciplinary (Psicología, Multidisciplinar)
Evolutionary Biology (Biología Evolutiva)	Psychology, Psychoanalysis (Psicología, Psicoanálisis)
Family Studies (Estudios de Familia)	Psychology, Social (Psicología, Social)
Film, Radio, Television (Radio, Televisión y Cine)	Public Administration (Administración Pública)
Fisheries (Pesquería)	Public, Environmental & Occupational Health (Salud Pública, Ambiental y Ocupacional)
Folklore (Folclore)	Radiology, Nuclear Medicine & Medical Imaging (Radiología, Imagen Médica y Medicina Nuclear)
Food Science & Technology (Ciencia y Tecnología de los Alimentos)	Rehabilitation (Rehabilitación)
Forestry (Ingeniería de Montes)	Religion (Religión)
Gastroenterology & Hepatology (Gastroenterología y Hepatología)	Remote Sensing (Teledetección)
Genetics & Heredity (Genética y Herencia Genética)	Reproductive Biology (Biología Reproductora)
Geochemistry & Geophysics (Geoquímica y Geofísica)	Respiratory System (Sistema Respiratorio)
Geography (Geografía)	Rheumatology (Reumatología)
Geography, Physical (Geografía, Física)	Robotics (Robótica)
Geology (Geología)	Social Issues (Asuntos Sociales)
Geosciences, Multidisciplinary (Ciencias de la Tierra, Multidisciplinar)	Social Sciences, Biomedical (Ciencias Sociales, Biomédica)
Geriatrics & Gerontology (Geriatría y Gerontología)	Social Sciences, Interdisciplinary (Ciencias Sociales, Interdisciplinar)
Gerontology (Gerontología)	Social Sciences, Mathematical Methods (Ciencias Sociales, Métodos Matemáticos)
Health Care Sciences & Services (Servicios y Ciencias de la Salud)	Social Work (Trabajo Social)
Health Policy & Services (Servicios y Política Sanitaria)	Sociology (Sociología)
Hematology (Hematología)	Soil Science (Ciencia del Suelo)
History (Historia)	
History & Philosophy of Science	

(Historia y Filosofía de la Ciencia)	Spectroscopy (Espectroscopia)
History of Social Sciences (Historia de las Ciencias Sociales)	Sport Sciences (Ciencias de la Actividad Física)
Horticulture (Horticultura)	Statistics & Probability (Estadística y Probabilidad)
Hospitality, Leisure, Sport & Tourism (Hospitalidad, Ocio, Deporte y Turismo)	Substance Abuse (Abuso de Sustancias)
Humanities, Multidisciplinary (Humanidades, Multidisciplinar)	Surgery (Cirugía)
Imaging Science & Photographic Technology (Tecnología Fotográfica y Ciencias de la Imagen)	Telecommunications (Telecomunicaciones)
Immunology (Inmunología)	Theater (Teatro)
Industrial Relations & Labor (Trabajo y Relaciones Laborales)	Thermodynamics (Termodinámica)
Infectious Diseases (Enfermedades Infecciosas)	Toxicology (Toxicología)
Information Science & Library Science (Ciencias de la Información y Biblioteconomía)	Transplantation (Trasplante)
Instruments & Instrumentation (Instrumentos e Instrumentación)	Transportation (Transporte)
Integrative & Complementary Medicine (Medicina Complementaria e Integral)	Transportation Science & Technology (Ciencia y Tecnología del Transporte)
	Tropical Medicine (Medicina Tropical)
	Urban Studies (Estudios Urbanos)
	Urology & Nephrology (Urología y Nefrología)
	Veterinary Sciences (Ciencias Veterinarias)
	Virology (Virología)
	Water Resources (Recursos Hídricos)
	Women's Studies (Estudios de la Mujer)
	Zoology (Zoología)

Anexo 2: Modelo de documentos de acuerdo de cesión de derechos con los archivos

CARTA DE AUTORIZACIÓN DE LOS FAMILIARES DE
.....
PARA LA CONSULTA Y REPRODUCCIÓN DEL ARCHIVO ACADÉMICO Y PERSONAL DEL CIENTÍFICO
D.
UBICADOS EN EL ARCHIVO DE
DE VALENCIA

-De acuerdo con 57.1c de la Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español, que determina que *los documentos que contengan datos personales de carácter policial, procesal, clínico o de cualquier otra índole que puedan afectar a la seguridad de las personas, a su honor, a la intimidad de su vida privada y familiar y a su propia imagen, no podrán ser públicamente consultados sin que medie consentimiento expreso de los afectados o hasta que haya transcurrido un plazo de veinticinco años desde su muerte, si su fecha es conocida, o, en otro caso, de cincuenta años, a partir de la fecha de los documentos*

SOLICITAMOS:

A nos permitan la consulta y la reproducción de la documentación del EXPEDIENTE ACADÉMICO Y DE LOS ARCHIVOS PERSONALES DE de con la finalidad de realizar un estudio biográfico y bibliométrico del insigne Académico, dentro de la línea de investigación abierta en la Universidad Católica de Valencia San Vicente Mártir: "Científicos valencianos del siglo XX"

HIJOS:

SOLICITANTES:

.....
Dirección:
DNI:
Teléfono y e.mail particular:
Teléfono y e.mail del trabajo:
Tno móvil:

Firmado:
.....
DNI

Anexo 3: Protocolo de trabajo para las fuentes orales

Yolanda Blasco Gil propone un protocolo de trabajo en las fuentes orales¹, que se puede resumir en cinco partes:

Entrevista inicial: toma de contacto con la persona que se va a entrevistar, que puede ser directamente o a través de un familiar, amigo o compañero. Tras una breve entrevista inicial, se elegirá el tema sobre el que versará la entrevista definitiva y que, dependiendo del interés que, el entrevistado haya mostrado en ese primer contacto, variará en cada caso concreto. El tema se podrá acotar temáticamente y cronológicamente.

Elaboración de un cuestionario o programa previo: se utiliza para delimitar las diferentes facetas o aspectos que van a abordarse. Se establecerán las líneas generales, los puntos a examinar, como un índice previo, sujeto a revisiones, de la cuestión. En este momento, se deberán consultar fuentes bibliográficas escritas, para conocer las hipótesis de partida y el contexto histórico, seleccionando la bibliografía más importante. Se determinarán los archivos que deben consultarse.

¹Blasco, Yolanda. *Protocolo de trabajo en las fuentes orales*. Este documento es inédito, se ha elaborado para realizar entrevistas a los personajes biografiados y sus familiares, en el proyecto “*científicos y humanistas valencianos del siglo XX. VESTIGIUM*”.

Cuestionario:

1- Relaciones familiares:

- ¿Cuál es su ciudad natal?
- ¿Su familia era originaria de allí?
- ¿Existía algún vínculo que le condicionó al estudio de su disciplina?
- ¿Su formación familiar fue importante para que tuviera la inquietud de entrar en la universidad?

2- La universidad:

- ¿Cómo fue el pase de la preparatoria a la facultad?
- ¿Tiene antecedentes familiares con su especialidad?
- ¿Cómo era la universidad que le tocó vivir?
- ¿Cómo fue llegar a la universidad de los años 60 con una visión como la suya? ¿Cómo lo vivió?
- ¿Y los profesores? Háblenos de algunos de los profesores que le influyeron y de alguno de sus compañeros.
- ¿Tuvo relación en sus años de estudiante con otros profesores europeos? ¿Viajó al exterior?

- ¿Dispuso de alguna beca de estudio durante la carrera o después?

3- Motivación de su especialidad:

- ¿Usted había considerado estudiar su especialidad?
- ¿Cómo surgió la vocación por la especialidad que desarrolla?
- ¿Usted se había formado en un ambiente relacionado con su campo de estudio?
- ¿Recuerda algún momento de impacto en su ámbito de trabajo?

4- Visión como mujer / varón:

- Desde el punto de vista de mujer o varón en el campo que desarrolla, ¿cuál es su implicación psicológica en los casos que lleva?
- ¿Cree que ha encontrado suficiente reconocimiento a la labor que usted desempeña?
- ¿Cree que es igual al mundo de masculino?
- ¿Cómo es su visión como mujer/varón de la universidad y de la ciencia en general?
- Y ¿cuál es su visión de la universidad actual y el papel de la mujer (en su caso)?

5- Publicaciones:

- ¿Cuál fue el tema de su tesis doctoral y su inquietud por ese tema? ¿Se publicó?
- Libros publicados.
- Artículos publicados en libros.
- Artículos publicados en revistas con arbitraje (Comité Científico).
- Artículos publicados en memorias extendidas.
- Tesis dirigidas.
- Publicaciones en soporte electrónico.
- Reseñas.
- Comunicaciones, ponencias, simposios.
- Publicaciones de difusión.
- Traducciones.

Tratamiento de la bibliografía y recogida de fuentes: en la anterior fase se ha seleccionado la bibliografía y las fuentes a consultar, que irán ampliándose constantemente. Se procederá a la lectura de la bibliografía y fuentes, siendo éstas la base esencial para decidir lo que ocurrió o los datos de la cuestión, ya que son escritos y documentos coetáneos.

La lectura de las fuentes nos dejará ir depurando los datos derivados en la entrevista. Se recogerán en fichas en donde se

extraerán, de forma literal o extractada, aquellos datos o noticias que nos interesen:

Tabla 25: Campos propuestos por la Dr. Yolanda Blasco para las biografías de Vestigium

Campo	Elementos del campo
Nombre	Apellidos y Nombre
Tipo de Autor y Clasificación: CDU	Identificación (profesores, políticos, investigadores)
Autoridad bibliográfica	Autoridad bibliográfica
Datos Personales	Fecha de nacimiento Lugar de nacimiento Condición socio-económica (origen social) Religión Ideología política
Otros datos biográficos: estudios, titulaciones, líneas de investigación, otras actividades profesionales	Biografía Estudios Carrera Otras actividades Lugares de vinculación Honores y premios
Datos de contacto	Información de contacto
Estudios suyos y sobre su obra	Publicaciones
Otros datos	Relaciones familiares o laborales (matrimonio, vínculos familiares y profesionales)
Fuentes	Fuentes y otros recursos web

Fuente: Yolanda Blasco²

Valoración crítica de las fuentes: para determinar su valor y sentido, deberá figurar el autor, lugar, fecha, tipos de documentos – orales o escritos- y comentar la fuente.

Transcripción – redacción: con los datos adquiridos, se podrán obtener conclusiones y realizar la transcripción y redacción del documento –entrevista en su caso o la composición audiovisual-.

² Este modelo de registro biográfico propone campos para las biografías, que coinciden con los propuestos en este proyecto: MARC 21/RDA, con sus correspondientes equivalencias de formatos.